Lecture 1

- Matrix Arithmetics
- Gaussian Elimination
- Matrix Inverses
- [Examples of Large Systems]
- Symmetric Matrices

Solving square nonsingular systems

One of the most fundamental objects in linear algebra (in fact, in all applied mathematics) is a <u>matrix</u>, and the most basic core problem is <u>solving systems of linear algebraic equations</u>. So, we start with a brief review of matrices and methods for solving systems of linear equations.

Def A matrix is a rectangular array of numbers : $A = \begin{array}{cc} & j \\ & \left[\begin{array}{ccc} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{array}\right] \end{array} i$

Here, m is the number of rows in A and n is the number of columns. A matrix is square if $m=n$.

general matrix of size $m \times n$

Matrix arithmetics involves three basic operations :

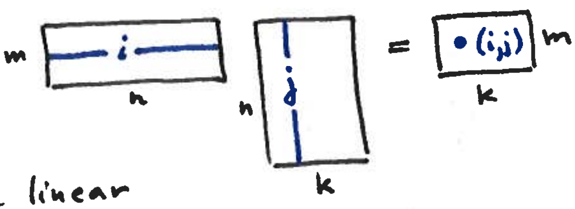
- <u>Matrix addition</u> : $C = A+B$, $c_{ij} = a_{ij} + b_{ij}$, $A, B, C \in \mathbb{M}_{m\times n}$ ← set of all $m\times n$ matrices

  It has usual properties : $A+B = B+A$ (commutative), $A+(B+C) = (A+B)+C$ (associative)

- <u>Scalar multiplication</u> $B = \alpha \cdot A$, $b_{ij} = \alpha a_{ij}$, $\alpha \in \mathbb{R}$

- <u>Matrix multiplication</u> $C = A \cdot B$, $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$, $A \in \mathbb{M}_{m\times n}$, $B \in \mathbb{M}_{n\times k}$, $C \in \mathbb{M}_{m\times k}$

  All usual properties are there, such as associativity distributivity, except $\boxed{AB \neq BA}$



The above defined matrix multiplication immediately allows to rewrite a general linear system of m equations with n unknowns

$$\begin{cases} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = b_1 \\ \ldots\ldots\ldots\ldots \\ a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n = b_m \end{cases}$$

Remark Why not to define it like $c_{ij} = a_{ij} \cdot b_{ij}$? You can, but it will be useless in applications.

in a compact matrix form : $\underline{Ax = b}$, where $A = [a_{ij}]$, $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

← an equality between two column vectors.

<u>A few test questions</u> :

1. $(A+B)^2 \overset{?}{=} A^2 + 2AB + B^2$
2. $AB = 0 \overset{?}{\Longrightarrow} A = 0$ or $B = 0$

   In particular, $A^2 = 0 \overset{?}{\Longrightarrow} A = 0$
3. The commutator of A and B

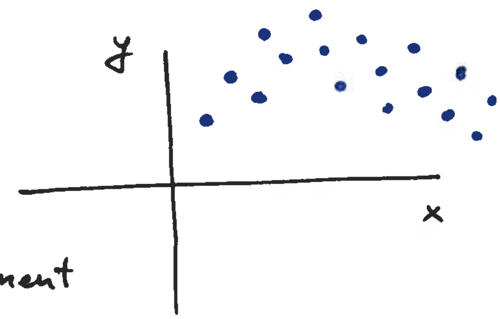
   $$[A,B] = AB - BA$$ ← plays important role in geometry, symmetry, quantum mechanics

   Check : it satisfies the Jacobi identity

   $$[[A,B],C] + [[C,A],B] + [[B,C],A] = 0$$

Before we proceed, let us discuss two examples, where large systems of linear equations appear in applications.

① <u>Curve Fitting</u>



Suppose we have a set of data points:

$(x_1, y_1) \ldots\ldots (x_n, y_n)$,

where (x_i, y_i) are measurements in a certain experiment

Suppose we want to fit a polynomial to the data,

e.g. find a polynomial $y = p(x)$ that passes through these points: $y_i = p(x_i)$

Given n points, it is enough to consider polynomials of deg $\leq n-1$.

Let $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$, where $a_i \in \mathbb{R}$

So the problem is to find the coefficients $a_0 \ldots a_{n-1}$ such that

$$\begin{cases} p(x_1) = a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_{n-1} x_1^{n-1} = y_1 \\ \vdots \\ p(x_n) = a_0 + a_1 x_n + a_2 x_n^2 + \ldots + a_{n-1} x_n^{n-1} = y_n \end{cases} \iff \underbrace{\begin{bmatrix} 1 & x_1 \ldots & x_1^{n-1} \\ 1 & x_2 \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots \\ 1 & x_n \ldots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{linear system of equations on } a_0 \ldots a_{n-1}}$$

Vandermonde matrix

The more data points we have, the larger the system.

② <u>Solving Differential Equations</u>

Very few differential equations can be solved analytically (intuitive reason: very few functions are analytically integrable)

In most applications, numerical solutions are required.

Consider the following equation (Poisson eq. in 1D)

$$-\frac{d^2 u}{dx^2} = f(x)$$

$$\left.\begin{aligned} u(0) &= 0 \\ u(1) &= 0 \end{aligned}\right\} \text{boundary condition}$$

It describes many simple physical phenomena:

- temperature distribution $(u(x))$ in a bar with a heat source $f(x)$
- deformation of an elastic bar
- deformation of aa string under tension.

In applications, the "source term" may not be even known in a closed form. We may just be able to measure f at any point x.

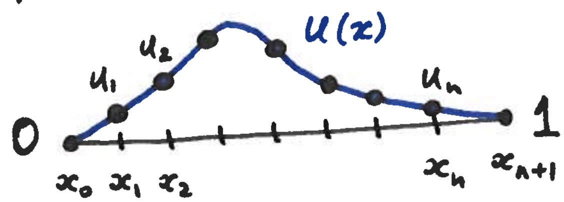
How to solve this problem?

We need to discretize it.

Remark: Numerical solutions of ODEs/PDEs are discussed in depth in ACM106b & ACM210ab.

Let us subdivide interval $[0,1]$ into $(n+1)$ equal subintervals:

$x_i = ih$, $i = 0, \dots, n+1$, $h = \frac{1}{n+1} \ll 1$.



Let $u_i = u(x_i)$. From the boundary condition, we know that $u_0 = 0$, $u_{n+1} = 0$.

If we find $u_1, \dots, u_n$, this will give us an approximation of $u(x)$.

The first step is to approximate $\frac{d^2u}{dx^2}$.

Assuming that h is small (n is large),

$$\frac{du}{dx} \approx \frac{u(x+h) - u(x-h)}{2h}$$

(this can be obtained as the average of two more direct approximations: $\frac{u(x+h)-u(x)}{h}$ and $\frac{u(x)-u(x-h)}{h}$)

$$\frac{d^2u}{dx^2} \approx \frac{u'(x+\frac{h}{2}) - u'(x-\frac{h}{2})}{h} =$$

$$= \frac{\frac{u(x+h)-u(x)}{h} - \frac{u(x)-u(x-h)}{h}}{h} = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2}$$

Therefore, $-\frac{d^2u}{dx^2} = f(x)$ leads to

$$\underbrace{-u_{i+1} + 2u_i - u_{i-1} = h^2 f_i}_{\text{difference equation}}, \text{ where } f_i = f(x_i), \quad i = 1, \dots, n.$$

This system of difference equations can be written in the matrix form:

$$\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ \vdots \\ u_n \end{bmatrix} = h^2 \begin{bmatrix} f_1 \\ \vdots \\ \vdots \\ \vdots \\ f_n \end{bmatrix}$$

To obtain an accurate approximation, the discretization step h should be small

$\Rightarrow$ n should be large

Numerical schemes for PDEs arising in fluid and solid mechanics, weather prediction, image and video processing, molecular dynamics, chemical processes, etc., often require $n \sim 10^6$ and more.

The design of efficient numerical algorithms for solving large systems (especially sparse $\uparrow$ many $a_{ij} = 0$) is an active area of research. For more: ACM 106a.

Let us focus on square matrices, $A \in \mathbb{M}_{n\times n}$

The inverse of a matrix is an analog of the reciprocal of a number.

Def Let $A \in \mathbb{M}_{n\times n}$. The inverse of A, denoted A^{-1}, is an $n\times n$ matrix that satisfies: $\underbrace{AA^{-1} = A^{-1}A = I_n}_{\text{we need this since multiplication, in general, is not commutative.}}$, where $I_n = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$ is the $n\times n$ identity matrix

Let us recall a few important properties of matrix inverses.

Remark If $A \in \mathbb{M}_{m\times n}$ we can define a right inverse of A as a matrix $X \in \mathbb{M}_{n\times m}$ such that $AX = I_m$; or left inverse of A as a matrix $Y \in \mathbb{M}_{n\times m}$ such that $YA = I_n$. But we will not discuss this now.

① If A^{-1} exists, then it is unique.

Proof: Suppose $AX = XA = I$ and $AY = YA = I$

Then $X = XI = X(AY) = (XA)Y = Y$ $\square$

② $(A^{-1})^{-1} = A$

Proof: All we need to check is that the definition is satisfied: $\underbrace{A^{-1}}_{\text{we are looking for an inverse of this } A^{-1}} \cdot \underbrace{A}_{\text{inverse candidate}} = I_n$, similarly $A \cdot A^{-1} = I_n$ since A^{-1} is the inverse of A $\square$

this is so tautological that may look confusing.

③ If $A, B \in \mathbb{M}_{n\times n}$ are invertible $\Rightarrow (AB)^{-1} = B^{-1}A^{-1}$

Proof: simple check. $(AB)\cdot(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$

$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$ $\square$

Remark In general $(A_1 \dots A_k)^{-1} = A_k^{-1} \dots A_1^{-1}$

From the theory of matrices, it follows that asking $AA^{-1} = I$ or $A^{-1}A = I$ in the definition is enough.

④ If $AX = I_n$, then $XA = I_n$ and thus $X = A^{-1}$

If $XA = I_n$, then $AX = I_n$ and thus $X = A^{-1}$.

Not all square matrices have the inverse (but most!)

Def A noninvertible matrix A is called singular. If A^{-1} exists $\Rightarrow$ A is nonsingular.

(noninvertible: A^{-1} does not exist)

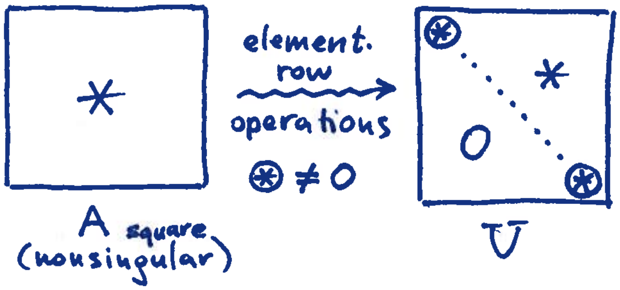
The original motivation for introducing the matrix inverse is that it allows to write the solution of any linear system in a compact way:

If A is nonsingular, then the unique solution of $Ax = b$ is $x = A^{-1}b$.

However, finding the inverse A^{-1} (using e.g. the Gauss-Jordan method) is computationally inefficient as compared to direct Gaussian Elimination, which provides a systematic method for solving linear systems. Nevertheless, A^{-1} is of great theoretical importance and provides insights into the design of practical algorithms.

Gaussian Elimination

Recall that any non-singular matrix A can be reduced to upper triangular matrix U with all non-zero diagonal elements by elementary row operations.



(1) $a^i \rightsquigarrow a^i + \alpha a^j$, $i>j$, where $\alpha \in \mathbb{R}$, a^i is the i^{th} row

(2) $\left.\begin{array}{l} a^i \rightsquigarrow a^j \\ a^j \rightsquigarrow a^i \end{array}\right\}$ interchange the i^{th} and the j^{th} rows.

Recall also that the elem. row operations can be realized by multiplication of the original matrix A by the so called elementary matrices.

Let

$$\underbrace{E^{\alpha}_{ij} = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cc} & \begin{array}{c} j \end{array} \\ & \begin{bmatrix} 1 & \cdots & 0 & 0 \\ & \ddots & \alpha & \cdots \\ & 0 & \ddots & 0 \\ & & & 1 \end{bmatrix} \begin{array}{c} \\ i \\ \\ \\ \end{array}\end{array}}_{\text{elem. matrix of type 1}} \quad i<j$$

$$\underbrace{P_{ij} = \begin{array}{cc} & \begin{array}{cc} i & j \end{array} \\ & \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 0 & \cdots & 1 & \\ & & & \vdots & 1 & \vdots & \\ & & & 1 & \cdots & 0 & \\ & & & & & & 1 \ddots 1 \end{bmatrix} \begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \end{array}\end{array}}_{\text{elem. matrix of type 2}}$$

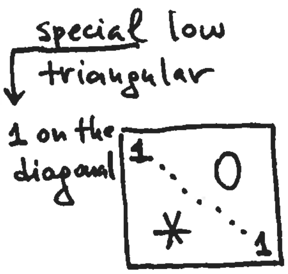
$\Rightarrow$ (1) $\Leftrightarrow$ $A \rightsquigarrow E^{\alpha}_{ij} A$

(2) $\Leftrightarrow$ $A \rightsquigarrow P_{ij} A$

Moreover, we can reduce A to U by first applying elementary row operations of the second type (permutation of rows), and then applying elem. row. oper. of type (1).

In other words, $\exists$ $\underbrace{P_1, \dots, P_k}_{\text{elem. matr. of type 2}}$ and $\underbrace{E_1 \dots E_m}_{\text{elem. mat. of type 1}}$ such that

$$\underbrace{E_1 \cdot \ldots \cdot E_m}_{\to\, E} \cdot \underbrace{P_1 \cdot \ldots P_k}_{P \,\leftarrow\, \text{permutation matrix}} A = U$$



since $i>j$, all $E_1 \dots E_m$ are special low triangular. $\Rightarrow$ so is the product and E^{-1}.

permutation matrix: It is a matrix where all rows and all columns contain all zeros except for a single 1.

Remark P is obtained from I_n by a finite number of row permutations.

Denoting $E^{-1} = L$, we obtain the well known permuted LU factorization of a nonsingular matrix:

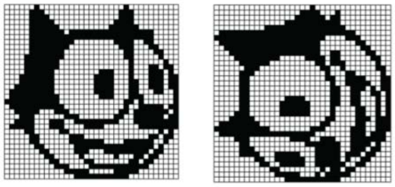
$$\boxed{PA = LU}$$

- U is upper triangular
- L is special lower triangular
- P is a permutation matrix.

Once the permuted LU factorization is obtained, it is easy to solve $Ax = b$

1) $PAx = Pb = \tilde{b} \Rightarrow L\underbrace{Ux}_{y} = \tilde{b}$.
2) Solve $Ly = \tilde{b}$ by forward substitution.
3) Solve $Ux = y$ by back substitution.

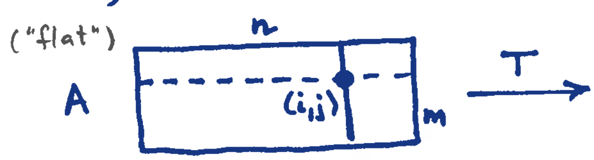
Check If $Ux = y \Rightarrow LUx = Ly = \tilde{b} = Pb$
$\Rightarrow PAx = Pb \Rightarrow Ax = b$.

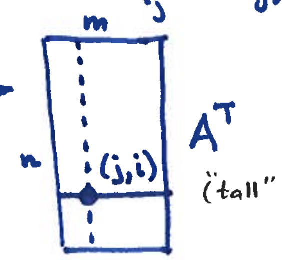


Matrix   Transpose of Matrix

Matrix operations that we discussed above, in one way or another, generalizations of the corresponding operations on scalars. A fundamentally new operation is <u>matrix transposition</u> : interchange of matrix rows and columns.

Def: Let A be an $m \times n$ matrix, then its transpose, denoted A^T, is the $n \times m$ matrix with $(A^T)_{ij} = A_{ji}$





Rows of A are columns of A^T
Columns of A are rows of A^T

<u>Remark</u>: In the square case, $A \to A^T$ is "reflecting" the matrix entries across the main diagonal:

<u>Basic properties</u> : $(A^T)^T = A$, $(A+B)^T = A^T + B^T$, $(\alpha A)^T = \alpha A^T$, $\boxed{(AB)^T = B^T A^T}$

$(ABC)^T = C^T B^T A^T$

Inversion and transposition respect each other :

Th: If A is a nonsingular matrix $\Rightarrow$ so is A^T, and $(A^T)^{-1} = (A^{-1})^T$

↑ it is often denoted A^{-T}

<u>Proof</u>: Let $B = (A^{-1})^T$. We need to show that $BA^T = I$.

$$BA^T = (A^{-1})^T \cdot A^T = (A^{-1} \cdot A)^T = I^T = I \qquad \square$$

<u>Test questions</u>:

1. when $(AB)^T = A^T B^T$? $(\Leftrightarrow AB = BA)$

   $(\Leftarrow)$ $(AB)^T = (BA)^T = A^T B^T$

   $(\Rightarrow)$ $(AB)^T = A^T B^T = (BA)^T \Rightarrow AB = BA$

2. True or False : every square matrix commutes with its transpose ? False $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

If $AA^T = A^T A$, then A is called normal

A particular important class of square matrices is symmetric matrices, which are invariant w.r.t. transposition

Def A (square) matrix is called symmetric if $A^T = A$.

special case of normal

<u>Test questions</u> :

1. True or False : if A is symmetric, then so is A^2 ? True : $(A^2)^T = (AA)^T = A^T A^T = AA = A^2$
2. True or False : if A is nonsingular symmetric, then so is A^{-1} ?

   True : $(A^{-1})^T \overset{Th}{=} (A^T)^{-1} = A^{-1}$
3. True or False : if A and B are symmetric $n \times n$ matrices, then so is AB ?

   False : $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$