Imperial College London

Maths for Signals and Systems

Session 2 (Week 3),

Pier Luigi Dragotti
EEE Department
Imperial College London
p.dragotti@imperial.ac.uk

Outline

The main goal of this set of lectures is to review fundamental notions in matrix theory and to explore their use in various contexts

- Determinant
- Rank and Trace of a matrix
- Eigenvectors and eigenvalues
- Matrix diagonalization
- Markov processes and circulant matrices
- Eigenvalues and eigenvectors of Symmetric and Hermitian matrices
- Positive Definite Matrices
- Graphs and the power method

Determinants

- The **Determinant** is a crucial number associated with square matrices.
- We denote it by $\det(\boldsymbol{A}) \ \ or \ \ |\boldsymbol{A}|$.
- The determinant can be used to determine invertibility. A matrix $\boldsymbol{A}$ is invertible if and only if $\det(\boldsymbol{A}) \neq 0$.
- An invertible matrix is called **non-singular**.
- For a $2{\times}2$ matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ the determinant is defined as $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

Determinants – Laplace's Formula

- The determinant of a $n \times n$ matrix is given by:

  - $det(\boldsymbol{A}) = \sum_{j=1}^{n}(-1)^{i+j}\, a_{ij}M_{ij}$ (for a fixed i) or
  - $\det(\boldsymbol{A}) = \sum_{i=1}^{n}(-1)^{i+j} a_{i,j}M_{ij}$ (for a fixed j)

- The '***minor***' $\boldsymbol{M}_{i,j}$ is defined to be the determinant of the $(n-1) \times (n-1)$-matrix that results from $\boldsymbol{A}$ by removing the i-th row and the j-th column.

- The '***cofactor***' $\boldsymbol{C}_{ij}$ is obtained by multiplying the minor by $(-1)^{(i+j)}$

Properties of determinants

1. $\det(I) = 1$. This is easy to show in the case of a 2×2 matrix using the formula of the previous slide.

2. If we exchange two rows of a matrix the sign of the determinant reverses. Therefore:
   - If we perform an even number of row exchanges the determinant remains the same.
   - If we perform an odd number of row exchanges the determinant changes sign.
   - Hence, the determinant of a **permutation** matrix is 1 or -1.

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \text{ and } \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1 \text{ as expected.}$$

Properties of determinants cont.

3a. If a row is multiplied with a scalar, the determinant is multiplied with that scalar too, i.e., $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

3b. A row of zeros leads to $\det = 0$. This can verified as follows for any matrix:

$$\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = \begin{vmatrix} 0 \cdot a & 0 \cdot b \\ c & d \end{vmatrix} = 0 \begin{vmatrix} a & b \\ c & d \end{vmatrix} = 0$$

3c. $\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

Note that $\det(A + B) \neq \det(A) + \det(B)$

We observe "linearity" only for a single row.

4. Two equal rows leads to $\det = 0$.
   - As mentioned, if we exchange rows the sign of the determinant changes.
   - In that case the matrix is the same and therefore, the determinant should remain the same. Therefore, the determinant must be zero.

Properties of determinants cont.

5. $$\begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ -la & -lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} - l\begin{vmatrix} a & b \\ a & b \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Therefore, the determinant after row reduction remains the same.

6. Consider an upper triangular matrix ($*$ is a random element)

$$\begin{vmatrix} d_1 & * & \dots & * \\ 0 & d_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{vmatrix} = d_1 d_2 \dots d_n$$

Properties of determinants cont.

7. $\det(\boldsymbol{A}) = 0$ when $\boldsymbol{A}$ is singular. This is because if $\boldsymbol{A}$ is singular we get a row of zeros by elimination. Using the same concept we can say that if $\boldsymbol{A}$ is invertible then $\det(\boldsymbol{A}) \neq 0$.

8. When both $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices then: $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$

9. *Moreover:*

$$\det(\boldsymbol{A}^{-1}) = \frac{1}{\det(\boldsymbol{A})}$$

$$\det(\boldsymbol{A}^2) = [\det(\boldsymbol{A})]^2$$

$\det(c\boldsymbol{A}) = c^n\det(\boldsymbol{A})$ where $\boldsymbol{A}: n \times n$ and c is a scalar.

10. $\det(\boldsymbol{A}^T) = \det(\boldsymbol{A})$.

Estimation of the inverse A^{-1} using cofactors

- For a 2×2 matrix it is quite easy to show that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- The general formula for the inverse A^{-1} is given by:

$$A^{-1} = \frac{1}{\det(A)} C^T$$

- *Where* C_{ij} is the cofactor of a_{ij}.
- In general

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} = \det(A)\cdot I$$

Imperial College London

The Polynomial Interpolation Problem

- Assume you want to find a polynomial of degree n which satisfies:

$$P(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_n x_i^n = y_i$$

for data points $(y_0, x_0), (y_1, x_1), \dots, (y_n, x_n)$.

- In matrix/vector form:

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

- Has this problem a solution?

Imperial College London

The Polynomial Interpolation Problem (cont'd)

- The matrix

$$\boldsymbol{V} = \begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^n \end{bmatrix}$$

is called *Vandermonde matrix* and satisfies this remarkable property:

$$\det(\mathrm{V}) = \prod_{0 \le j < i \le n} (x_i - x_j)$$

- So the problem has always a unique solution if the points are distinct

Rank of a Matrix and Trace of a Matrix

- **Rank of a matrix**: Given an $m \times n$ matrix $\boldsymbol{A}$, the rank of $\boldsymbol{A}$, denoted $rank(\boldsymbol{A})$, is the number of linearly independent rows or columns; in particular, if $m = n$, $rank(\boldsymbol{A}) = n$ if and only if $\det(\boldsymbol{A}) \neq 0$
  - **Properties** (without proof):
    - For a rectangular matrix $rank(\boldsymbol{A}) \leq \min(n, m)$
    - $rank(\boldsymbol{AB}) \leq \min(rank(\boldsymbol{A}), rank(\boldsymbol{B}))$

- **Trace** of a $n \times n$ matrix is defined as the sum of the elements along the main diagonal. Moreover
  - $trace(\boldsymbol{ABC}) = trace(\boldsymbol{BCA}) = trace(\boldsymbol{CAB})$
  - (Frobenius norm): $||\boldsymbol{A}|| = \sqrt{\sum |a_{i,j}|^2} = [trace(A^H A)]^{\frac{1}{2}}$

Eigenvectors and eigenvalues

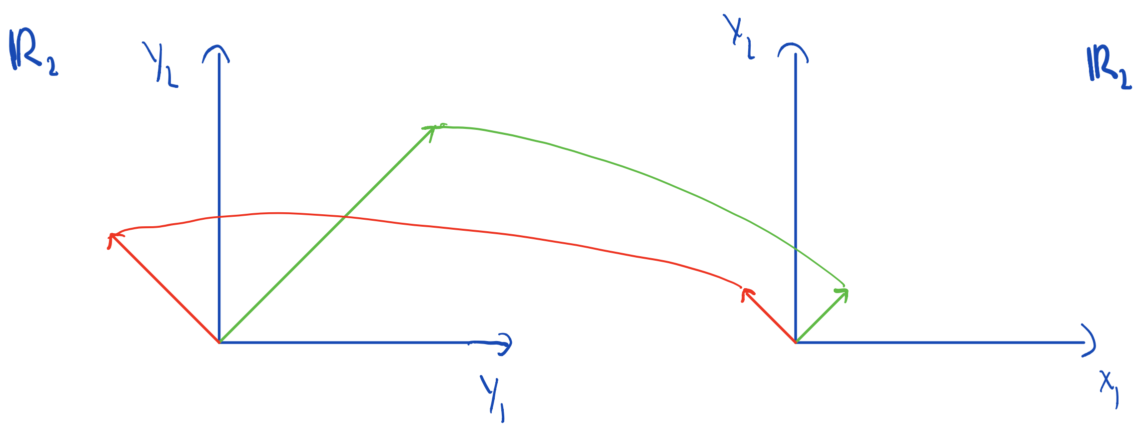
- Consider a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{x}$.
- The operation $\boldsymbol{Ax}$ produces a vector $\boldsymbol{y}$ at some direction.
- We are interested in vectors $\boldsymbol{y}$ which lie in the same direction as $\boldsymbol{x}$.
- In that case we have $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ with λ being a scalar.
- When the above relationship holds, $\boldsymbol{x}$ is called an **eigenvector** and λ is called an **eigenvalue** of matrix $\boldsymbol{A}$.
- If $\boldsymbol{A}$ is singular then $\lambda = 0$ is an eigenvalue.
- **Problem:** How do we find the eigenvectors and eigenvalues of a matrix?

Eigenvectors and eigenvalues: motivation

- Why are we interested in the the eigenvectors and eigenvalues of a matrix? Because they make the solution of complex problems simpler.
- For example if we know the eigenvectors and eigenvalues of a linear mapping $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ we can characterize that mapping easily

Eigenvectors and eigenvalues: motivation

- For example if we know the eigenvectors and eigenvalues of a linear mapping $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ we can characterize that mapping easily
- For example, $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, eigenvalues: $\lambda_1 = 4,\ \lambda_2 = 2$, eigenvectors: $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\boldsymbol{u}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$
- Then $\boldsymbol{x} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2$ implies $\boldsymbol{y} = a_1\lambda_1\boldsymbol{u}_1 + a_2\lambda_2\boldsymbol{u}_2$



Eigenvectors and eigenvalues: motivation

- Why are we interested in the the eigenvectors and eigenvalues of a matrix? Because they make the solution of complex problems simpler.
- In Signals and Systems we have already seen a few examples of this (in the continuous-time case), e.g.,
    - the eigen-functions of a LTI systems leading to the convolution formula in the Fourier domain
    - the solutions of linear differential equations
- In the discrete-time finite dimensional case, eigenvalues and eigenvectors are very important for discrete-time convolution and for difference equations

Eigenvectors and eigenvalues of a permutation matrix

- Consider the permutation matrix $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

  Problem: Can you give an eigenvector of the above matrix? Or can you think of a vector that if permuted is still a multiple of itself?

  Answer: YES. It is the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and the corresponding eigenvalue is $\lambda = 1$.

  And furthermore, the vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ with eigenvalue $\lambda = -1$.

- $n \times n$ matrices will have n eigenvalues.
- It is not easy to find them.
- **The sum of the eigenvalues, called the trace of a matrix, equals the sum of the diagonal elements of the matrix.**
- **The product of the eigenvalues equals the determinant of the matrix.**
- Therefore, in the previous example, once we found an eigenvalue $\lambda = 1,$ we should know that there is another eigenvalue $\lambda = -1$.

Problem: Solve $Ax = \lambda x$

- Consider an eigenvector $\boldsymbol{x}$ of matrix $\boldsymbol{A}$. In that case $\boldsymbol{Ax} = \lambda\boldsymbol{x} \Rightarrow \boldsymbol{Ax} - \lambda\boldsymbol{x} = \boldsymbol{0}$ ($\boldsymbol{0}$ is the zero vector). Therefore, $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$.

  In order for the above set of equations to have a non-zero solution, the nullspace of $(\mathbf{A} - \lambda\mathbf{I})$ must be non-zero, i.e., the matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$ must be singular. Therefore, $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$.

- We now have an equation for λ. It is called the **characteristic equation**, or the **eigenvalue equation**. From the roots of this equation we can find the eigenvalues.

- We might have repeated λs. **This might cause problems but we will not deal with this issue.**

- After we find λ, we can find $\boldsymbol{x}$ from $(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{x} = \boldsymbol{0}$. Basically, we will be looking for the nullspace of $(\boldsymbol{A} - \lambda\boldsymbol{I})$.
- The eigenvalues of $\boldsymbol{A}^T$ are obtained through the equation $\det(\boldsymbol{A}^T - \lambda\boldsymbol{I}) = 0$. But: $\det(\boldsymbol{A}^T - \lambda\boldsymbol{I}) = \det(\boldsymbol{A}^T - \lambda\boldsymbol{I}^T) = \det[(\boldsymbol{A} - \lambda\boldsymbol{I})^T] = \det(\boldsymbol{A} - \lambda\boldsymbol{I})$.
- Therefore, the eigenvalues of $\boldsymbol{A}^T$ are the same as the eigenvalues of $\boldsymbol{A}$.

Solve $Ax = \lambda x$. An example.

- Consider the matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.
- $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (3-\lambda)^2 - 1 = 0 \Rightarrow 3 - \lambda = \pm 1 \Rightarrow \lambda = 3 \pm 1 \Rightarrow \lambda_1 = 4,\ \lambda_2 = 2.$
  Or $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda$^$2 - 6\lambda + 8 = 0$. Note that $6 = \lambda_1 + \lambda_2$ and $8 = \det(A) = \lambda_1 \lambda_2$.
- Find the eigenvector for $\lambda_1 = 4$.
  $\mathbf{A} - 4\boldsymbol{I} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x = y$
- Find the eigenvector for $\lambda_2 = 2$.
  $\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x = -y$
- Notice that there are **families of eigenvectors**, not single eigenvectors.

Compare the two matrices given previously

- Consider the matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. As shown it has eigenvectors $\begin{bmatrix} x \\ x \end{bmatrix}$ and $\begin{bmatrix} -x \\ x \end{bmatrix}$ with eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 2$.

- Consider the matrix $\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, also with eigenvectors $\begin{bmatrix} x \\ x \end{bmatrix}$ and $\begin{bmatrix} -x \\ x \end{bmatrix}$ and eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$.

- **We observe that $A = B + 3I$. The eigenvalues of A are obtained from the eigenvalues of B if we increase them by 3.**

- **The eigenvectors of A and B are the same.**

Generalization of the above observation

- Consider the matrix $\boldsymbol{A} = \boldsymbol{B} + c\boldsymbol{I}$.
- Consider an eigenvector $\boldsymbol{x}$ of $\boldsymbol{B}$ with eigenvalue λ. Then $\boldsymbol{B}\boldsymbol{x} = \lambda\boldsymbol{x}$ and therefore,
  $\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{B} + c\boldsymbol{I})\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + c\boldsymbol{I}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + c\boldsymbol{x} = \lambda\boldsymbol{x} + c\boldsymbol{x} = (\lambda + c)\boldsymbol{x}$
  $\boldsymbol{A}$ has the same eigenvectors with $\boldsymbol{B}$ with eigenvalues $\lambda + c$.
- **This is one of the few cases where we can find the eigenvalues of $\boldsymbol{A} + \boldsymbol{B}$ directly.**

Example

- Consider $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.
- $\lambda_1 + \lambda_2 = 6$ and $\det(\lambda_1 \lambda_2) = 9$.
- $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det \begin{bmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} = (3-\lambda)^2 = 0 \Rightarrow \lambda_{1,2} = 3$
- **The eigenvalues of a triangular matrix are the values of the diagonal.**
- For that particular case we have

  $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x \\ 3y \end{bmatrix} \Rightarrow y = 0$ and x can be any number.

Imperial College London

Matrix diagonalization
The case of independent eigenvectors

- Suppose we have n independent eigenvectors of a matrix $\boldsymbol{A}$. We call them $\boldsymbol{x_i}$.
- The associated eigenvalues are λ_i.
- We put them in the columns of a matrix $\boldsymbol{S}$.
- We form the matrix:

$$\boldsymbol{AS} = \boldsymbol{A}[\boldsymbol{x_1} \quad \boldsymbol{x_2} \quad \cdots \quad \boldsymbol{x_n}] = [\lambda_1\boldsymbol{x}_1 \quad \lambda_2\boldsymbol{x}_2 \quad \dots \quad \lambda_n\boldsymbol{x}_n] =$$

$$[\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_n]\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = \boldsymbol{S\Lambda} \Rightarrow \boldsymbol{AS} = \boldsymbol{S\Lambda}$$

$\boldsymbol{S}^{-1}\boldsymbol{AS} = \boldsymbol{\Lambda}$ or $\boldsymbol{A} = \boldsymbol{S\Lambda S}^{-1}$

- The above formulation of $\boldsymbol{A}$ is very important in Mathematics and Engineering and is called **matrix diagonalization**.

Imperial College London

Matrix diagonalization

- Theorem: If $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonalizable they share the same eigenvector matrix $\boldsymbol{S}$ if and only if $\boldsymbol{AB} = \boldsymbol{BA}$

- *Sketch of the proof*: $\boldsymbol{AB} = \boldsymbol{S\Lambda_1 S^{-1} S\Lambda_2 S^{-1}}$ and $\boldsymbol{BA} = \boldsymbol{S\Lambda_2 S^{-1} S\Lambda_1 S^{-1}}$, since $\boldsymbol{\Lambda_1 \Lambda_2} = \boldsymbol{\Lambda_2}\boldsymbol{\Lambda}_1$ it follows that $\boldsymbol{AB} = \boldsymbol{BA}$. The converse is omitted.

Matrix diagonalization: Eigenvalues of A^k

- If $\boldsymbol{Ax} = \lambda \boldsymbol{x} \Rightarrow \boldsymbol{A}^2\boldsymbol{x} = \lambda \boldsymbol{Ax} \Rightarrow \boldsymbol{A}^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$.
- Therefore, the eigenvalues of $\boldsymbol{A}^2$ are λ^2.
- The eigenvectors of $\boldsymbol{A}^2$ remain the same since $\boldsymbol{A}^2 = \boldsymbol{S\Lambda S}^{-1}\boldsymbol{S\Lambda S}^{-1} = \boldsymbol{S\Lambda}^2\boldsymbol{S}^{-1}$
- In general $\mathbf{A}^k = \boldsymbol{S\Lambda}^k\boldsymbol{S}^{-1}$
- A matrix has n independent eigenvectors and therefore is diagonalizable if all the eigenvalues are different.
- If there are repeated eigenvalues a matrix may, or may not have independent eigenvectors. As an example consider the identity matrix. Its eigenvectors are the row (or column) vectors. They are all independent. However, the eigenvalues are all equal to 1.

- Assume $\boldsymbol{A} = \boldsymbol{S\Lambda S}^{-1}$ is square and invertible, can you compute $\boldsymbol{A}^{-1}$?

Difference Equations and the powers of A^k

- Consider the Fibonacci sequence: 0,1,1,2,3,5,8,13,…
- It is described by the difference equation $f_{k+2} = f_{k+1} + f_k$
- By setting $\boldsymbol{u}_k = \begin{bmatrix} f_{k+1} \\ f_k \end{bmatrix}$ and using $f_{k+1} = f_{k+1}$ we have $\boldsymbol{u}_{k+1} = \boldsymbol{A}\boldsymbol{u}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k$
- If we know $\boldsymbol{u}_0$ then $\boldsymbol{u}_k = \boldsymbol{A}^k \boldsymbol{u}_0 = \boldsymbol{S}\boldsymbol{\Lambda}^k \boldsymbol{S}^{-1} \boldsymbol{u}_0$
- Therefore, the knowledge of the eigenvectors and eigenvalues of $\boldsymbol{A}$ makes the solution of the equation easier.

Difference Equations and the powers of A^k

- Consider the system of difference equations
- $$\begin{aligned} y_1[n+1] &= \quad y_1[n] \quad -1.5y_2[n] \\ y_2[n+1] &= 0.5y_1[n] \quad + \quad y_2[n] \end{aligned}$$

- By setting $\boldsymbol{u}_n = \begin{bmatrix} y_1[n] \\ y_2[n] \end{bmatrix}$ we have $\boldsymbol{u}_{n+1} = \boldsymbol{A}\boldsymbol{u}_n = \begin{bmatrix} 1 & -1.5 \\ 0.5 & 1 \end{bmatrix} \boldsymbol{u}_n$
- If we know $\boldsymbol{u}_0$ then $\boldsymbol{u}_n = \boldsymbol{A}^n \boldsymbol{u}_0 = \boldsymbol{S}\boldsymbol{\Lambda}^n \boldsymbol{S}^{-1} \boldsymbol{u}_0$
- Therefore, also in this case the knowledge of the eigenvectors and eigenvalues of $\boldsymbol{A}$ makes the solution of the system easier.

Markov Processes

- This example is taken from Strang's book
- *"Assume that each year 1/10 of the people outside California move in and 2/10 of the people inside California move out"*

- Assuming y_0 is the initial population outside and z_0 is the initial population inside we have

- $\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix}$

- If we know the eigenvectors and eigenvalues of $\boldsymbol{A}$ then it is easy to calculate the population at instant k

- We note that in this case $\lambda_1 = 1$ and $\lambda_2 = 0.7$, moreover $\boldsymbol{u}_1 = \left(\frac{2}{3}, \frac{1}{3}\right)^T, \boldsymbol{u}_2 = \left(\frac{1}{3}, -\frac{1}{3}\right)$.
- For large k the contribution of the second eigenvalue vanishes and so the steady-state solution is:

- $\begin{bmatrix} y_\infty \\ z_\infty \end{bmatrix} = (y_0 + z_0) \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$

Imperial College London

Markov Processes (cont'd)

- By using $\boldsymbol{A}^k = \boldsymbol{S}\boldsymbol{\Lambda}^k\boldsymbol{S}^{-1}$ we find that the population at instant k is

- $\begin{bmatrix} y_k \\ z_k \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (0.7)^k \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} = (y_0 + z_0) \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} + (y_0 - 2z_0)(0.7)^k \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$

- For large k the contribution of the second eigenvalue vanishes and so the steady-state solution is:

- $\begin{bmatrix} y_\infty \\ z_\infty \end{bmatrix} = (y_0 + z_0) \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$

Markov Processes General Properties

- From this example we can derive some general properties of Markov or “transition” matrices:

    - All entries are non-negative
    - Each column of the matrix adds up to one
    - $\lambda_1 = 1$ is an eigenvalue, its eigenvector $\boldsymbol{u}_1$ is a steady state
    - The other eigenvalues satisfy $|\lambda_i| \leq 1$

- That $\lambda_1 = 1$ is always an eigenvalue can be easily shown by observing that each column of $\boldsymbol{A} - \boldsymbol{I}$ adds up to 0 and therefore $\boldsymbol{A} - \boldsymbol{I}$ is singular and $\lambda_1 = 1$ is an eigenvalue.

- The corresponding eigenvector is a steady state since $\boldsymbol{A}\boldsymbol{u}_1 = \boldsymbol{u}_1$

Diagonalization of Circulant Matrices

- Recall that the circular convolution between $\boldsymbol{y}$ and $\boldsymbol{x}$ using a filter $\boldsymbol{h}$ can be modelled as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is circulant.

- The Discrete Fourier Transform (DFT) matrix of size $n \times n$ has entry (i, k) given by $W_n^{(i-1)(k-1)} = e^{-j2\pi(i-1)(k-1)/n}$. We previously denoted it as $\boldsymbol{F}^H$

- The eigenvectors of a circulant matrix are the columns of $\boldsymbol{F}$:

  $u_i = \frac{1}{\sqrt{n}}(1 \quad \omega_i \quad \omega_i^2 \quad \cdots \quad \omega_i^{n-1})^H$ with $\omega_i = e^{-j2\pi(i-1)/n}$

- The eigenvalues are the DFT of $\boldsymbol{h}$

- So the DFT diagonalizes circulant matrices

Symmetric matrices

- In case of real matrices, symmetry is defined as $\boldsymbol{A} = \boldsymbol{A}^T$.
- In case of complex matrices, symmetry is defined as $\boldsymbol{A}^{*T} = \boldsymbol{A}$. A matrix which possesses this property is called **Hermitian**.
- Remember that we use the symbol $\boldsymbol{A}^H = {\boldsymbol{A}^*}^T$.
- Symmetric matrices appear in many contexts in particular in statistics (e.g., covariance matrices)
- We will prove that
  - the eigenvalues of a symmetric matrix are real.
  - The eigenvectors of a symmetric matrix that correspond to different eigenvalues **are** orthogonal.

Real symmetric matrices

Problem:

Prove that the eigenvalues of a symmetric matrix occur in complex conjugate pairs.

Solution:

Consider $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$.

If we take complex conjugate on both sides we get

$$(\mathbf{Ax})^{*} = (\lambda\mathbf{x})^{*} \Rightarrow \boldsymbol{A}^{*}\ \boldsymbol{x}^{*} = \lambda^{*}\ \mathbf{x}^{*}$$

If $\boldsymbol{A}$ is real then $\boldsymbol{A}\boldsymbol{x}^{*} = \lambda^{*}\ \boldsymbol{x}^{*}$. Therefore, if λ is an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector $\boldsymbol{x}$ then λ^{*} is an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector x^{*}.

Real symmetric matrices cont.

Problem:

Prove that the eigenvalues of a symmetric matrix are real.

Solution:

We proved that if $\boldsymbol{A}$ is real then $Ax^* = \lambda^* \, x^*$.

If we take transpose in both sides we get

$$x^{*T} A^T = \lambda^* x^{*T} \Rightarrow x^{*T} A = \lambda^* x^{*T}$$

We now multiply both sides from the right with x and we get

$$x^{*T} A x = \lambda^* x^{*T} x$$

We take now $Ax = \lambda x$. We now multiply both sides from the left with x^{*T} and we get

$$x^{*T} A x = \lambda x^{*T} x.$$

From the above we see that $\lambda x^{*T} x = \lambda^* x^{*T} x$ and since $x^{*T} x \neq 0$, we see that $\lambda = \lambda^*$.

Real symmetric matrices cont.

Problem:

Prove that the eigenvectors of a symmetric matrix **which correspond to different eigenvalues** are always orthogonal.

Solution:

Suppose that $Ax = \lambda_1 x$ and $Ay = \lambda_2 y$ with $\lambda_1 \neq \lambda_2$.

$$(\lambda_1 x)^T y = x^T \lambda_1 y = (Ax)^T y = x^T A y = x^T \lambda_2 y$$

The conditions $x^T \lambda_1 y = x^T \lambda_2 y$ and $\lambda_1 \neq \lambda_2$ give $x^T y = 0$.

The eigenvectors x and y are orthogonal.

Complex matrices. Complex symmetric matrices.

- Let us find which complex matrices have real eigenvalues and orthogonal eigenvectors.
- Consider $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ with $\boldsymbol{A}$ possibly complex.
- If we take complex conjugate in both sides we get
$$(Ax)^{*} = (\lambda x)^{*} \Rightarrow A^{*}\ x^{*} = \lambda^{*}\ x^{*}$$
- If we take transpose in both sides we get
$$x^{*T}A^{*T} = \lambda^{*}x^{*T}$$
- We now multiply both sides from the right with x we get
$$x^{*T}A^{*T}x = \lambda^{*}x^{*T}x$$
- We take now $Ax = \lambda x$. We now multiply both sides from the left with x^{*T} and we get
$$x^{*T}Ax = \lambda x^{*T}x.$$
- From the above we see that if $A^{*T} = A$ then $\lambda x^{*T}x = \lambda^{*}x^{*T}x$ and since $x^{*T}x \neq 0$, we see that $\lambda = \lambda^{*}$.

Complex symmetric matrices cont.

Problem:

Prove that the eigenvectors of a complex symmetric matrix (Hermitian matrix) **which correspond to different eigenvalues** are always perpendicular.

Solution:

Suppose that $Ax = \lambda_1 x$ and $Ay = \lambda_2 y$ with $\lambda_1 \neq \lambda_2$.

$$(\lambda_1 x)^H y = x^H \lambda_1 y = (Ax)^H y = x^H A y = x^H \lambda_2 y$$

The conditions $x^H \lambda_1 y = x^H \lambda_2 y$ and $\lambda_1 \neq \lambda_2$ give $x^H y = 0$.

The eigenvectors x and y are perpendicular.

Summary for Symmetric and Hermitian Matrices

- We proved that:
  - The eigenvalues of a symmetric matrix, either real or complex, are real.
  - The eigenvectors of a symmetric real or complex matrix that correspond to different eigenvalues are orthogonal.
- We conclude this part with the following statement (without proof):
  - (Spectral theorem): Every real symmetric matrix can be diagonalized by an orthogonal matrix, and every Hermitian matrix can be diagonalized by a unitary matrix:

(real case) $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\boldsymbol{T}}$ (complex case) $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\boldsymbol{H}}$

Positive Definite Hermitian Matrices

- A matrix $\boldsymbol{B}$ is *positive definite* when $\boldsymbol{x}^H\boldsymbol{B}\boldsymbol{x} > 0$ for any $\boldsymbol{x} \neq 0$. In the case of Hermitian matrices it is easy to test whether they are positive definite.

- **Statement without proof**: A Hermitian Matrix $\boldsymbol{A}$ is *positive definite* if and only if:

  - $\boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x} > 0$

  - All the eigenvalues of $\boldsymbol{A}$ satisfy $\lambda_i > 0$

  - All the pivots (without row exchanges) satisfy $d_i > 0$

Imperial College London

Positive Definite Hermitian Matrices

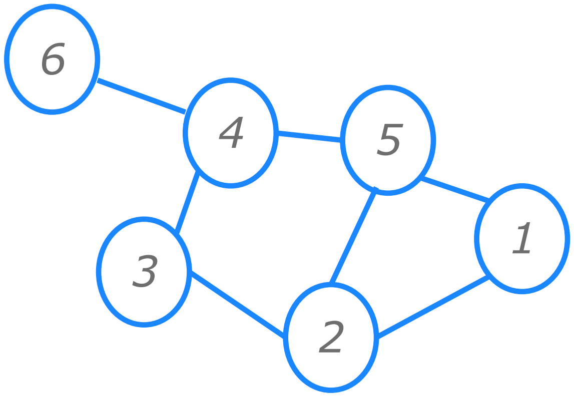
- Why is the positive definite test important?

- Consider the function $f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$
  - Clearly $f(0,0) = 0$ and at $(0,0)$ we also have $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = 0$
  - So is (0,0) a minimum, a maximum or a saddle point?
  - If $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$ is positive definite, then $(0,0)$ is a minimum, it is a maximum is $\boldsymbol{A}$ is negative definite and it is a saddle point otherwise.
- In general, if you have a function in many variables of the form $f(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j$ you can write it in compact form as $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ and then by testing whether $\boldsymbol{A}$ is positive definite you can verify if the point at $\boldsymbol{x} = 0$ is a global minimum.

Introducing graphs and their link to eigenvectors

- Any time we observe a set of objects which are related, these objects form a graph, e.g.,
  - A social network: people are nodes, friendships are links
  - Correlated variables: variables are nodes, correlations are links
  - Web-pages and hyperlinks

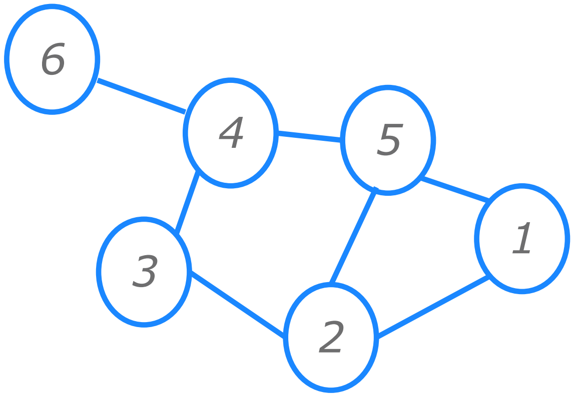
- More formally, a graph $G = (V, E)$ comprises:
  - A set V of nodes (vertices): $\{1,2,3,4,5,6\}$
  - A set E of pairs of nodes denoting
    edges $\{(1,2), (1,5), (2,5), (4,5), (2,3), (3,4), (4,6)\}$



Imperial College London

Directed, Weighted Graphs

- Graphs can be *directed* if edges have directions (like arrows), otherwise they are *undirected*
- Graphs are *weighted* if we associate each edge with a value (weight), otherwise they are *unweighted*



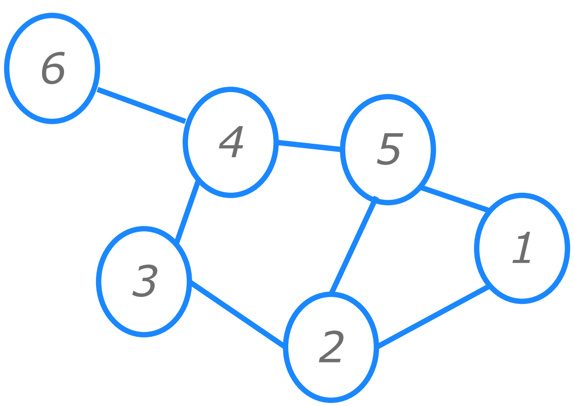
Adjacency Matrix and Degrees of a Graph

- The adjacency matrix $\boldsymbol{A}$ is defined as follows:

$$a_{i,j} = \begin{cases} w_{i,j} & if\ i\ is\ connected\ to\ j \\ 0 & otherwise \end{cases}$$

- In our example (unweighted and undirected graph) we have:

- $A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

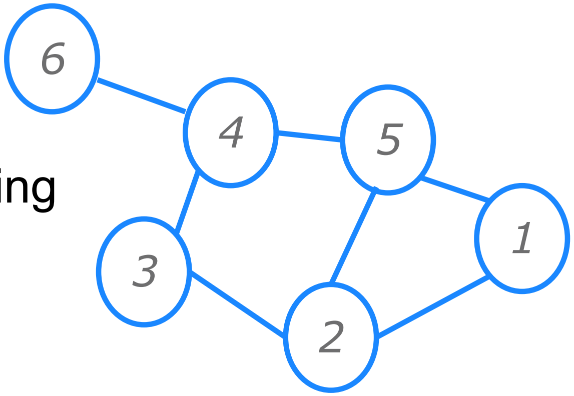


Adjacency Matrix and Degrees of a Graph

- The adjacency matrix $\boldsymbol{A}$ is defined as follows:

$$a_{i,j} = \begin{cases} w_{i,j} & if\ i\ is\ connected\ to\ j \\ 0 & otherwise \end{cases}$$

- The degree d_i of a node is the number of nodes to which it is adjacent (i.e., number of neighbours)
- E.g. the degree of node 4 is $d_4 = 3$
- For an undirected and unweighted graph with no self-loops the degree of node i is obtained by summing across the i-th column or row of $\boldsymbol{A}$
- This means that in this case the degree sequence is given by $\boldsymbol{A}\mathbf{1}$ with $\mathbf{1} = [1,1, \dots, 1]^T$



Node Centrality and PageRank Algorithm

- The degree d_i of a node gives us a first idea of the importance of that node in a network (e.g., a webpage is more important if many pages link to it)
- However, the best way to infer centrality of a node is that it should be proportional to the sum of the importances of its neighbours
- Assume the vector $\boldsymbol{u}$ is defined such that its entry u_i corresponds to the centrality (importance) of node i
- Then the sum of importances over node i's neighbours is the i-th entry of $\boldsymbol{Au}$ since it is given by $\sum_{j=1}^{n} a_{ij} u_j$
- We expect to reach convergence at a certain point which implies that we are trying to solve this eigenvector problem $\boldsymbol{Au} = \lambda\, \boldsymbol{u}$
- We also want all the entries of $\boldsymbol{u}$ to be non-negative which means that we are after the largest eigenvalue of $\boldsymbol{A}$ (proof omitted)
- This is the basic idea behind pagerank algorithm

The Power Method

- A standard (but not ideal) algorithm to find the principal eigenvector is the *Power Method*
- Assume a matrix $\boldsymbol{A}$ with a full set of eigenvectors $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n)$ and no multiple eigenvalues.
- We start with a first guess $\boldsymbol{x}_0$ and then iterate as follows $\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1}$
- We have $\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} = \boldsymbol{A}^{\boldsymbol{k}}\boldsymbol{x}_{\boldsymbol{0}} = \boldsymbol{U}\boldsymbol{\Lambda}^k\boldsymbol{U}^{-1}\ \boldsymbol{x}_{\boldsymbol{0}} = c_1\lambda_1^k\boldsymbol{u}_1 + c_2\lambda_2^k\boldsymbol{u}_2 + \cdots + c_n\lambda_n^k\boldsymbol{u}_n$
- Assuming that the eigenvalues are in decreasing order of magnitude, then the first term (linked to the main eigenvector) becomes more and more dominant
- To avoid that the iteration diverges it is better to normalize it at each iteration leading to the standard power method routine:

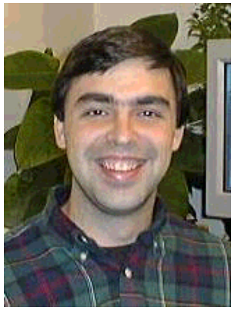
$$\boldsymbol{x}_k = \frac{\boldsymbol{A}\boldsymbol{x}_{\boldsymbol{k-1}}}{\|\boldsymbol{A}\boldsymbol{x}_{\boldsymbol{k-1}}\|} \to \boldsymbol{u}_1$$

To probe further

PageRank algorithm:

- Original paper: S.Brin and L. Page: "The anatomy of a large scale hypertextual Web search engine," Compt. Netw. 1998
- Patent: L.Page, "Methods for node ranking in a linked database", 1998
- John MacCormick, "Nine Algorithms That Changed the Future", 2012

Sergey Brin received his B.S. degree in mathematics and computer science from the University of Maryland at College Park in 1993. Currently, he is a Ph.D. candidate in computer science at Stanford University where he received his M.S. in 1995. He is a recipient of a National Science Foundation Graduate Fellowship. His research interests include search engines, information extraction from unstructured sources, and data mining of large text collections and scientific data.



Lawrence Page was born in East Lansing, Michigan, and received a B.S.E. in Computer Engineering at the University of Michigan Ann Arbor in 1995. He is currently a Ph.D. candidate in Computer Science at Stanford University. Some of his research interests include the link structure of the web, human computer interaction, search engines, scalability of information access interfaces, and personal data mining.

Condition number of a matrix

- Assume $\boldsymbol{A}$ is non-singular and that $\boldsymbol{A} = \boldsymbol{S\Lambda}\,\boldsymbol{S}^{-1}$
- In theory you can always solve $\boldsymbol{Ax} = \boldsymbol{b}$, but in practice you could have e.g rounding errors. You want to understand the effect of a small perturbation of $\boldsymbol{b}$ to the solution: $\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{\delta x}) = (\boldsymbol{b} + \boldsymbol{\delta b})$
- Assume $\boldsymbol{A}$ is symmetric and positive definite (all eigenvalues are real and positive) with a full set of eigenvectors $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n)$
- We are interested in $\frac{\|\boldsymbol{\delta x}\|}{\|\boldsymbol{x}\|}$ and worst case is when $\|\boldsymbol{\delta x}\|$ is large and $\|\boldsymbol{x}\|$ is small
- First focus on $\boldsymbol{A}(\boldsymbol{\delta x}) = \boldsymbol{\delta b}$. Solution: $\delta\boldsymbol{x} = c_1\lambda_1^{-1}\boldsymbol{u}_1 + c_2\lambda_2^{-1}\boldsymbol{u}_2 + \cdots + c_n\lambda_n^{-1}\boldsymbol{u}_n$ and worst case $\delta\boldsymbol{x} = c_1\lambda_1^{-1}\boldsymbol{u}_1 = \mathrm{c}_1\lambda_{min}^{-1}\boldsymbol{u}_1$.
- Similarly, $\|\boldsymbol{x}\| \geq \frac{\|\boldsymbol{b}\|}{\lambda_{max}}$. Therefore $\frac{\|\boldsymbol{\delta x}\|}{\|\boldsymbol{x}\|} \leq \frac{\lambda_{max}\|\boldsymbol{\delta b}\|}{\lambda_{min}\|\boldsymbol{b}\|}$, the ratio $\lambda_{max}/\lambda_{min}$ is called the condition number of $\boldsymbol{A}$
- In the more general case the condition number is $\|\boldsymbol{A}\|\left\|\boldsymbol{A}^{-1}\right\|$