Basics in Applied Mathematics – Part I: Numerics

Lecture 7: Eigenvalue problems — Gershgorin discs, the Rayleigh quotient, the power method, inverse iteration

JProf. Dr. Diyora Salimova

Department of Applied Mathematics, University of Freiburg

10 November 2026

Theorem 4.1 (singular value decomposition)

Every $A \in \mathbb{R}^{m\times n}$ can be written as $A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i\, u_i v_i^\top$ with U, V orthogonal, $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, $r = \operatorname{rank} A$.

Proposition 4.3: $\|A\|_2 = \sigma_1$, $\quad \|A\|_F = \big(\sum_i \sigma_i^2\big)^{1/2}$, $\quad \operatorname{cond}_2(A) = \sigma_1/\sigma_n$ for regular A.

Pseudoinverse (Definition 4.4, Theorem 4.5): $A^+ = V\Sigma^+ U^\top = \sum_{i=1}^{r} \sigma_i^{-1}\, v_i u_i^\top$; $\quad A^+ b$ is the least squares solution of minimal norm.

Theorem 4.10 (Eckart–Young–Mirsky)

$A_k := \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ is a best approximation of rank at most k:

$$\min_{\operatorname{rank} B \le k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}, \qquad \min_{\operatorname{rank} B \le k} \|A - B\|_F = \|A - A_k\|_F = \Big(\sum_{i>k} \sigma_i^2 \Big)^{1/2}.$$

Image compression: A_{20} of a 600×512 picture is clearly recognisable with $7.2\,\%$ of the storage.

Today: Chapter 5. The SVD rests on the eigenvalues of $A^\top A$ — so far a black box. How are eigenvalues *computed*?

Not from the characteristic polynomial — iteratively, with the matrix.

$Av = \lambda v$ with $v \neq 0$: $\lambda \in \mathbb{C}$ **eigenvalue**, $v \in \mathbb{C}^n$ **eigenvector**; the eigenvalues are the roots of $p_A(\lambda) = \det(\lambda I - A)$.

- A, $S^{-1}AS$ and $A^\top$ have the same eigenvalues
- A **diagonalisable** (over $\mathbb{C}$): $A = V\Lambda V^{-1} \Leftrightarrow$ n linearly independent eigenvectors
- A **symmetric**: real eigenvalues, orthonormal eigenvectors, $A = Q\Lambda Q^\top$

Goals of the chapter

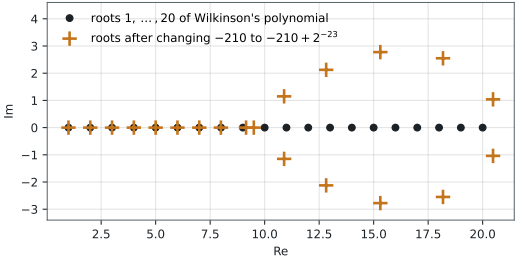
(i) why eigenvalues are *not* computed as roots of p_A
(ii) localise them: Gershgorin discs
(iii) estimate them: the Rayleigh quotient
(iv) compute them: the power method and inverse iteration with shifts
(v) apply them: PageRank (Thursday)

Why not the characteristic polynomial? (1/2)

Wilkinson's polynomial $p(\lambda) = \prod_{k=1}^{20}(\lambda - k) = \lambda^{20} - 210\,\lambda^{19} + \ldots$ roots $1, \ldots, 20$.

Change the coefficient -210 to $-210 + 2^{-23}$ (relative change $\approx 6 \cdot 10^{-10}$).

Where are the roots now?



Ten roots are non-real, with imaginary parts up to 2.8. But p is the characteristic polynomial of $\mathrm{diag}(1, \ldots, 20)$, and every eigenvalue of $\mathrm{diag}(1, \ldots, 20) + \Delta A$ lies within $\|\Delta A\|_\infty$ of one of $1, \ldots, 20$ (Gershgorin, Theorem 5.2). $\Rightarrow$ **work with the matrix, not with p_A.**

Why not the characteristic polynomial? (2/2)

Lemma 5.1 (companion matrix)

For $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ the **companion matrix**

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix} \quad \text{satisfies} \quad \det(\lambda I - C) = p(\lambda).$$

Check $n = 2$: $\det \begin{pmatrix} \lambda & a_0 \\ -1 & \lambda + a_1 \end{pmatrix} = \lambda^2 + a_1\lambda + a_0$. ✓ (general n: induction, notes)

Every monic polynomial is a characteristic polynomial. A formula for eigenvalues (arithmetic operations and radicals) would solve every polynomial equation — impossible for $n \geq 5$ (Abel–Ruffini).

$\Rightarrow$ **General-purpose eigenvalue algorithms are iterative.**

Theorem 5.2 (Gershgorin)

Let $A \in \mathbb{C}^{n\times n}$ and $D_i := \{z \in \mathbb{C} : \ |z - a_{ii}| \le r_i\}$ with $r_i := \sum_{j\neq i} |a_{ij}|, \ \ i = 1, \dots, n.$
Then every eigenvalue of A lies in $D_1 \cup \dots \cup D_n$.

Blackboard: proof.

$A^\top$ has the same eigenvalues: the theorem also holds with the **column** sums $\sum_{i\neq j} |a_{ij}|$ as radii.

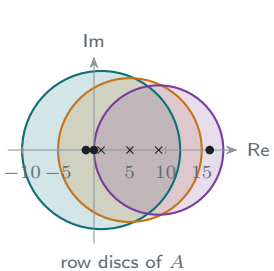
Corollary 5.3 (diagonally dominant matrices)

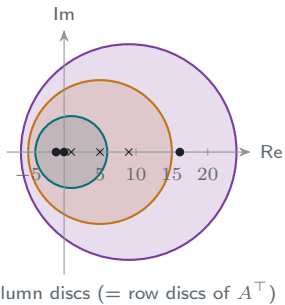
If $|a_{ii}| > r_i$ for all i (**strictly diagonally dominant**), then A is regular.
If moreover A is symmetric with positive diagonal entries, then A is SPD and $\operatorname{cond}_2(A) \le \max_i(a_{ii} + r_i) / \min_i(a_{ii} - r_i)$.

Example 5.4: Gershgorin discs of a 3×3 matrix

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}.$$ Row discs: centres $1, 5, 9$, radii $r_i = ?$ Column discs: radii $= ?$

$r = (11, 10, 9)$; columns: $(5, 10, 15)$.





Eigenvalues $(15 \pm \sqrt{297})/2 \approx 16.12,\ -1.12$ and 0: in the intersection of the two unions. 0 lies inside, so the discs cannot show regularity — indeed A is singular.

Definition 5.5

For $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n \setminus \{0\}$: $r_A(x) := \dfrac{x^\top A x}{x^\top x}$. For an eigenvector v with eigenvalue λ: $r_A(v) = \lambda$.

Theorem 5.6 (A symmetric)

Eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, orthonormal eigenvectors $v_1, \ldots, v_n$.

(i) $\lambda_n \leq r_A(x) \leq \lambda_1$ for all $x \neq 0$, with $r_A(v_1) = \lambda_1$ and $r_A(v_n) = \lambda_n$.

(ii) If $\|x\|_2 = 1$ and $\|x - v_j\|_2 = \varepsilon$, then $|r_A(x) - \lambda_j| \leq 2\,\|A\|_2\,\varepsilon^2$.

Blackboard: proof.

Eigenvector error ε $\Rightarrow$ eigenvalue error $O(\varepsilon^2)$: the eigenvalue estimator of all algorithms below.

Multiply a vector by A again and again. For diagonalisable A and $x^{(0)} = \sum_i c_i v_i$ with $c_1 \neq 0$:

$$A^k x^{(0)} = \sum_{i=1}^{n} c_i \lambda_i^k v_i = c_1 \lambda_1^k \Big(v_1 + \sum_{i=2}^{n} \frac{c_i}{c_1} \Big(\frac{\lambda_i}{\lambda_1} \Big)^k v_i \Big).$$

If $|\lambda_1| > |\lambda_i|$ for $i \geq 2$, the direction of $A^k x^{(0)}$ tends to that of v_1. Normalise in every step.

Algorithm 5.7 (power method)

Choose $x^{(0)}$ with $\|x^{(0)}\|_2 = 1$. For $k = 0, 1, 2, \ldots$: compute $y^{(k+1)} := Ax^{(k)}$; if $y^{(k+1)} = 0$, stop ($x^{(k)}$ is an eigenvector for 0); otherwise

$$x^{(k+1)} := \frac{y^{(k+1)}}{\|y^{(k+1)}\|_2}, \qquad \mu^{(k+1)} := r_A\big(x^{(k+1)}\big).$$

Convergence of the power method

Theorem 5.8

Let $A \in \mathbb{R}^{n\times n}$ be diagonalisable with $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$ (a **dominant** eigenvalue λ_1; it is real), with eigenvectors $v_1, \dots, v_n$, v_1 real, and let $x^{(0)} = \sum_i c_i v_i$ with $c_1 \neq 0$. Put $q := |\lambda_2/\lambda_1| < 1$. Then there are signs $s_k \in \{\pm 1\}$ and a constant C such that for all $k \geq 0$

$$\left\| s_k x^{(k)} - \frac{v_1}{\|v_1\|_2} \right\|_2 \leq C\, q^k \qquad \text{and} \qquad \frac{\|Ax^{(k)}\|_2}{\|x^{(k)}\|_2} \to |\lambda_1|.$$

If A is symmetric, then $|\mu^{(k)} - \lambda_1| \leq C' q^{2k}$ for all $k \geq 1$.

Blackboard: proof.

Eigenvectors: linear convergence with the rate $q = |\lambda_2/\lambda_1|$. Rayleigh quotients (symmetric A): the squared rate q^2.

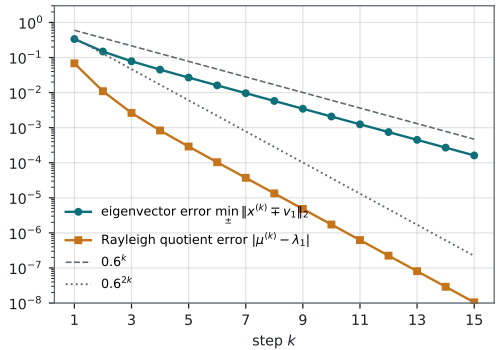
Remark 5.9 and an experiment

(a) $c_1 \neq 0$ cannot be checked in advance — choose $x^{(0)}$ at random (then $c_1 \neq 0$ with probability one).

(b) $|\lambda_1| = |\lambda_2|$ with $\lambda_1 \neq \lambda_2$: no convergence in general. $\mathrm{diag}(1, -1)$, $x^{(0)} = (1, 1)^\top / \sqrt{2}$: the iterates alternate.

(c) Cost: one matrix–vector product per step — the simplest method for very large sparse A; slow when $|\lambda_2/\lambda_1|$ is close to 1.

Experiment: symmetric 4×4, eigenvalues $1,\ 0.6,\ 0.3,\ -0.2$. **Rates?**



A 2×2 example by hand

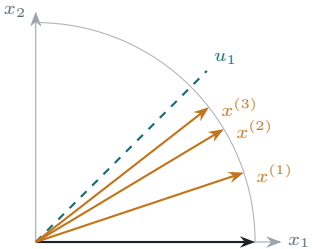
$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$: eigenvalues 4 and 2, $u_1 = \frac{1}{\sqrt{2}}(1,1)^\top$, $u_2 = \frac{1}{\sqrt{2}}(1,-1)^\top$; $x^{(0)} = e_1$, $q = \frac{1}{2}$.

Compute the directions of $Ax^{(0)}$, $A^2x^{(0)}$, $A^3x^{(0)}$.

Multiples of $(3,1)^\top$, $(5,3)^\top$, $(9,7)^\top$. In general, with $e_1 = \frac{1}{\sqrt{2}}(u_1 + u_2)$:
$A^k x^{(0)} = \frac{1}{\sqrt{2}}\big(4^k u_1 + 2^k u_2\big) = 2^{k-1}\,(2^k+1,\, 2^k-1)^\top$,

$$x^{(k)} = \frac{(2^k+1,\, 2^k-1)^\top}{\sqrt{2(4^k+1)}}, \qquad \mu^{(k)} = 4 - \frac{2}{4^k+1}.$$

$\tan\angle(x^{(k)}, u_1) = 2^k/4^k = q^k$, eigenvalue error $2/(4^k+1) < 2q^{2k}$: the squared rate.



$x^{(0)} = e_1$
$x^{(1)} \propto (3,1)^\top$, $x^{(2)} \propto (5,3)^\top$,
$x^{(3)} \propto (9,7)^\top$

The power method finds only the eigenvalue of largest modulus. Change the matrix:

Proposition 5.10

Let A have the eigenvalues $\lambda_1, \dots, \lambda_n$ with eigenvectors $v_1, \dots, v_n$, and let $\mu \in \mathbb{R}$. Then A^{-1}, $A - \mu I$ and $(A - \mu I)^{-1}$ have, with the *same* eigenvectors, the eigenvalues

$$\frac{1}{\lambda_i} \quad (A \text{ regular}), \qquad \lambda_i - \mu, \qquad \frac{1}{\lambda_i - \mu} \quad (\mu \notin \sigma(A)).$$

Which eigenvalue of $(A - \mu I)^{-1}$ is dominant? The one belonging to the eigenvalue λ_a of A **closest to the shift** μ.

Power method for $(A - \mu I)^{-1}$ $\Rightarrow$ eigenvector for λ_a, rate $q = |\lambda_a - \mu| / |\lambda_b - \mu|$ (λ_b: second closest to μ),

provided Theorem 5.8 applies to $(A - \mu I)^{-1}$: A diagonalisable, λ_a the unique eigenvalue closest to μ, $x^{(0)}$ with a component along v_a.

Inverse iteration with shift

Algorithm 5.11 (inverse iteration with shift)

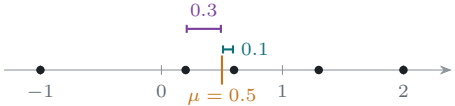
Let $\mu \notin \sigma(A)$ and $\|x^{(0)}\|_2 = 1$. Compute **once** an LU decomposition of $A - \mu I$.
For $k = 0, 1, 2, \ldots$: solve $(A - \mu I)\, y^{(k+1)} = x^{(k)}$ by two substitutions,
$x^{(k+1)} := y^{(k+1)} / \|y^{(k+1)}\|_2$, $\mu^{(k+1)} := r_A(x^{(k+1)})$.

- Convergence (Theorem 5.8 for $(A - \mu I)^{-1}$): rate q for the vectors; for symmetric A rate q^2 for the $\mu^{(k)}$.
- Cost: $O(n^3)$ once, then $O(n^2)$ per step. A good shift makes q tiny: a few steps suffice — this is how eigenvectors are computed once the eigenvalues are known.
- $\mu \approx \lambda_a$: $A - \mu I$ is nearly singular, the computed $y^{(k+1)}$ may be inaccurate. Harmless for symmetric A: the error mainly changes the *length* of $y^{(k+1)}$, and the normalisation removes it.

Experiment: inverse iteration with the shift $\mu = 0.5$

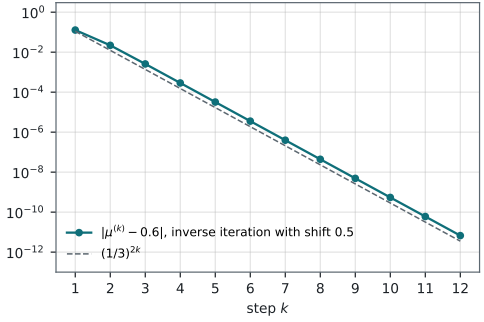
Symmetric 5×5 matrix, eigenvalues $-1,\ 0.2,\ 0.6,\ 1.3,\ 2,$ shift $\mu = 0.5$.

Which eigenvalue is found, and how fast?



$\lambda_a = 0.6$ (distance 0.1), $\lambda_b = 0.2$ (distance 0.3): $q = \frac{0.1}{0.3} = \frac{1}{3}$, so the $\mu^{(k)}$ gain a factor $q^2 = \frac{1}{9}$ per step — about one digit.

After 12 steps: error $6.7 \cdot 10^{-12}$; the measured error ratios are $0.111 = q^2$ from the third step on.



Summary

Sections 5.1–5.5

- Eigenvalues are computed **iteratively, from the matrix** — not as roots of p_A (Wilkinson; Abel–Ruffini).
- **Gershgorin:** $\sigma(A) \subset \bigcup_i D_i$, $D_i = \{|z - a_{ii}| \le r_i\}$ (rows or columns); strictly diagonally dominant $\Rightarrow$ regular.
- **Rayleigh quotient** (symmetric A): eigenvector error ε $\Rightarrow$ eigenvalue error $\le 2\|A\|_2\varepsilon^2$.
- **Power method:** dominant eigenpair, rate $|\lambda_2/\lambda_1|$. **Inverse iteration** with shift μ: eigenvalue closest to μ, rate $|\lambda_a - \mu|/|\lambda_b - \mu|$; one LU decomposition, two substitutions per step.
- **Thursday:** PageRank (Section 5.6), then Chapter 6 (polynomial interpolation). Sheet 4 on Thursday; test 1 on Friday 13 November, 13:45–14:15 (Chapters 1–4).