Numerical methods for eigenvalues and eigenvectors part I, MAT 3110 UiO

Håkon Hoel

Fall 2023

Table of Contents

Big question I

How does one compute eigenvalues and eigenvectors of a matrix $A \in \mathbb{R}^{n \times n}$?

A natural idea:

1. For $\lambda \in \mathbb{R}$, compute the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$
2. Find an eigenvalue of A by solving $p(\lambda) = 0$ using some iteration method
3. Having obtained a numerical solution $\bar{\lambda} = \lambda_K$, compute eigenvector $\bar{x} \in \mathbb{R}^n_*$ (if you also seek eigenvector) by solving $(A - \bar{\lambda} I)\bar{x} = 0$

Big question II

How does one compute eigenvalues and eigenvectors of a matrix $A \in \mathbb{R}^{n \times n}$?

A natural idea:

1. For $\lambda \in \mathbb{R}$, compute the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$
2. Find an eigenvalue of A by solving $p(\lambda) = 0$ using some iteration method
3. Having obtained a numerical solution $\bar{\lambda} = \lambda_K$, compute eigenvector $\bar{x} \in \mathbb{R}^n_*$ (if you also seek eigenvector) by solving $(A - \bar{\lambda} I)\bar{x} = 0$

But it is a bad idea because every step in your iteration method, for example,

$$\lambda_{k+1} = \lambda_k - p(\lambda_k) \left(\frac{\lambda_k - \lambda_{k-1}}{p(\lambda_k) - p(\lambda_{k-1})} \right) \qquad k = 1, 2, \ldots$$

requires that you compute the determinant $p(\lambda_k) = \det(A - \lambda_k I)$. This costs $\mathcal{O}(n^3)$ operations per iteration.

And eigenvalue λ may be complex-valued, complicating things ...

Estimates of eigenvalues

Notation: For $A \in \mathbb{R}^{n\times n}$, let $\sigma(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ denote its set of eigenvalues.

Theorem (Gershgorin's circle theorem)

Consider a matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ and associate its i-th row to the off-diagonal radius

$$r_i = \sum_{j\neq i} |a_{ij}|, \quad \text{and the } i\text{-th Gershgorin disc} \quad D_i := \{z \in \mathbb{C} \mid |z - a_{ii}| \leq r_i\}.$$

Then each eigenvalue lies inside some Gershgorin disc, $\lambda \in D_i$ for some $1 \leq i \leq n$, and thus also $\sigma(A) \subset \cup_{i=1}^n D_i$.

Theorem (Gershgorin's circle theorem (abbrv.))

For any $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, any eigv. $\lambda \in D_i$ for some $i = 1, 2, \ldots, n$.

Proof.

Let (λ, v) be an arbitrary eigenpair of A. Set $\tilde{v} = \pm v/\|v\|_\infty$ with the sign $\pm$ chosen so that $\tilde{v}_i = 1$ for some $i \in \{1, \ldots, n\}$. Then

$$(A\tilde{v})_i = (\lambda \tilde{v})_i = \lambda \qquad \text{and also} \qquad (A\tilde{v})_i = a_{ii} + \sum_{j \neq i} a_{ij} \tilde{v}_j.$$

We conclude that

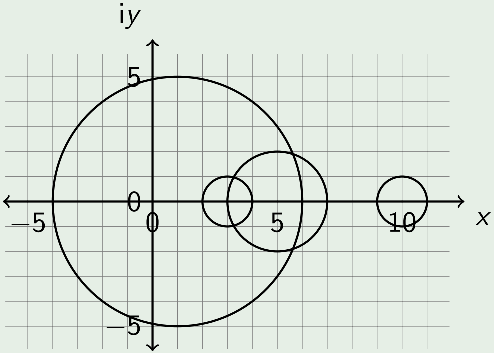
$$|\lambda - a_{ii}| \leq \Big| \sum_{j \neq i} a_{ij} \tilde{v}_j \Big| \leq \sum_{j \neq i} |a_{ij}| \underbrace{|\tilde{v}_j|}_{\leq 1} \leq \sum_{j \neq i} |a_{ij}| = r_i.$$

□

Example

$$A = \begin{bmatrix} 1 & 0 & 5 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 1 & 5 & 1 \\ 0 & 1 & 0 & 10 \end{bmatrix} \quad \text{with (approx.)} \quad \sigma(A) = \{1.80 \pm 0.61i, 5.38, 10.02\}$$

$$D_1 = B((1,0),5), \quad D_2 = B((3,0),1), \quad D_3 = B((5,0),2), \quad D_4 = B((10,0),1).$$



Observations

Remark

a) **Note:** *The theorem does not say that each Gershgorin disc contains an eigenvalue. Some discs may contain many, others none.*

b) *If all the Gershgorin discs are disjoint, then one can show that each disc D_i must contain one and only one eigenvalue.*

Theorem (Extension of Gershgorin's thm)

If the Gershgorin discs of a matrix $A \in \mathbb{R}^{n \times n}$ for some ordering satisfies that $B_1 = \cup_{i=1}^{k} D_i$ is disjoint from $B_2 = \cup_{i=k+1}^{n} D_i$ (meaning $B_1 \cap B_2 = \emptyset$), then k eigenvalues belong to B_1 and $n-k$ eigenvalues belong to B_2.

And if all discs are disjoint, then each disc contains one and only one eigenvalue.

Applications of Gershgorin's

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be strictly diagonally dominant if

$$|a_{ii}| > \underbrace{\sum_{j\neq i} |a_{ij}|}_{=r_i} \qquad \forall i \in \{1, \dots, n\}.$$

Theorem (Diagonal dominance)

Every strictly diagonally dominant matrix is non-singular.

Applications of Gershgorin's II

Theorem (Diagonal dominance)

Every strictly diagonally dominant matrix is non-singular.

Proof.

Every eigenvalue of A lies inside the union of Gershgorin discs, meaning

$$\sigma(A) \subset \cup_{i=1}^{n} D_i$$

The disc $D_i = B((a_{ii}, 0), r_i)$ does not contain point $z = (0,0) \in \mathbb{C}$ since

$$|a_{ii} - 0| = |a_{ii}| \underbrace{>}_{\text{strict diagonal dominance}} r_i.$$

$$\text{Holds for all } i \in \{1, \dots, n\} \implies 0 \notin \cup_{i=1}^{n} D_i \implies 0 \notin \sigma(A).$$

$$\text{And} \qquad \det(A) = \prod_{\lambda \in \sigma(A)} \lambda \neq 0.$$

Example

$$A = \begin{bmatrix} 2 & 1 & -1/2 \\ -1 & 3 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

is non-singular as

$$|a_{11}| = 2 > 1 + |-1/2|$$

$$|a_{22}| = 3 > |-1| + 1$$

$$|a_{33}| = |-2| > 1 + 0.$$

Similarity transformations

If $T \in \mathbb{R}^{n \times n}$ is invertible, then we recall that $T^{-1}AT$ is called a similarity transformation of A.

Note: Similarity transformations preserve the matrix spectrum $\sigma(T^{-1}AT) = \sigma(A)$, since the characteristic polynomial is preserved:

$$\begin{aligned} p_{T^{-1}AT}(\lambda) &= \det(T^{-1}AT - \lambda I) \\ &= \det(T^{-1}(A - \lambda I)T) \\ &= \underbrace{\det(T^{-1})\det(T)}_{=1}\det(A - \lambda I) \\ &= p_A(\lambda) \end{aligned}$$

Gershgorin cobined with similarity transformations

Core idea: Eigenvalues must be contained inside Gershgorin discs of A, but also inside of Gershgorin discs of $\tilde{A} = T^{-1}AT$. Information of set of discs from both matrices can give more information.

Example

The matrix

$$A = \begin{bmatrix} 10 & 2 & 3 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix}$$

has $\sigma(A) = \{10.226, 0.387 \pm 2.216i\}$ (that we assume unknown and try to estimate).

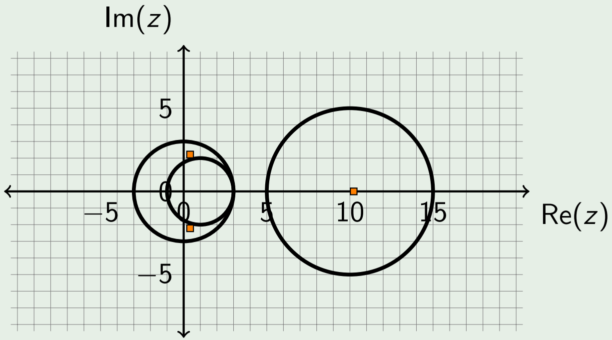
By Gershgorin's theorem,

$$D_1 = B((10,0),5), \quad D_2 = D((0,0),3), \quad D_3 = D((1,0),2).$$

Since D_1 does not intersect with $D_2 \cup D_3$, D_1 must contain one eigenvalue (and must thus be real-valued).

Gershgorin cobined with similarity transformations II

Example



To improve estimate of λ_1, consider for some $\alpha > 0$,

$$\tilde{A} = T^{-1}AT \quad \text{with} \quad T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{bmatrix} \implies \tilde{A} = \begin{bmatrix} 10 & 2\alpha & 3\alpha \\ -1/\alpha & 0 & 2 \\ 1/\alpha & -1 & 1 \end{bmatrix}$$

Gershgorin cobined with similarity transformations II

Example

Since $\sigma(\tilde{A}) = \sigma(A)$, we apply Gershgorin's theorem on $\tilde{A}$ to obtain the discs

$$\tilde{D}_1 = B((10,0), 5\alpha), \quad \tilde{D}_2 = B((0,0), 2 + 1/\alpha), \quad \tilde{D}_3 = B((1,0), 1 + 1/\alpha).$$

Choose $\alpha > 0$ so small that

$$\tilde{D}_1 \cap (\tilde{D}_2 \cup \tilde{D}_3) = \emptyset \implies 10 - 5\alpha > 2 + 1/\alpha.$$

- One valid choice: $\alpha = 1/7$
- This yields $\lambda_1 \in \tilde{D}_1(\alpha = 1/7) = B((10,0), 5/7)$
- and tells us that $\lambda_1 \in [10 - 5/7, 10 + 5/7]$

Example

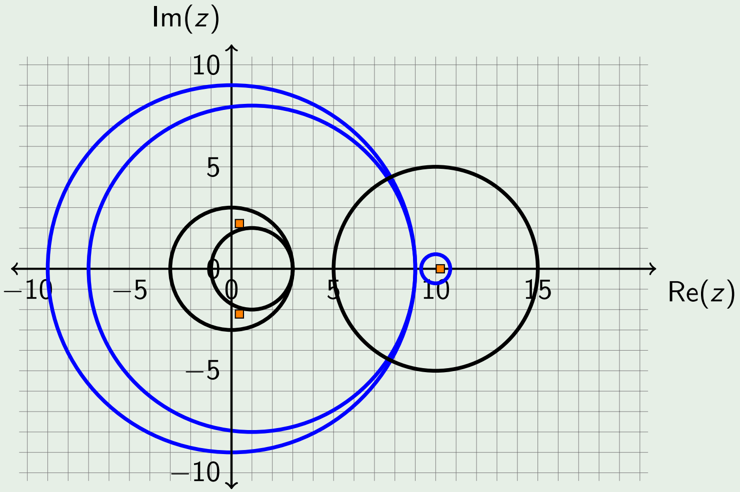


Figure: Gershgorin discs $\tilde{D}_1, \tilde{D}_2$ and $\tilde{D}_3$ of $\tilde{A}$ for $\alpha = 1/7$ in blue, and black Gershgorin discs for D_1, D_2 and D_3 for $\tilde{A} = A$ with $\alpha = 1$.

Power iteration I

Is an algorithm that computes the dominating (largest in absolute value) eigenvalue of a matrix.

Algorithm 1: Power iteration

Data: $A \in \mathbb{R}^{n \times n}$

Choose a start vector $x^{(0)} = x_0 \in \mathbb{R}^n \setminus \{0\}$.

for $k = 1, 2, \dots$ **do**

$$x^{(k)} \leftarrow Ax^{(k-1)} \tag{1}$$

Compute the normalized vector

$$z^{(k)} \leftarrow \frac{x^{(k)}}{\|x^{(k)}\|_2}$$

and the so-called Rayleigh quotient

$$\lambda^{(k)} \leftarrow (z^{(k)})^T A z^{(k)}.$$

end

Power iteration II

Remark

Remarks:

a) *Let* λ_1 *denote the dominating eigenvalue. Then under some assumptions,*

$$\lim_{k\to\infty} \lambda^{(k)} = \lambda_1$$

and $z^{(k)}$ *will asymptotically belong to eigenspace of* λ_1.

b) *Normally, one replaces the step* (1) *by*

$$z^{(k+1)} = \frac{Az^{(k)}}{\|Az^{(k)}\|_2},$$

to avoid the need for storing the $(x^{(k)})$ *sequence.*

Power iteration III

Example

Consider

$$A = \begin{bmatrix} 7/2 & 5 \\ 5/2 & 1 \end{bmatrix} \quad \text{with} \quad \sigma(A) = \{6, -3/2\} \quad \text{and } v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\text{Start vector} \quad x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{yields} \quad x^{(1)} = Ax^{(0)} = \frac{1}{2} \begin{bmatrix} 7 \\ 5 \end{bmatrix},$$

$$x^{(2)} = Ax^{(1)} = \frac{1}{4} \begin{bmatrix} 99 \\ 45 \end{bmatrix}, \qquad x^{(3)} = \frac{1}{8} \begin{bmatrix} 1143 \\ 585 \end{bmatrix} \dots$$

and

$$\lambda^{(1)} = (z^{(1)})^T A z^{(1)} = 6.2027, \quad \lambda^{(2)} = 5.8973, \dots, \quad \lambda^{(8)} = 6.0001$$

Limits: $\lambda^{(k)} \to \lambda_1$ and $z^{(k)} \to v_1$ as $k \to \infty$.

Theorem (Convergence of the power iteration)

Let $A \in \mathbb{R}^{n\times n}$ *be symmetric with real-valued eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ *such that*

$$\lambda_1 = \lambda_2 = \ldots = \lambda_r, \qquad \text{and} \qquad |\lambda_r| > |\lambda_{r+1}| > \ldots > |\lambda_n| \qquad \text{for some} \quad r$$

and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$*. Let further*

$$x^{(0)} = \sum_{j=1}^{n} \alpha_j v_j \qquad \text{with} \sum_{j=1}^{r} |\alpha_j| \neq 0.$$

Then, the normalized power iteration sequence $z^{(k)} := x^{(k)}/\|x^{(k)}\|$ *satisfies*

$$Az^{(k)} = \lambda_1 z^{(k)} + \mathcal{O}(q^k) \qquad \text{where} \quad q := \frac{|\lambda_{r+1}|}{|\lambda_1|},$$

and

$$\lambda^{(k)} = \left(z^{(k)}\right)^T Az^{(k)} = \lambda_1 + \mathcal{O}(q^{2k}).$$

Core proof idea

If $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$ and

$$x^{(0)} = \sum_{j=1}^{n} \alpha_j v_j, \quad \text{with} \quad \alpha_1 > 0,$$

then for $k \gg 1$

$$A^k x^{(0)} = \lambda_1^k \sum_{j=1}^{n} \alpha_j (\lambda_j/\lambda_1)^k v_j = \lambda_1^k \alpha_1 v_1 + \sum_{j=2}^{n} \alpha_j \underbrace{(\lambda_j/\lambda_1)^k}_{\mathcal{O}(q^k)} v_j$$

and one can show that

$$z^{(k)} = \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|_2} = \frac{\lambda_1^k v_1}{|\lambda_1|^k |\alpha_1|} + \mathcal{O}(q^k) = \pm v_1 + \mathcal{O}(q^k)$$

leading also to $\lambda^{(k)} = z^{(k)} A z^{(k)} = \lambda_1 + \mathcal{O}(q^{2k})$.

Remark

a) *The smaller the ratio* $q = \frac{|\lambda_{r+1}|}{|\lambda_1|}$*, the faster the convergence.*

b) *The assumption* $\sum_{j=1}^{r} |\alpha_j| \neq 0$ *for* $x^{(0)}$ *is difficult to verify, as one typically do not know the eigenvectors of A.*
Good strategy: *draw* $x^{(0)} \in \mathbb{R}^n_*$ *randomly.*

c) *The normalized vector*

$$z^{(k)} = \text{sgn}(\lambda_1^k) \frac{\tilde{v}_1}{\|\tilde{v}_1\|} + \mathcal{O}(q^k),$$

will converge as $k \to \infty$ *iff* $\lambda_1 \geq 0$. *Otherwise, it may for instance oscillate like* $z^{(k)} \approx (-1)^k v_1$.

Example (Sensitivity with respect to the start vector)

Consider the matrix in the previous example. How sensitive is the power iteration to the start vector $x^{(0)}$?
Experiment: draw the components in $x^{(0)}$ as independent $U[0,1]$ random variables, and compute

$$z^{(10)} = x^{(10)}/\|x^{(10)}\|_2$$

We repeat the experiment 5 times. Matlab code:

```
x0 = rand(2,5) %5 columns with x0 vectors
x10 = A^(10) * x0;
z =zeros(2,5);
for i = 1:5
     z(:,i) = x10(:,i)/norm(x10(:,i));
end
```

Example (Output runs)

This yields the output

```
x0 =

    0.7513    0.5060    0.8909    0.5472    0.1493
    0.2551    0.6991    0.9593    0.1386    0.2575



z =

    0.8944    0.8944    0.8944    0.8944    0.8944
    0.4472    0.4472    0.4472    0.4472    0.4472
```

Definition (Rayleigh quotient)

For a symmetric $A \in \mathbb{R}^{n \times n}$, the Rayleigh quotient for a nonzero vector $x \in \mathbb{R}^n$ is defined by

$$R(x) := \frac{x^T A x}{x^T x}$$

Theorem

Let $A \in \mathbb{R}^{n \times n}_{sym}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Then it holds that

$$\lambda_1 = \sup_{x \in \mathbb{R}^n_*} R(x) \qquad \lambda_n = \inf_{x \in \mathbb{R}^n_*} R(x)$$

Proof.

Let $\{(\lambda_i, v_i)\}_{i=1}^n$ denote the eigenpairs of A, with $v_1, \ldots, v_n$ orthonormal. Any $x \in \mathbb{R}^n$ can be written

$$x = \sum_{j=1}^{n} \alpha_j v_j$$

and

$$R(x) = \frac{x^T A x}{x^T x} = \frac{\sum_{j,k} \alpha_j \alpha_k v_k^T A v_j}{\sum_{j,k} \alpha_j \alpha_k v_k^T v_j} = \frac{\sum_{j=1}^n \alpha_j^2 \lambda_j}{\sum_{j,k} \alpha_j^2} \leq \frac{\sum_{j=1}^n \alpha_j^2 \lambda_1}{\sum_j \alpha_j^2} = \lambda_1.$$

By choosing $x = v_1$, we obtain $R(x) = \lambda_1$.
The lower bound $R(x) \geq \lambda_n$ and $R(v_n) = \lambda_n$ is proved similarly. □

Exercise

Show that

a) For any eigenvector v_j, it holds that

$$R(v_j) = \lambda_j.$$

b) For the approximation of an eigenvector $\tilde{v} = v_j + \Delta v$ where $\Delta v = \sum_{i=1}^{n} \varepsilon_i v_i$, it holds that (exercise)

$$|R(\tilde{v}) - \lambda_j| \leq 2(n-1)\|A\|_2 \|\Delta v\|_2^2.$$

Inverse iteration I

If A has eigenvalues $\{\lambda_i\}_{i=1}^n$ with

$$|\lambda_n| < |\lambda_{n-1}| \leq \ldots \leq \lambda_1,$$

then A^{-1} has eigenvalues $\{\lambda_i^{-1}\}_{i=1}^n$ with

$$|\lambda_n^{-1}| > |\lambda_{n-1}^{-1}| > \ldots$$

Motivation:

$$Av_k = \lambda_k v_k \implies \lambda_k^{-1} v_k = A^{-1} v_k$$

Inverse iteration II

Approximate smallest eigenpair (λ_n, v_n) of A, by applying inverse power iteration

$$Ax^{(k+1)} = x^{(k)} \qquad k = 0, 1, \dots$$

Note: this is power iteration for A^{-1} with largest eigenpair (λ_n^{-1}, v_n).

Know from convergence of power iteration that $x^{(k)}/\|x^{(k)}\|_2 \approx \pm v_n$ for large k, and therfore also that

$$\lambda^{(k)} = R(x^{(k)}) \to \lambda_n \qquad \text{as } k \to \infty.$$

Example

Consider again

$$A = \begin{bmatrix} 7/2 & 5 \\ 5/2 & 1 \end{bmatrix} \quad \text{with} \quad \sigma(A) = \{6, -3/2\} \quad \text{and } v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Inverse iterations applied to the start vector $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ yields

```
x = [1; 0]; % startvector
z = zeros(2,7);
for i = 1:7
    x = A\x;
    z(:,i) = x/norm(x);
end
```

```
z =
   -0.3714    0.7682   -0.6902    0.7112   -0.7061    0.7074   -0.7070
    0.9285   -0.6402    0.7236   -0.7030    0.7081   -0.7068    0.7072
```

Inverse iteration with shift

Inverse iteration (and also power iteration) may also compute the eigenvalue of A closest to a $\mu \in \mathbb{R}$ through:

1. Apply inverse iterations to the shifted matrix $(A - \mu I)$:

$$(A - \mu I)x^{(k+1)} = x^{(k)} \qquad \text{for } k = 0, 1, \ldots$$

2. The dominant eigenvalue of $(A - \mu I)^{-1}$ equals

$$\hat{\lambda}^{-1} = (\lambda_i - \mu)^{-1},$$

where $\lambda_i = \arg\min_{\lambda \in \sigma(A)} |\lambda - \mu| = \lambda_{\text{closest to } \mu}$.

By same argument as before

$$\frac{x^{(k)}}{\|x^{(k)}\|_2} \approx \pm v_i \quad \text{for} \quad k \gg 1, \quad \text{and} \quad R(x^{(k)}) \to \lambda_i$$

Summary

- Gershgorin's theorem may be used in combination with similarity transformations to improve eigenvalue estimates.
- Power iteration and inverse iteration are methods for computing eigenpairs of matrices, respectively for largest and smallest eigenvalue, in absolute value.
- Shifting combined with power/inverse iteration makes it possible to compute eigenpairs for eigenvalue closest to shift.