Numerical methods

System of linear equations

OUTLINE

1. System of linear equations (linear system)
2. Gaussian elimination
3. Pivoting
4. LU decomposition
5. Cholesky decomposition
6. Effect of round-off errors
7. Conditionality
8. System of homogeneous linear equations

Basic definitions

$$
\begin{aligned}
x_1 a_{11} &+ \cdots + x_n a_{1n} = b_1 \\
x_1 a_{21} &+ \cdots + x_n a_{2n} = b_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
x_1 a_{m1} &+ \cdots + x_n a_{mn} = b_m
\end{aligned}
\tag{1}
$$

We can assign two matrixes to the system (1):

- **coefficient matrix**
- **augmented matrix**

$$
A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
\qquad
A' = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}
$$

Denote

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Then the linear system (1) reads

$$\begin{pmatrix} \mathrm{a}_{11} & \cdots & \mathrm{a}_{1\mathrm{n}} \\ \mathrm{a}_{21} & \cdots & \mathrm{a}_{2\mathrm{n}} \\ \cdots & \cdots & \cdots \\ \mathrm{a}_{\mathrm{m}1} & \cdots & \mathrm{a}_{\mathrm{mn}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$\mathbf{Ax} = \mathbf{b}$$

If $\mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ then the system (1) is called **homogeneous**

if $\mathbf{b} \neq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ then it is called **nonhomogeneous**

System of linear equations (linear system)

Definition We say that the n-tuple $(r_1, \ldots, r_n)$ is **solution of system** (1), if $\mathbf{Ar} = \mathbf{b}$, where

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}.$$

Note

System (1) is **consistent**,
if it has at least one solution;
otherwise it is **inconsistent**.

Direct methods

attempt to solve the problem
by a finite sequence of operations.
In the absence of rounding errors,
direct methods would deliver an exact solution.

Iterative methods

delivers only approximate solution.
The number of steps depends
on required precision.

Example: Find the solution of linear system in real numbers.

$$\begin{aligned} x-2y+4z+t &= -6 \\ 2x+3y-z+2t &= 13 \\ 2x+5y+z+t &= 8 \\ 3x+y+3z+t &= 1 \end{aligned}$$

$$\begin{aligned} x-2y+4z+t &= -6 \\ 7y-9z &= 25 \\ 9y-7z-t &= 20 \\ 7y-9z-2t &= 19 \end{aligned}$$

$$\begin{aligned} x-2y+4z+t &= -6 \\ 7y-9z &= 25 \\ 32/7z-t &= -85/7 \\ -2t &= -6 \end{aligned}$$

$$x = 1,\ y = 1,\ z = -2,\ t = 3$$

Definition: **Elementary row operations** (ERO) on a matrix are:

- row switching $R_j \leftrightarrow R_j$
- row multiplication $kR_i \to R_i$
- row addition $R_i + kR_j \to R_i$

Definition: We say that matrix $\mathbf{A}$ of type $m \times n$ is **row equivalent** with the matrix $\mathbf{B}$ of type $m \times n$ if it is possible to transform $\mathbf{A}$ into $\mathbf{B}$ by a sequence of elementary row operations.

Definition: We say that squared matrix **U** is **upper triangular (**or **right triangular)** if all the entries below the main diagonal are zero.

$$\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1,n-1} & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2,n-1} & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3,n-1} & u_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & u_{n-1,n-1} & u_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & u_{nn} \end{pmatrix}$$

Theorem: Each matrix is row equivalent
with some upper triangular matrix.

Definition: **The rank of matrix A** is
the number of nonzero rows
of triangular matrix equivalent
to the matrix **A**,
we will denote it $h(\mathbf{A})$.

or

The rank of A is the maximal number
of linearly independent rows of **A**.

Consequence: Row equivalent matrices have the same rank.

Example: Find the solution of linear system in real numbers.

```
 x-2y+4z+ t =-6
2x+3y- z+2t =13 /(2)-2*(1)
2x+5y+ z+ t = 8 /(3)-2*(1)
3x+ y+3z+ t = 1 /(4)-3*(1)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & -\mathbf{6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 2 & 5 & 1 & 1 & \mathbf{8} \\ 3 & 1 & 3 & 1 & \mathbf{1} \end{pmatrix}$$

```
 x-2y+4z+ t =  -6
   7y-9z    =  25
   9y-7z- t =  20 /(3)-9/7*(2)
   7y-9z-2t =  19 /(4)-7/7*(2)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 9 & -7 & -1 & \mathbf{20} \\ 0 & 7 & -9 & -2 & \mathbf{19} \end{pmatrix}$$

```
x-2y+  4z+ t =   -6
  7y-  9z    =   25
    32/7z- t =-85/7
          -2t =   -6
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 0 & 32/7 & -1 & \mathbf{-85/7} \\ 0 & 0 & 0 & -2 & \mathbf{-6} \end{pmatrix}$$

$$x = 1,\ \ y = 1,\ \ z = -2,\ \ t = 3$$

We showed a principle of
Gaussian elimination (GE):

Using the finite number of elementary row operations
on the coefficient matrix we get
the **upper triangular matrix**
and
using back substitution we calculate the vector of unknowns.

We showed a principle of
Gaussian elimination (GE):

Using the finite number of elementary row operations
on the coefficient matrix we get
the **upper triangular matrix**
and
using back substitution we calculate the vector of unknowns.

Forward elimination:

Let
$\mathbf{A}^{(0)} = \mathbf{A}$, $\mathbf{b}^{(0)} = \mathbf{b}$
elements of matrix $\mathbf{A}^{(0)}$ will be $a_{ij}^{(0)} = a_{ij}$
elements of vector $\mathbf{b}^{(0)}$ will be $b_i^{(0)} = b_i$

algorithm of GE

for $k := 1$ **to** $n-1$ **do**
begin
$\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)}$; $\mathbf{b}^{(k)} := \mathbf{b}^{(k-1)}$;
for $i := k+1$ **to** n **do**
begin
$m_{ik} := a_{ik}^{(k)} / a_{kk}^{(k)}$;
for $j := k+1$ **to** n **do** $a_{ij}^{(k)} := a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}$;
$b_i^{(k)} := b_i^{(k)} - m_{ik} b_k^{(k)}$;
end
end

algorithm of GE

for $k := 1$ **to** $n-1$ **do**
begin
  $\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)}$; $\mathbf{b}^{(k)} := \mathbf{b}^{(k-1)}$
  for $i := k+1$ **to** n **do**
  begin
    $m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$;
    for $j := k+1$ **to** n **do** $a_{ij}^{(k)} := a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}$;
    $b_i^{(k)} := b_i^{(k)} - m_{ik} b_k^{(k)}$;
  end
end

multiplicator assures that
element (i,k) of matrix $\mathbf{A}^{(k)}$
will be zero

algorithm of GE

for $k := 1$ **to** $n-1$ **do**
begin
$\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)}$; $\mathbf{b}^{(k)} := \mathbf{b}^{(k-1)}$;
for $i := k+1$ **to** n **do**
begin
$m_{ik} := a_{ik}^{(k)} / a_{kk}^{(k)}$;
for $j := k+1$ **to** n **do** $a_{ij}^{(k)} := a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}$;
$b_i^{(k)} := b_i^{(k)} - m_{ik} b_k^{(k)}$;
end
end

leading coefficient (pivot)
has to be non-zero

Example:
Solve the linear system

$$\begin{pmatrix} 10 & -7 & 0 \\ -3 & 2{,}099 & 6 \\ 5 & -1{,}1 & 4{,}8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 3{,}901 \\ 5{,}9 \end{pmatrix}$$

on the hypothetical computer, which works in decimal system with five digits mantissa.

The exact solution is

$$\mathbf{x} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Gaussian elimination

$$\begin{pmatrix} 10 & -7 & 0 \\ -3 & 2{,}099 & 6 \\ 5 & -1{,}1 & 4{,}8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 3{,}901 \\ 5{,}9 \end{pmatrix}$$

$$\begin{pmatrix} 10 & -7 & 0 \\ 0 & -0{,}001 & 6 \\ 0 & 2{,}4 & 4{,}8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 6{,}001 \\ 2{,}4 \end{pmatrix}$$

$$(4{,}8 + 6 \cdot 2400) \cdot x_3 = 2{,}4 + 6{,}001 \cdot 2400\,.$$

$$4{,}8 + 6 \cdot 2400 = 14404{,}8 \qquad 6{,}001 \cdot 2400 = 14402{,}4$$

$$14405\, x_3 = 14404\,.$$

$$x_3 = \frac{14404}{14\,405} \doteq 0{,}99993$$

Gaussian elimination

$$\begin{pmatrix} 10 & -7 & 0 \\ 0 & -0{,}001 & 6 \\ 0 & 2{,}4 & 4{,}8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 6{,}001 \\ 2{,}4 \end{pmatrix}$$

$$x_3 = \frac{14404}{14\,405} \doteq 0{,}99993$$

$$-0{,}001\, x_2 + 6 \cdot 0{,}99993 = 6{,}001$$

$$6 \cdot 0{,}99993 \doteq 5{,}9996$$

$$x_2 = \frac{0{,}0014}{-0{,}001} = -1{,}4$$

$$10x_1 - 7 \cdot (-1{,}4) = 7$$

$$x_1 = -0{,}28$$

$$\tilde{\mathbf{x}} = \begin{pmatrix} -0{,}28 \\ -1{,}4 \\ 0{,}99993 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Partial pivoting

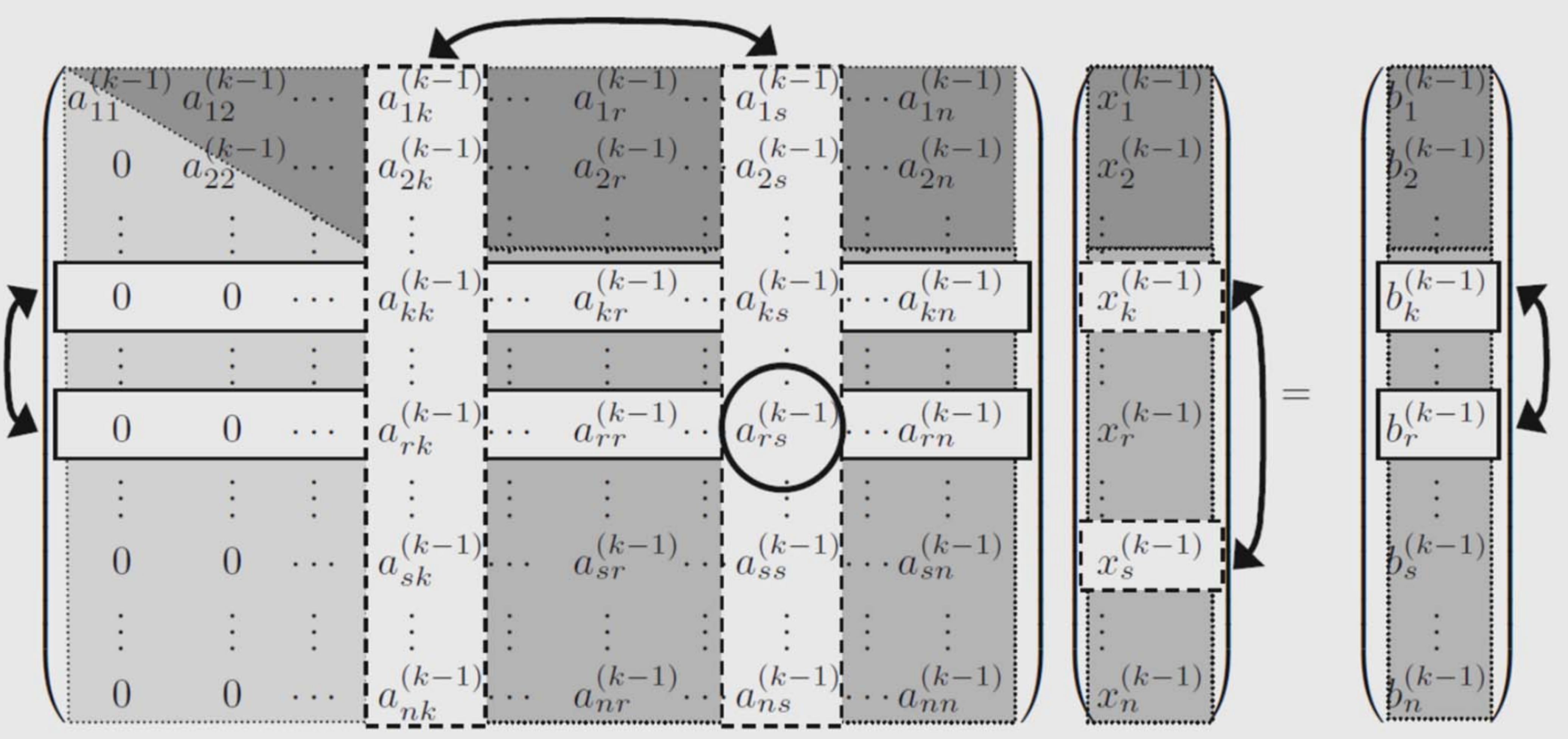
is the modification of **GE** which assures
that the absolute values of multiplicators
are less than or equal to 1.

The algorithm selects the entry
with largest absolute value
from the column of the matrix
that is currently being considered
as the pivot element.

Gaussian elimination – partial pivoting

$$
\begin{pmatrix}
a_{11}^{(k-1)} & a_{12}^{(k-1)} & \cdots & a_{1k}^{(k-1)} & \cdots & a_{1n}^{(k-1)} \\
0 & a_{22}^{(k-1)} & \cdots & a_{2k}^{(k-1)} & \cdots & a_{2n}^{(k-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & a_{kk}^{(k-1)} & \cdots & a_{kn}^{(k-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & a_{rk}^{(k-1)} & \cdots & a_{rn}^{(k-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & a_{nk}^{(k-1)} & \cdots & a_{nn}^{(k-1)}
\end{pmatrix}
\begin{pmatrix}
x_1 \\ x_2 \\ \vdots \\ x_k \\ \vdots \\ x_r \\ \vdots \\ x_n
\end{pmatrix}
=
\begin{pmatrix}
b_1^{(k-1)} \\ b_2^{(k-1)} \\ \vdots \\ b_k^{(k-1)} \\ \vdots \\ b_r^{(k-1)} \\ \vdots \\ b_n^{(k-1)}
\end{pmatrix}
$$

Gaussian elimination – **complete** (maximal) pivoting



Partial or **complete pivoting?**

usually partial pivoting is enough

Partial pivoting is generally sufficient to adequately reduce round-off error.

Complete pivoting is usually not necessary
to ensure numerical stability and,
due to the additional cost
of searching for the maximal element,
the improvement in numerical stability
is typically outweighed
by its reduced efficiency
for all but the smallest matrices.

Computational efficiency

The number of arithmetic operations in forward elimination: $\approx \dfrac{2}{3} n^3$

The number of arithmetic operations in back substitution : $\approx n^2$

LU decomposition (factorization)

```
 x-2y+4z+ t =-6
2x+3y- z+2t =13 /(2)-2*(1)
2x+5y+ z+ t = 8 /(3)-2*(1)
3x+ y+3z+ t = 1 /(4)-3*(1)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & -\mathbf{6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 2 & 5 & 1 & 1 & \mathbf{8} \\ 3 & 1 & 3 & 1 & \mathbf{1} \end{pmatrix}$$

```
 x-2y+4z+ t =  -6
   7y-9z    =  25
   9y-7z- t =  20 /(3)-9/7*(2)
   7y-9z-2t =  19 /(4)-7/7*(2)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 9 & -7 & -1 & \mathbf{20} \\ 0 & 7 & -9 & -2 & \mathbf{19} \end{pmatrix}$$

```
x-2y+  4z+ t =   -6
  7y-  9z    =   25
     32/7z- t =-85/7
          -2t =   -6
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 0 & 32/7 & -1 & \mathbf{-85/7} \\ 0 & 0 & 0 & -2 & \mathbf{-6} \end{pmatrix}$$

$$x = 1,\ \ y = 1,\ \ z = -2,\ \ t = 3$$

LU decomposition (factorization)

```
 x-2y+4z+ t =-6
2x+3y- z+2t =13  /(2)-2*(1)
2x+5y+ z+ t = 8  /(3)-2*(1)
3x+ y+3z+ t = 1  /(4)-3*(1)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 2 & 5 & 1 & 1 & \mathbf{8} \\ 3 & 1 & 3 & 1 & \mathbf{1} \end{pmatrix}$$

```
 x-2y+4z+ t =  -6
   7y-9z    =  25
   9y-7z- t =  20  /(3)-9/7*(2)
   7y-9z-2t =  19  /(4)-7/7*(2)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 9 & -7 & -1 & \mathbf{20} \\ 3 & 7 & -9 & -2 & \mathbf{19} \end{pmatrix}$$

```
x-2y+  4z+ t =    -6
  7y-  9z    =    25
     32/7z- t =-85/7
          -2t =    -6
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 0 & 32/7 & -1 & \mathbf{-85/7} \\ 3 & 0 & 0 & -2 & \mathbf{-6} \end{pmatrix}$$

$$x = 1,\ y = 1,\ z = -2,\ t = 3$$

LU decomposition (factorization)

```
 x-2y+4z+ t =-6
2x+3y- z+2t =13  /(2)-2*(1)
2x+5y+ z+ t = 8  /(3)-2*(1)
3x+ y+3z+ t = 1  /(4)-3*(1)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 2 & 5 & 1 & 1 & \mathbf{8} \\ 3 & 1 & 3 & 1 & \mathbf{1} \end{pmatrix}$$

```
 x-2y+4z+ t =  -6
   7y-9z    =  25
   9y-7z- t =  20  /(3)-9/7*(2)
   7y-9z-2t =  19  /(4)-7/7*(2)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 9 & -7 & -1 & \mathbf{20} \\ 3 & 7 & -9 & -2 & \mathbf{19} \end{pmatrix}$$

```
x-2y+  4z+ t =   -6
  7y-  9z    =   25
     32/7z- t =-85/7
          -2t =   -6
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 9/7 & 32/7 & -1 & \mathbf{-85/7} \\ 3 & 1 & 0 & -2 & \mathbf{-6} \end{pmatrix}$$

$$x = 1, \; y = 1, \; z = -2, \; t = 3$$

LU decomposition (factorization)

```
 x-2y+4z+ t =-6
2x+3y- z+2t =13 /(2)-2*(1)
2x+5y+ z+ t = 8 /(3)-2*(1)
3x+ y+3z+ t = 1 /(4)-3*(1)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 2 & 5 & 1 & 1 & \mathbf{8} \\ 3 & 1 & 3 & 1 & \mathbf{1} \end{pmatrix}$$

```
 x-2y+4z+ t =  -6
   7y-9z    =  25
   9y-7z- t =  20 /(3)-9/7*(2)
   7y-9z-2t =  19 /(4)-7/7*(2)
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 9 & -7 & -1 & \mathbf{20} \\ 3 & 7 & -9 & -2 & \mathbf{19} \end{pmatrix}$$

```
x-2y+  4z+ t =   -6
  7y-  9z    =   25
     32/7z- t =-85/7
          -2t =   -6
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 7 & -9 & 0 & \mathbf{25} \\ 2 & 9/7 & 32/7 & -1 & \mathbf{-85/7} \\ 3 & 1 & 0 & -2 & \mathbf{-6} \end{pmatrix}$$

$$x = 1,\ \ y = 1,\ \ z = -2,\ \ t = 3$$

LU decomposition (factorization)

Multiplicators m_{ij} from the forward elimination are placed into lower triangular matrix $\mathbf{L}$)

$$\mathbf{L} = \begin{pmatrix} \ell_{11} & 0 & 0 & \cdots & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 & \cdots & 0 & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{n-1,1} & \ell_{n-1,2} & \ell_{n-1,3} & \cdots & \ell_{n-1,n-1} & 0 \\ \ell_{n1} & \ell_{n2} & \ell_{n3} & \cdots & \ell_{n,n-1} & \ell_{nn} \end{pmatrix},$$

$$\ell_{ik} := \begin{cases} 0 & \text{pro } i = 1, 2, \ldots, k-1\,, \\ 1 & \text{pro } i = k\,, \\ m_{ik} & \text{pro } i = k+1, k+2, \ldots, n\,, \end{cases} \qquad k = 1, 2, \ldots, n\,.$$

We obtained **LU** decomposition of matrix $\mathbf{A}$

(so called Doolittle version – ones on the diagonal of matrix $\mathbf{L}$)

because

$$\mathbf{A} = \mathbf{L} \cdot \mathbf{U}$$

$$\underbrace{\begin{pmatrix} 1 & -2 & 4 & 1 \\ 2 & 3 & -1 & 2 \\ 2 & 5 & 1 & 1 \\ 3 & 1 & 3 & 1 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 9/7 & 1 & 0 \\ 3 & 1 & 0 & 1 \end{pmatrix}}_{\mathbf{L}} \cdot \underbrace{\begin{pmatrix} 1 & -2 & 4 & 1 \\ 0 & 7 & -9 & 0 \\ 0 & 0 & 32/7 & -1 \\ 0 & 0 & 0 & -2 \end{pmatrix}}_{\mathbf{U}}$$

Doolittle algorithm
makes the **LU** decomposition
using the half number
of arithmetic operations

$$\approx \frac{1}{3} n^3$$

LU decomposition (factorization)

LU decomposition of **A**
is useful for solving of sequences of tasks

$$\mathbf{A}\mathbf{x}_i = \mathbf{b}_i$$

if the new right-hand-side $\mathbf{b}_i$ could be estimated only after
we solve the previous linear systems

$$\mathbf{A}\mathbf{x}_k = \mathbf{b}_k$$

for $k < i$.

To solve the system $\mathbf{L}\mathbf{U}\mathbf{x} = \mathbf{b}$ means to solve

$$\mathbf{L}\mathbf{y} = \mathbf{b}\,, \qquad \mathbf{U}\mathbf{x} = \mathbf{y}\,.$$

LU decomposition **with partial pivoting**
is standard routine of each library of codes.

Input parameter is matrix **A**.

Output are three matrices:
L, **U** and permutation matrix **P**
where
LU = **PA.**

Permutation matrix comes form identity matrix
after row exchange.

Solution of **Ax**=**b** is given by solution of

$$\mathbf{z} = \mathbf{P}\mathbf{b}\,, \qquad \mathbf{L}\mathbf{y} = \mathbf{z}\,, \qquad \mathbf{U}\mathbf{x} = \mathbf{y}\,.$$

(show this)

Instead of matrix **P** we work with
the **vector of row permutations p**.

At the beginning we put

$$\mathbf{p}^{(0)} = (1, 2, \ldots, n)^T$$

and we just exchange the elements of vector **p**.

Definition: We say that square matrix $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{n}$ is **diagonally dominant** if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \qquad i = 1, 2, \ldots, n$$

Definition: Symmetric matrix $\mathbf{A}$ is **positive-definite** if

for any non-zero vector x it holds

$$x^T \cdot \mathbf{A} \cdot x > 0$$

If $\mathbf{A}$ is regular then

$$\mathbf{A}^T \cdot \mathbf{A}$$

is positive-definite.

Sylvester's criterion: Symmetric matrix $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{n}$ is **positive-definite**, if and only if the leading principal minors are positive, i.e. if

$$a_{11} > 0\,, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0\,, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0\,, \ \ldots\,, \ \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} > 0\,.$$

Gaussian elimination
is numerically stable for
diagonally dominant
or
positive-definite matrices.

Cholesky decomposition theorem: There is only one way, how the decompose the positive-definite matrix $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{L} \cdot \mathbf{L}^T$$

where $\mathbf{L}$ is the lower triangular matrix.

Cholesky decomposition theorem: There is only one way, how the decompose the positive-definite matrix $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{L} \cdot \mathbf{L}^T$$

where $\mathbf{L}$ is the lower triangular matrix.

Cholesky algorithm:

$$\ell_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} \ell_{kj}^2}\,,$$

$$\ell_{ik} = \frac{1}{\ell_{kk}} \left(a_{ik} - \sum_{j=1}^{k-1} \ell_{ij} \ell_{kj} \right), \qquad i = k+1, k+2, \ldots, n\,.$$

The number of arithmetic operations: $\approx \frac{1}{3} n^3$

The effect of round-off errors

Example:

On the hypothetical computer working decimal system with three digits mantissa solve the following system

$$\begin{pmatrix} 3{,}96 & 1{,}01 \\ 1 & 0{,}25 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5{,}03 \\ 1{,}25 \end{pmatrix}.$$

The effect of round-off errors

GE with partial pivoting
assures the small residua

However, the small residuum does not give
an estimation of the error of solution
$\tilde{\mathbf{x}} - \mathbf{x}$.

Conditionality

In order to evaluate the conditionality of problem

„to find solution **x** of system **Ax** = **b**“

we need to asses

the effect of change of **A** a **b**

to the solution **x**.

Norm of the vector

class of vector norms l_p,

$1 \leq p \leq \infty$:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

The most often we use p=1, p=2, or $p \longrightarrow \infty$:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \,, \qquad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \,, \qquad \|\mathbf{x}\|_\infty = \max_i |x_i| \,.$$

Manhattan Euclid Chebyshev

Properties

1. $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|,$ (*absolute homogeneity* or *absolute scalability*)

2. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (*triangle inequality* or *subadditivity*)

3. If $\|\mathbf{v}\| = 0$ then $\mathbf{v}$ is the zero vector (*separates points*)

Condition number of a matrix

Let

$$M = \max_{\mathbf{x} \neq \mathbf{o}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}, \qquad m = \min_{\mathbf{x} \neq \mathbf{o}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|},$$

where maximum and minimum is considered for all nonzero **x**.

Ratio M/m is called **condition number of matrix A**:

$$\kappa(\mathbf{A}) = \frac{M}{m} = \left(\max_{\mathbf{x} \neq \mathbf{o}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \right) \cdot \left(\min_{\mathbf{x} \neq \mathbf{o}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \right)^{-1}.$$

Conditionality

Lets consider linear system

$$\mathbf{Ax} = \mathbf{b}$$

and other system, which could be obtained by changing the r.h.s.:

$\mathbf{A}(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$, , or $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$.

we can consider $\Delta\mathbf{b} = \tilde{\mathbf{b}} - \mathbf{b}$ as an error of **b**

and

$\Delta\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{x}$ corresponding error of solution **x**.

Conditionality

Because $\mathbf{A}\Delta\mathbf{x} = \Delta\mathbf{b}$ then
from the definition of *M* and *m*
it follows that

$$\|\mathbf{b}\| \leq M\|\mathbf{x}\|\,, \qquad \|\Delta\mathbf{b}\| \geq m\|\Delta\mathbf{x}\|\,,$$

so for $m \neq 0$,

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = \kappa(\mathbf{A})\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}\,,$$

where $\mathbf{r} = \mathbf{b} - \mathbf{A}(\mathbf{x} + \Delta\mathbf{x})$ is a residuum.

Condition number of matrix acts as an amplifier of relative error !

It is possible to show, that also changes in matrix **A** leads to the similar effects on results.

Norm of matrix

the number *M* is also known as the norm of matrix

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{o}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$$

and also it holds

$$\|\mathbf{A}^{-1}\| = 1/m$$

Equivalently we could write the condition number of matrix as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|$$

Conditionality

“Entrywise” norm of matrix

$$\left\|\mathbf{A}\right\|_p = \left\|vec\left(\mathbf{A}\right)\right\|_p = \left(\sum_{j=1}^{n}\sum_{j=1}^{n}\left|a_{ij}\right|^p\right)^{\frac{1}{p}}$$

$$\left\|\mathbf{A}\right\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2}$$

Frobenius norm (consistent with l_2)

$$\left\|\mathbf{A}\right\|_\infty = \max_{i,j}\left|a_{ij}\right|$$

max norm

Example: Find the solution of linear system in real numbers.

$$
\begin{aligned}
x+2y+z-t &= 1\\
2x+3y-z+2t &= 3\\
4x+7y+z &= 5\\
5x+7y-4z+7t &= 10
\end{aligned}
$$

Example: Find the solution of linear system in real numbers.

$$\begin{aligned} x+2y+z-t &= 1\\ 2x+3y-z+2t &= 3\\ 4x+7y+z &= 5\\ 5x+7y-4z+7t &= 10 \end{aligned}$$

$$\begin{pmatrix} 1 & 2 & 1 & -1 & \mathbf{1}\\ 2 & 3 & -1 & 2 & \mathbf{3}\\ 4 & 7 & 1 & 0 & \mathbf{5}\\ 5 & 7 & -4 & 7 & \mathbf{10} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 1 & -1 & \mathbf{1}\\ 0 & -1 & -3 & 4 & \mathbf{1}\\ 0 & -1 & -3 & 4 & \mathbf{1}\\ 0 & -3 & -9 & 12 & \mathbf{5} \end{pmatrix}$$

$$\begin{aligned} x+2y+z-t &= 1\\ -y-3z+4t &= 1\\ 0z+0t &= 2\\ 0t &= 0 \end{aligned}$$

system has no solution

$$\begin{pmatrix} 1 & 2 & 1 & -1 & \mathbf{1}\\ 0 & -1 & -3 & 4 & \mathbf{1}\\ 0 & 0 & 0 & 0 & \mathbf{2}\\ 0 & 0 & 0 & 0 & \mathbf{0} \end{pmatrix}$$

Another example: Find the solution of linear system in rational numbers.

$$
\begin{aligned}
x - 2y + 4z + t &= -6 \\
2x + 3y - z + 2t &= 13 \\
3x + y + 3z + 3t &= 7 \\
x + 5y - 5z + t &= 19
\end{aligned}
$$

Another example: Find the solution of linear system in rational numbers.

$$\begin{aligned} x-2y+4z+t &= -6 \\ 2x+3y-z+2t &= 13 \\ 3x+y+3z+3t &= 7 \\ x+5y-5z+t &= 19 \end{aligned}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 2 & 3 & -1 & 2 & \mathbf{13} \\ 3 & 1 & 3 & 3 & \mathbf{7} \\ 1 & 5 & -5 & 1 & \mathbf{19} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 7 & -9 & 0 & \mathbf{25} \end{pmatrix}$$

$$\begin{aligned} x-2y+4z+t &= -6 \\ 7y-9z+0t &= 25 \\ 0z+0t &= 0 \\ 0t &= 0 \end{aligned}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{pmatrix}$$

Another example: Find the solution of linear system in rational numbers.

```
x-2y+4z+ t =-6
   7y-9z+0t =25
      0z+0t = 0
         0t = 0
```

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{-6} \\ 0 & 7 & -9 & 0 & \mathbf{25} \\ 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{pmatrix}$$

System has infinite number of solutions

$$\mathbf{t = p, \quad z = q,} \qquad p,q \in Q$$

$$\mathbf{y = \frac{25}{7} + \frac{9}{7}q} \qquad \mathbf{x = \frac{8}{7} - \frac{10}{7}q - p}$$

$$\mathbf{S = \left\{ \left(\frac{8}{7} - \frac{10}{7}q - p, \frac{25}{7} + \frac{9}{7}q, q, p \right), \ p, q \in Q \right\}}$$

For example: if $q = 1, p = 0$, we get particular solution $\left(\frac{-2}{7}, \frac{34}{7}, 1, 0 \right)$

Frobenius theorem: System of nonhomogeneous equations
has solution
if and only if,
the **rank of coefficient matrix**
is equal to
the **rank of augmented matrix**.

Consequence 1: If $h(\mathbf{A}) = h(\mathbf{A}') = n$ (n is the number of unknowns),
then the system has just one solution.

Consequence 2: If $h(\mathbf{A}) = h(\mathbf{A}') < n$ (n is the number of unknowns),
then the system has infinite number of solutions
and
n-h unknowns could be freely chosen.

Consequence 3: If $h(\mathbf{A}) \neq h(\mathbf{A}')$ then the system has no solution.

System of homogeneous linear equations

System of homogeneous linear equations

$$x_1 a_{11} + \ldots + x_n a_{1n} = 0$$
$$x_1 a_{21} + \ldots + x_n a_{2n} = 0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_1 a_{m1} + \ldots + x_n a_{mn} = 0$$

The system has always solution because $h(\mathbf{A}) = h(\mathbf{A}')$

Theorem: System of homogeneous equations has nontrivial solution if and only if $h(\mathbf{A}) < n$.
(trivial solution is (0,0,...,0))

Example: Find the solution of linear system in rational numbers.

$$\begin{aligned} x-2y+4z+t&=0\\ 2x+3y-z+2t&=0\\ 3x+y+3z+3t&=0\\ x+5y-5z+t&=0 \end{aligned}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{0} \\ 2 & 3 & -1 & 2 & \mathbf{0} \\ 3 & 1 & 3 & 3 & 0 \\ 1 & 5 & -5 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{0} \\ 0 & 7 & -9 & 0 & \mathbf{0} \\ 0 & 7 & -9 & 0 & 0 \\ 0 & 7 & -9 & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} x-2y+4z+t&=0\\ 7y-9z+0t&=0\\ 0z+0t&=0\\ 0t&=0 \end{aligned}$$

$$\begin{pmatrix} 1 & -2 & 4 & 1 & \mathbf{0} \\ 0 & 7 & -9 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{pmatrix}$$

Example: Find the solution of linear system in rational numbers.

$$
\begin{aligned}
x-2y+4z+\ t\ &=0\\
7y-9z+0t\ &=0\\
0z+0t\ &=0\\
0t\ &=0
\end{aligned}
$$

$$\mathbf{t=p,\quad z=q,}\quad p,q\in Q$$

$$\mathbf{y = \frac{9}{7}q}$$

$$\mathbf{x=-\frac{10}{7}q-p}$$

$$\mathbf{S=\left\{\left(-\frac{10}{7}q-p,\frac{9}{7}q,q,p\right),\ p,q\in Q\right\}}$$

Summary of notions

- homogeneous and nonhomogeneous system of linear equations
- coefficient matrix and augmented matrix
- elementary row operations
- row equivalent matrices
- rank of a matrix
- upper and lower triangular matrix
- Gaussian elimination (pivoting)
- LU decomposition (Doolittle method)
- Cholesky decomposition
- Frobenius theorem
- solution of homogeneous linear systems