Matrix

Fndt'n of IS & Data Anlys

Contents

- Matrix notation
- Special Matrices
- Transpose, Addition and Multiplication
- Determinant and Inverse
- Linear Equations
- Eigenvector and Eigenvalue
- Orthonormal Matrix and Application

What is Matrix?

A matrix is a rectangular array of numbers written between rectangular brackets or round brackets

$$\begin{bmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{bmatrix} \quad \begin{pmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{pmatrix}$$

Size : number of rows (m) and number of columns (n)

A m-by-n ($m \times n$) matrix $\boldsymbol{A}$ has m rows and n columns

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

a_{ij} is the element in the i-th row and the j-th column.

Square matrix $m = n$

In theoretical mathematics, we call $\boldsymbol{A} \subset \mathbb{R}^{m \times n}$ ($m \times n$ real number array)

In some textbooks, the indices i, j may start from zero.

Column and row vectors

An n-vector can be interpreted as an $n \times 1$ matrix.

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

A matrix with only one row, i.e., with $1 \times n$, is called a row vector. n-row-vector .

$$\boldsymbol{b} = (b_1 \quad b_2 \quad b_3)$$

Columns and rows of a matrix

Given a Matrix indexing

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

The j-th column is $\boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$ is a column vector

$$\begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

The i-th row $\boldsymbol{b}_i = [a_{i1} \quad \cdots \quad a_{in}]$ is a row vector

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Block matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{D} & \boldsymbol{E} \end{bmatrix}$$

where $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}$, are matrices with suitable sizes

$$B = \begin{bmatrix} 0 & 2 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} -1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 5 \end{bmatrix}, \quad E = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 2 & 3 & -1 \\ 2 & 2 & 1 & 4 \\ 1 & 3 & 5 & 4 \end{bmatrix}$$

Submatrix

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 2 & 3 & -1 \\ 2 & 2 & 1 & 4 \\ 1 & 3 & 5 & 4 \end{bmatrix}.$$

$$\boldsymbol{A}_{p:q,r:s} = \begin{bmatrix} a_{pr} & a_{p,r+1} & \cdots & a_{a_{ps}} \\ a_{p+1,r} & a_{p+1,r+1} & \cdots & a_{p+1,s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{qr} & a_{q,r+1} & \cdots & a_{qs} \end{bmatrix}$$

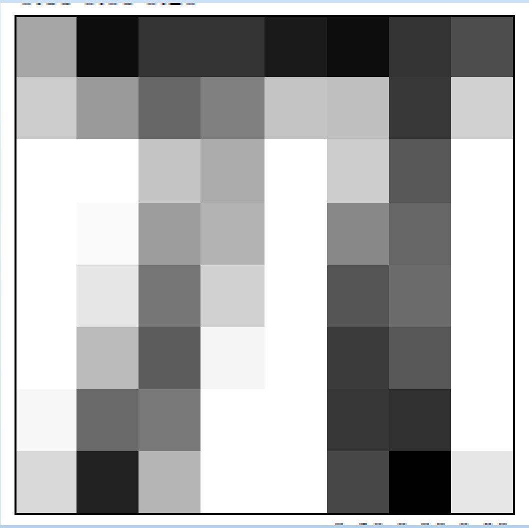
$$A_{2:3,3:4} = \begin{bmatrix} 1 & 4 \\ 5 & 4 \end{bmatrix}.$$

- Use colon notation to denote submatrices.
- p, q, r, s are integers, $1 \leq p < q \leq m, 1 \leq r < s \leq n$
- $\boldsymbol{A}_{p:q,r:s}$ is obtained from rows p through q and rows r through s

Examples

Images:

A black and white image with MN pixels is naturally represented as an $M \times N$ matrix.



Examples

- Asset returns. A $T \times n$ matrix $\boldsymbol{R}$ gives the returns of a collection of n assets (called the universe of assets) over T periods, with r_{ij} giving the return of asset j in period i.

Date	AAPL	GOOG	MMM	AMZN
March 1, 2016	0.00219	0.00006	−0.00113	0.00202
March 2, 2016	0.00744	−0.00894	−0.00019	−0.00468
March 3, 2016	0.01488	−0.00215	0.00433	−0.00407

Table 6.1 Daily returns of Apple (AAPL), Google (GOOG), 3M (MMM), and Amazon (AMZN), on March 1, 2, and 3, 2016 (based on closing prices).

Examples

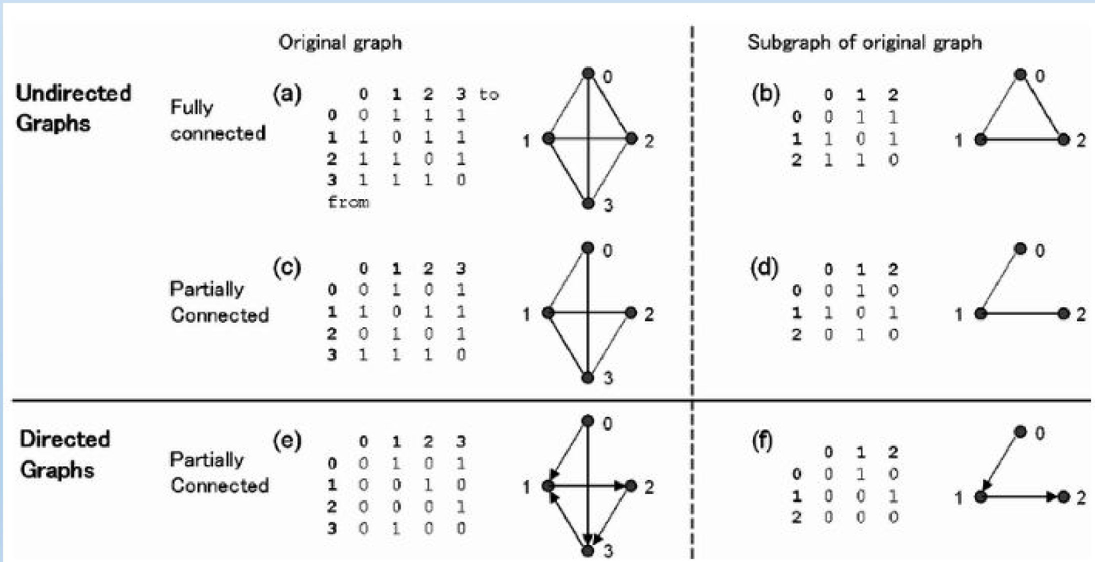
Matrix representation of a collection of vectors.

- Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N$ are n -vectors that give that give the n feature values for each of N objects, we can collect them all into one $n \times N$ matrix

- $\boldsymbol{X} = [\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_N]$ (data matrix or feature matrix)

adjacency matrix

- If G is a graph with vertices labelled $\{1,2,\ldots\ldots,n\}$, its **adjacency array A** is the *n* x *n matrix* whose ij-th entry is the number of edges (or the weight of the edge) joining vertex i and vertex j.

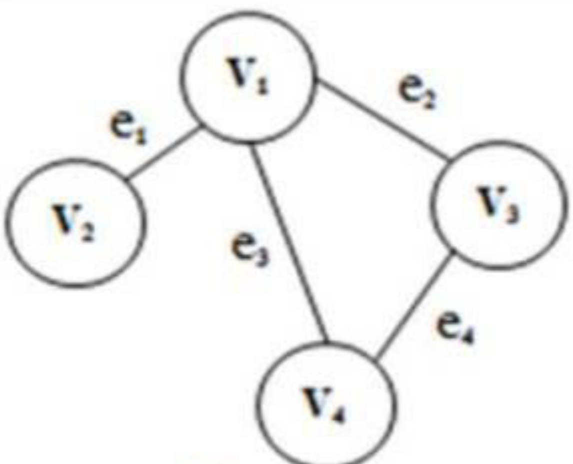


Adjacency matrix and graph presentations of different undirected and directed graphs.

Examples

Incidence matrix

- n nodes and m edges (
- Its **incidence array $\boldsymbol{M}$** is the *n* x *m matrix* whose ij*-th* entry is 1 if vertex i is incident to edge; and 0 otherwise.



(i)

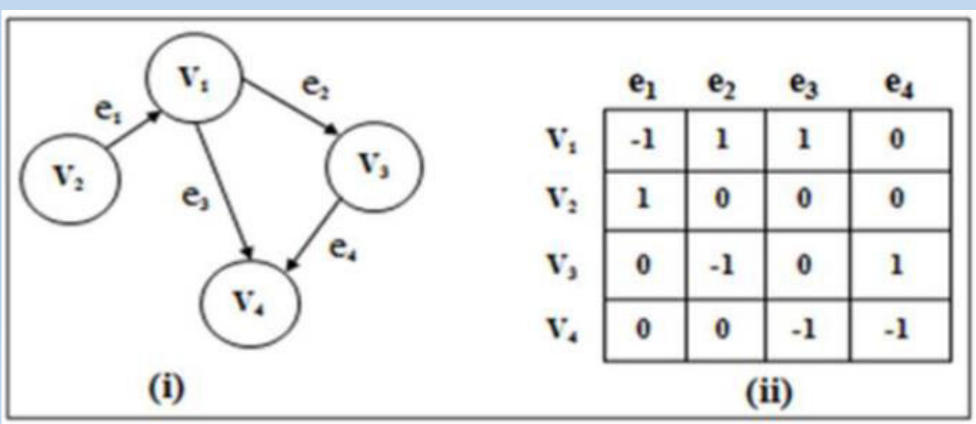
	e_1	e_2	e_3	e_4
V_1	1	1	1	0
V_2	1	0	0	0
V_3	0	1	0	1
V_4	0	0	1	1

(ii)

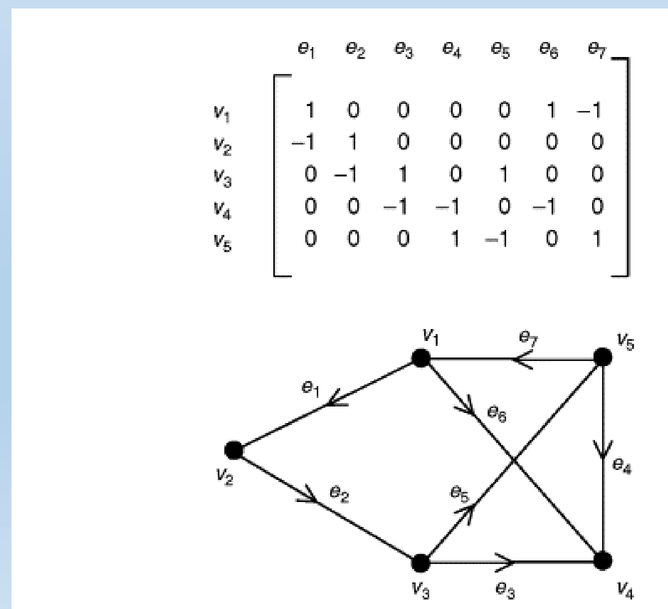
Incidence matrix

- n nodes and m edges
- Its **incidence array** $\boldsymbol{M}$ is a n × m matrix
- The ij-*th* entry of $\boldsymbol{M}$ is 1, the j-th edge leaves the i-th node.
- The ij-*th* entry of $\boldsymbol{M}$ is -1, the j-th edge entries the i-th node.

(Note that some authors use opposite sign notation)



	e1	e2	e3	e4
V1	-1	1	1	0
V2	1	0	0	0
V3	0	-1	0	1
V4	0	0	-1	-1



$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{c} \begin{array}{ccccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix} \end{array}$$

Special matrices

Square matrix : Number of rows equal to Number of columns

$$\begin{bmatrix} 1 & 4 & 0 \\ 8 & 15 & 3 \\ 1 & 9 & 2 \end{bmatrix} \qquad \begin{bmatrix} \mathbf{a}_{1,1} & \mathbf{a}_{1,2} & \mathbf{a}_{1,3} \\ \mathbf{a}_{2,1} & \mathbf{a}_{2,2} & \mathbf{a}_{2,3} \\ \mathbf{a}_{3,1} & \mathbf{a}_{3,2} & \mathbf{a}_{3,3} \end{bmatrix}$$

Zero matrix (denoted as $\mathbf{0}$ or $\emptyset$ **or** $\mathbf{0}_{mn}$ **or** $\emptyset_{mn}$

All elements are *equal to zero*

$$O_{2\,x\,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$O_{3x3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$O_{3x2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$O_{1x4} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

Identity matrix:(denoted as $\boldsymbol{I}_n$ or $\boldsymbol{I}$)

It is always square $m = n$.

Its diagonal elements equal to 1,

Other elements equal to 0

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Special matrices

Identity matrix:(denoted as $\boldsymbol{I}_n$ or $\boldsymbol{I}$)

$\boldsymbol{I}_n = [\boldsymbol{e}_1 \;\; \boldsymbol{e}_2 \;\; \dots \;\; \boldsymbol{e}_n]$

- The column vectors of the $n \times n$ identity matrix are the standard unit vectors of size n.
- Sometimes, the size is omitted and follows from the context.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 5 & 6 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Special matrices

Diagonal matrix:

It is always square $m = n$.

At least one diagonal element not equal to zero.

Other elements equal to 0

$$\begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$$

Diag$\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ or **Diag**$(1,2)$

The notation diag$(a_1, \dots, a_n)$ is used to compactly describe the $n \times n$ diagonal matrix

$$\boldsymbol{A} = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & a_n \end{pmatrix}$$

Special matrices

Triangular matrices.

- A square $n \times n$ matrix $\boldsymbol{A}$ is upper triangular if $a_{ij} = 0$ = 0 for i > j,

- It is lower triangular if $a_{ij} = 0$ = 0 for j > i.
- diagonal matrix refer to either lower or upper triangular.

$$\begin{bmatrix} 1 & -1 & 0.7 \\ 0 & 1.2 & -1.1 \\ 0 & 0 & 3.2 \end{bmatrix}, \quad \begin{bmatrix} -0.6 & 0 \\ -0.3 & 3.5 \end{bmatrix}$$

Special matrices

Symmetric Matrices

A $n \times n$ (square) matrix $\boldsymbol{A}$ is **Symmetric**

if $a_{ij} = a_{ji}$

$$\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 3 & 0 \\ 3 & 5 & 2 \\ 0 & 2 & -4 \end{bmatrix}$$

Transpose, addition, Scalar Multiplication

A $m \times n$ matrix $\boldsymbol{A}$, its transpose, denoted $\boldsymbol{A}^{\mathrm{T}}$ (or sometimes $\boldsymbol{A}'$), is a $n \times m$ matrix given by $\boldsymbol{A}_{ij} = [\boldsymbol{A}^{\mathrm{T}}]_{ji}$

- $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}$$

The i-th column vector of $\boldsymbol{A}$ becomes the i-th row vector of $\boldsymbol{A}^{\mathrm{T}}$.

The j-th row vector of $\boldsymbol{A}$ becomes the j-th column vector of $\boldsymbol{A}^{\mathrm{T}}$.

For Symmetric Matrices ($n \times n$ (square))

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$$

For Block Matrix

- $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}^T & \boldsymbol{C}^T \\ \boldsymbol{B}^T & \boldsymbol{D}^T \end{bmatrix}$

Transpose, addition, Scalar Multiplication

Matrix addition

Two matrices of the same size can be added together. The result is another matrix of the same size, obtained by adding the corresponding elements of the two matrices.

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}$$

Matrix subtraction is similar. As an example,

$$\begin{bmatrix} 1 & 6 \\ 9 & 3 \end{bmatrix} - \boldsymbol{I} = \begin{bmatrix} 0 & 6 \\ 9 & 2 \end{bmatrix}$$

Transpose, addition, Scalar Multiplication

Given $\boldsymbol{A} = \begin{bmatrix} 4 & -3 & 6 \\ -8 & 5 & -9 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} -5 & 6 & -2 \\ 3 & 7 & -4 \end{bmatrix}$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 4-5 & -3+6 & 6-2 \\ -8+3 & 5+7 & -9-4 \end{bmatrix} = \begin{bmatrix} -1 & 3 & 4 \\ -5 & 12 & -13 \end{bmatrix}$$

Given $\boldsymbol{A} = \begin{bmatrix} 6 & -7 \\ -4 & 5 \\ -3 & 2 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} -8 & 3 \\ 3 & -1 \\ 2 & -8 \end{bmatrix}$

$$\boldsymbol{A} - \boldsymbol{B} = \begin{bmatrix} 14 & -10 \\ -7 & 6 \\ -5 & 10 \end{bmatrix}$$

Transpose, addition, Scalar Multiplication

- Commutativity

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}.$$

- Associativity.

$$(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = \mathbf{A} + \mathbf{B} + \mathbf{C}$$

- Addition with zero matrix.

$$\boldsymbol{A} + \emptyset = \boldsymbol{A}$$

- Transpose of sum. $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$

Transpose, addition, Scalar Multiplication

Given $\boldsymbol{A} = \begin{bmatrix} 4 & -3 & 6 \\ -8 & 5 & -9 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} -5 & 6 & -2 \\ 3 & 7 & -4 \end{bmatrix}$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 4-5 & -3+6 & 6-2 \\ -8+3 & 5+7 & -9-4 \end{bmatrix} = \begin{bmatrix} -1 & 3 & 4 \\ -5 & 12 & -13 \end{bmatrix}$$

Given $\boldsymbol{A} = \begin{bmatrix} 6 & -7 \\ -4 & 5 \\ -3 & 2 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} -8 & 3 \\ 3 & -1 \\ 2 & -8 \end{bmatrix}$

$$\boldsymbol{A} - \boldsymbol{B} = \begin{bmatrix} 14 & -10 \\ -7 & 6 \\ -5 & 10 \end{bmatrix}$$

Transpose, addition, Scalar Multiplication

Given an square matrix $\boldsymbol{A}$, show that

$\boldsymbol{A} + \boldsymbol{A}^T$ is a symmetric matrix

Consider $(\boldsymbol{A} + \boldsymbol{A}^T)^{\boldsymbol{T}} = \boldsymbol{A}^T + \boldsymbol{A} = \boldsymbol{A} + \boldsymbol{A}^T$

That means $\boldsymbol{A} + \boldsymbol{A}^T$ is a symmetric matrix

Transpose, addition, Scalar Multiplication

Scalar-matrix multiplication

- Scalar multiplication of matrices is defined by multiplying every element of the matrix by the scalar. For example

$$(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}$$

- **Distributive law:** For all real numbers α and all matrices $\boldsymbol{A}, \boldsymbol{B}$ in $\mathbb{V}$, $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$
- **Associative law:** For all real numbers α, β and all matrices $\boldsymbol{A}$,

$$\alpha\beta\boldsymbol{A} = (\alpha\beta)\boldsymbol{A} = \alpha(\beta\boldsymbol{A}) = \beta(\alpha\boldsymbol{A})$$

Transpose, addition, Scalar Multiplication

The norm of an $m \times n$ matrix $\boldsymbol{A}$

$$\|\boldsymbol{A}\| = \sqrt{\sum_{i,j} a_{ij}^2}$$

Note that

$$\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$$

For example,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$\|\boldsymbol{A} + \boldsymbol{B}\| = \sqrt{10}$=3.16???

$\|\boldsymbol{A}\| = 2, \|\boldsymbol{B}\| = \sqrt{2}, \|\boldsymbol{A}\| + \|\boldsymbol{B}\| = 3.41???$

Can you prove $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$?

In many engineering problems, we use the inequality to prove the convergence of the methods, or the bound of our solution.

Matrix-vector Multiplication

An m-by-n ($m \times n$) matrix $\boldsymbol{A}$

An n-vector $\boldsymbol{x}$

The Matrix-vector Multiplication is

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

$\boldsymbol{y}$ is m-vector: $y_i = \sum_{j=1}^{n} a_{ij} x_j$

$$\begin{bmatrix} 0 & 2 & -1 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} (0)(2) + (2)(1) + (-1)(-1) \\ (-2)(2) + (1)(1) + (1)(-1) \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{a} & b \\ \mathrm{c} & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{bmatrix}$$

Matrix-vector Multiplication

Row and column interpretations: We can express the matrix-vector product in terms of the rows or columns of the matrix.

m-by-n $\boldsymbol{A}$, n-vector $\boldsymbol{x}$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are column vectors of $\boldsymbol{A}$

$\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_m$ are rows of $\boldsymbol{A}$ (Note that they are row vectors)

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$
$$y_i = \boldsymbol{b}_i \boldsymbol{x}$$

$$\begin{bmatrix} \mathrm{a} & b \\ \mathrm{c} & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{bmatrix}$$

Matrix-vector Multiplication

Row and column interpretations: We can express the matrix-vector product in terms of the rows or columns of the matrix.

m-by-n $\boldsymbol{A}$, n-vector $\boldsymbol{x}$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

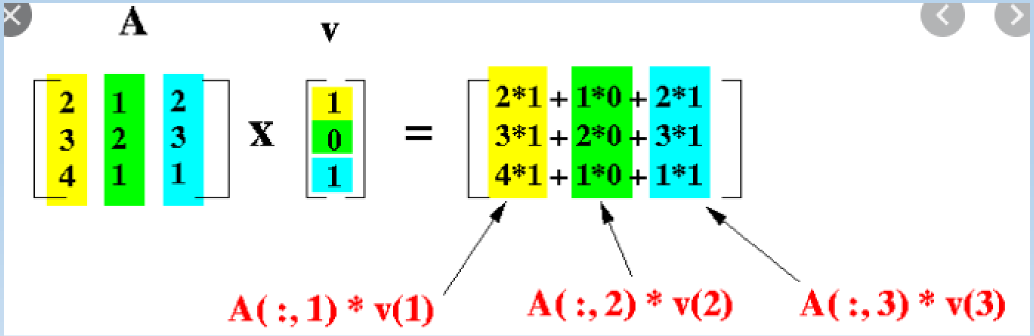
$\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are column vectors of $\boldsymbol{A}$

$\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_m$ are rows of $\boldsymbol{A}$ (Note that they are row vectors)

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

$$\boldsymbol{y} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

$\boldsymbol{y}$ is linear combination of column vectors of $\boldsymbol{A}$ and the coefficients are given by $\boldsymbol{x}$.



Matrix-vector Multiplication

Feature matrix and dissimilarity

Suppose $\boldsymbol{A}$ is a feature matrix, where its columns $\boldsymbol{a}_1, \boldsymbol{a}_2,\ldots, \boldsymbol{a}_n$are feature -$m$-vectors for n objects.

This is, m features and n objects.

Also the norm $\|\boldsymbol{a}_i\| = 1$ for all objects

Given $\boldsymbol{x}$ is the feature vector (m-vector) of a new object.$\|\boldsymbol{a}_i\| = 1$

$s_i = \boldsymbol{a}_i{}^T\boldsymbol{x}$ is a similarity measure between the i -object and the feature vector of the new object

All the similarity measures from the new object $\boldsymbol{x}$ to all data objects are given by

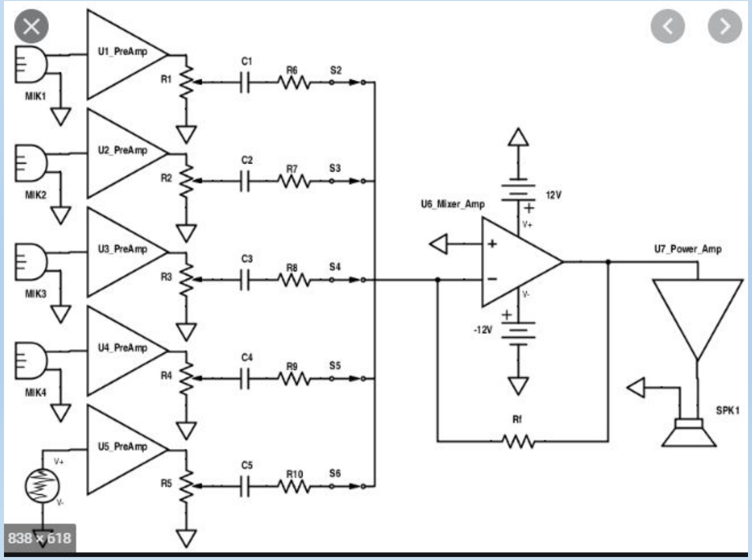
$\boldsymbol{s} = \boldsymbol{A}^T\boldsymbol{x}$

Matrix-vector Multiplication

Audio mixing

Suppose the n columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ of $\boldsymbol{A}$ are vectors representing audio signals or tracks of length m, and $\boldsymbol{w}$ is a n-vector.

Then $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{w}$ represents the mix of the audio signals, with track weights given by the vector $\boldsymbol{w}$.



Application Example

Feature matrix and weight vector.

Suppose $\boldsymbol{X}$ is a feature matrix, where its N columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are feature n -vectors for N objects or examples. Let the n-vector $\boldsymbol{w}$ be a weight vector, and let $s_i = \boldsymbol{x}_i^T \boldsymbol{w}$ be the score associated with object i using the weight vector $\boldsymbol{w}$. Then we can write $\boldsymbol{s} = \boldsymbol{X}^T \boldsymbol{w}$, where $\boldsymbol{s}$ is the N-vector of scores of the objects.

Portfolio return time series.

Suppose that $\boldsymbol{R}$ is a $T \times n$ asset return matrix, that gives the returns of n assets over T periods. A common trading strategy maintains constant investment weights given by the n -vector $\boldsymbol{w}$ over the T periods. For example, $w_4 = 0.15$ means that 15% of the total portfolio value is held in asset 4. Then $\boldsymbol{R}\boldsymbol{w}$, which is a T -vector, is the time series of the portfolio returns over the periods $1, \ldots, T$.

Application Example

Total price from multiple suppliers.

Suppose the $m \times n$ matrix $\boldsymbol{P}$ gives the prices of n goods from m suppliers. If $\boldsymbol{q}$ is an n -vector of quantities of the n goods (sometimes called a basket of goods), then $\boldsymbol{c} = \boldsymbol{P}\boldsymbol{q}$ an m -vector that gives the total cost of the goods, from each of the m suppliers.

Document scoring.

Suppose $\boldsymbol{A}$ in an $N \times n$ document-term matrix, which gives the word counts of a corpus of N documents using a dictionary of n words, so the row vectors of $\boldsymbol{A}$ are the word count vectors for the documents. Suppose that $\boldsymbol{w}$ in an n -vector that gives a set of weights for the words in the dictionary.

Then $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{w}$ is an N -vector that gives the scores of the documents, using the weights and the word counts. A search engine, for example, might choose $\boldsymbol{w}$ (based on the search query) so that the scores are predictions of relevance of the documents (to the search).

Matrix-Matrix Multiplication

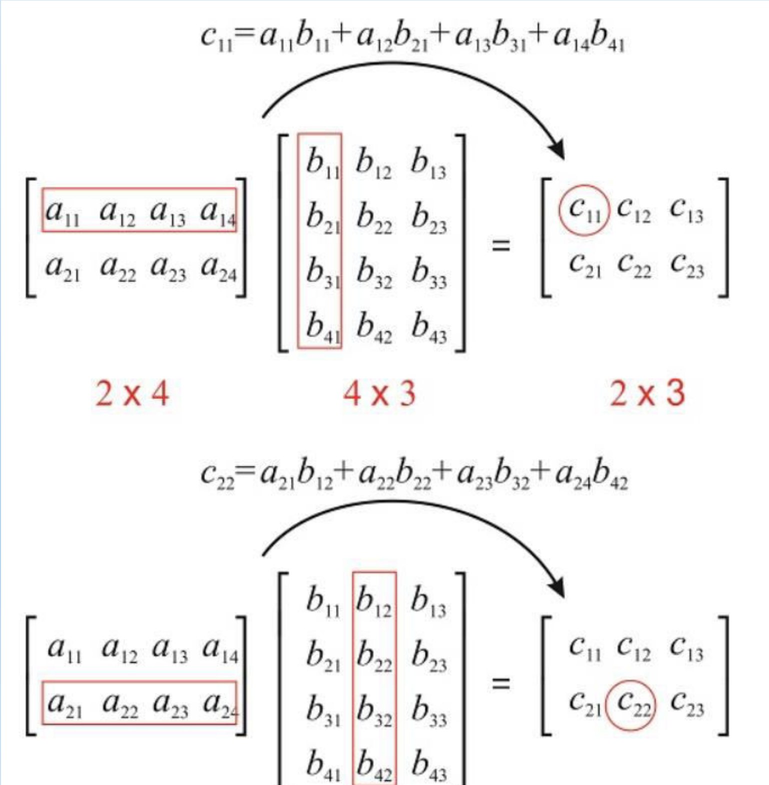
m-by-p $\boldsymbol{A}$, p-by-n $\boldsymbol{B}$

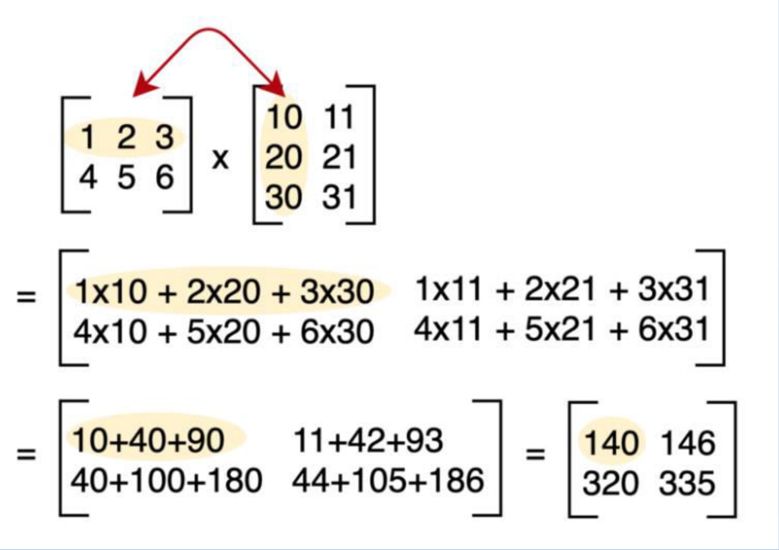
Then we have

$$\boldsymbol{C} = \boldsymbol{AB}$$

The ij element $\boldsymbol{C}$: $c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$

Note that : the number of columns of $\boldsymbol{A}$ must be equal to the number of rows of $\boldsymbol{B}$





Matrix-Matrix Multiplication

$$\begin{pmatrix} 1 & -2 & 4 \\ 5 & 0 & 3 \\ 0 & 1/2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 5 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & 4 \\ 5 & 0 & 3 \\ 0 & 1/2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 5 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \rightarrow a = 1 \times 1 + (-2) \times 5 + 4 \times (-1) = -13$$

$$\begin{pmatrix} 1 & -2 & 4 \\ 5 & 0 & 3 \\ 0 & 1/2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 5 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \rightarrow b = 1 \times 0 + (-2) \times 3 + 4 \times 0 = -6$$

$$\begin{pmatrix} 1 & -2 & 4 \\ 5 & 0 & 3 \\ 0 & 1/2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 5 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -13 & -6 \\ 2 & 0 \\ -13/2 & 3/2 \end{pmatrix}$$

Matrix-Matrix Multiplication

$$A = \begin{pmatrix} 6 & -2 & 3 \\ -4 & 2 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & -3 \\ 4 & -5 \\ 1 & -6 \end{pmatrix}$$

$$(A)(B) \Rightarrow \begin{pmatrix} 6\times 2 + -2\times 4 + 3\times 1 & 6\times -3 + -2\times -5 + 3\times -6 \\ -4\times 2 + 2\times 4 + 5\times 1 & -4\times -3 + 2\times -5 + 5\times -6 \end{pmatrix} \Rightarrow \begin{pmatrix} 7 & -26 \\ 5 & -28 \end{pmatrix}$$

Matrix-Matrix Multiplication

Some special case:

Matrix-vector multiplication.

m-by-n ($m \times n$) matrix $\boldsymbol{A}$, n-vector $\boldsymbol{x}$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

Vector Outer Product:

m-vector column $\boldsymbol{a}$, n-vector column $\boldsymbol{b}$

$$\text{An } m \times n \text{ matrix } \boldsymbol{C} = \boldsymbol{a}\boldsymbol{b}^{\boldsymbol{T}} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

Note that $\boldsymbol{a}\boldsymbol{b}^{\boldsymbol{T}} \neq \boldsymbol{b}\boldsymbol{a}^{\boldsymbol{T}}$

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 \end{bmatrix}.$$

Matrix-Matrix Multiplication

Matrix multiplication order matters

In general $\boldsymbol{AB} \neq \boldsymbol{BA}$ (Even $\boldsymbol{A}$ and $\boldsymbol{B}$ are with suitable sizes)

$$A = \begin{bmatrix} 1 & 6 \\ 9 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix}$$

$$AB = \begin{bmatrix} -6 & 11 \\ -3 & -3 \end{bmatrix}, \qquad BA = \begin{bmatrix} -9 & -3 \\ 17 & 0 \end{bmatrix}$$

Matrix multiplication with identity matrix

$\boldsymbol{AI = A, IA = A}$

$$M \times I = \begin{bmatrix} -4 & -3 \\ -6 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -4\times1+-3\times0 & -4\times0+-3\times1 \\ -6\times1+5\times0 & -6\times0+5\times1 \end{bmatrix}$$
$$= \begin{bmatrix} -4 & -3 \\ -6 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 3 & 1 & 2 \\ 4 & 5 & 1 & 7 \\ 2 & 9 & 3 & 6 \\ 8 & 7 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 & 2 \\ 4 & 5 & 1 & 7 \\ 2 & 9 & 3 & 6 \\ 8 & 7 & 5 & 1 \end{bmatrix}$$

Matrix-Matrix Multiplication

Properties

Associativity:

$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) = \boldsymbol{ABC}$

Note that compute $\boldsymbol{AB}$ first or $\boldsymbol{BC}$?

Different orders produce the same result but require different computational complexity

Associativity with scalar multiplication:

$$\alpha(\boldsymbol{AB}) = (\alpha\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(\alpha\boldsymbol{B})$$

Distributivity with addition:

$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$, Also $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$

=> $(\boldsymbol{A} + \boldsymbol{B})(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD} + \boldsymbol{BC} + \boldsymbol{BD}$

Transpose of product:

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{T}$$

Matrix-Matrix Multiplication

Let $\boldsymbol{A}$ and $\boldsymbol{P}$ be square matrices, let μ be a scalar and $\boldsymbol{P}$ is symmetric.

Show that $(\boldsymbol{APA}^T + \mu\boldsymbol{I})$ is symmetric.

$$(\boldsymbol{APA}^T + \mu\boldsymbol{I})^{\boldsymbol{T}} = (\boldsymbol{APA}^T)^{\boldsymbol{T}} + \mu\boldsymbol{I}^T = (\boldsymbol{APA}^T)^{\boldsymbol{T}} + \mu\boldsymbol{I}$$
$$= (\boldsymbol{A}^T)^{\boldsymbol{T}}(\boldsymbol{AP})^{\boldsymbol{T}} + \mu\boldsymbol{I} = \boldsymbol{AP}^T\boldsymbol{A}^T + \mu\boldsymbol{I} = \boldsymbol{APA}^T + \mu\boldsymbol{I}$$

Matrix-Matrix Multiplication

Matrix power: $\boldsymbol{A}$ is a square matrix

$\boldsymbol{AA}$ is denoted as $\boldsymbol{A}^2$

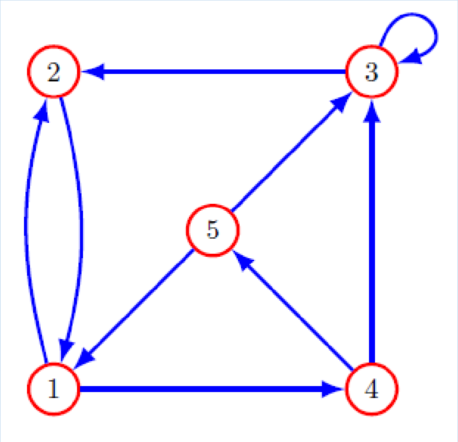
$\underbrace{\boldsymbol{AA}\ldots\boldsymbol{A}}_{l\ times}$ is denoted as $\boldsymbol{A}^l$

Adjacency matrix :

Suppose $\boldsymbol{A}$ is the $n \times n$ adjacency matrix of a directed graph with n vertices.

If there is an edge from node-i to node-j, $a_{ij} = 1$. (I follow the common notation not the notation of the textbook)

Otherwise, $a_{ij} = 0$. $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$



Matrix-Matrix Multiplication

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}, \boldsymbol{A}^2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & 1 & 0 \end{bmatrix}$$

$(\boldsymbol{A}^2)_{52} = \mathbf{1} \cdot 1 + \mathbf{0} \cdot 0 + \mathbf{1} \cdot 1 + \mathbf{0} \cdot 0 + \mathbf{0} \cdot \mathbf{0} =$ 2=>

There are 2 paths from node-5 to node-2 with length-2.

$\left(\boldsymbol{A}^l\right)_{ik}$: The number of paths with length l from node-i to node-k.

Why??

Matrix-Matrix Multiplication

Why??

By induction,

$a_{ij} = 1$, there is a path with length from node-i to node j. So it true for l=1.

Assume that

$\left(\boldsymbol{A}^l\right)_{ik}$: The number of paths with length l from node-i to node-k.

$\boldsymbol{A}^{l+1} = \boldsymbol{A}^l\boldsymbol{A},\quad \left(\boldsymbol{A}^{l+1}\right)_{ij} = \sum_k^n \left(\boldsymbol{A}^l\right)_{ik} a_{kj}$

a_{kj}: indicate that there a path from node-k to node-j. (1 or 0)

$\left(\boldsymbol{A}^l\right)_{ik} a_{kj} \neq 0$ indicates that there are $\left(\boldsymbol{A}^l\right)_{ij}$ paths (with length $l + 1$) from node-i to node-j through node-k

$\sum_k^n \left(\boldsymbol{A}^l\right)_{ik} a_{kj}$, the number of paths from node-i to node-j.

Back Home study $\boldsymbol{A}^3$.

What are the paths ? Ans: During computation, we need to some steps.

Matrix-Matrix Multiplication

Let $\boldsymbol{P}$ be a square matrix

(a) Show that $(\boldsymbol{P}-5\boldsymbol{I})(\boldsymbol{P}+\boldsymbol{I})=\boldsymbol{P}^2-4\boldsymbol{P}-5\boldsymbol{I}$,

(b) Given that $\boldsymbol{A}=\begin{pmatrix}1 & 2 & 2\\ 2 & 1 & 2\\ 2 & 2 & 1\end{pmatrix}$, verify that $\boldsymbol{A}^2-4\boldsymbol{A}-5\boldsymbol{I}=\emptyset$

(c) If $\boldsymbol{P}^2-4\boldsymbol{P}-5\boldsymbol{I}=\emptyset$, can we conclude that $\boldsymbol{P}=5\boldsymbol{I}$ or $\boldsymbol{P}=-\boldsymbol{I}$

(a) $(\boldsymbol{P}-5\boldsymbol{I})(\boldsymbol{P}+\boldsymbol{I})=\boldsymbol{P}^2-5\boldsymbol{P}+\boldsymbol{P}-5\boldsymbol{I}=\boldsymbol{P}^2-4\boldsymbol{P}-5\boldsymbol{I}$

(b) You can easily verify this.

(c) Cannot. from (b) $\boldsymbol{P}=\begin{pmatrix}1 & 2 & 2\\ 2 & 1 & 2\\ 2 & 2 & 1\end{pmatrix}$ ***is also a solution.***

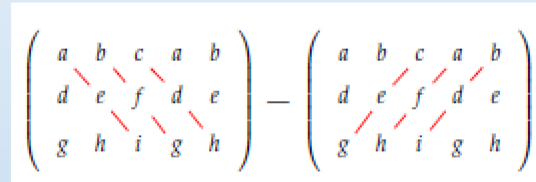
Determinant and Inverse

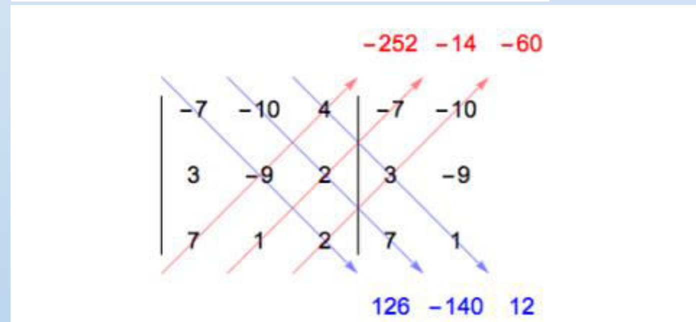
Determinant: a scalar value of a [square matrix]

Discussing the definition is quite time consuming, in this course, we show the way to compute this only.

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

$$\begin{aligned} |A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a\begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b\begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c\begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix} \\ &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$





$$\therefore \begin{vmatrix} -7 & -10 & 4 \\ 3 & -9 & 2 \\ 7 & 1 & 2 \end{vmatrix} = 126 + -140 + 12 - (-252 + -14 + -60) = 324$$

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} = a\begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - b\begin{vmatrix} e & g & h \\ i & k & l \\ m & o & p \end{vmatrix} + c\begin{vmatrix} e & f & h \\ i & j & l \\ m & n & p \end{vmatrix} - d\begin{vmatrix} e & f & g \\ i & j & k \\ m & n & o \end{vmatrix}.$$

Determinant and Inverse

The above computation is called Laplace expansion

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$|\boldsymbol{A}| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Determinant and Inverse

The above computation is called Laplace expansion

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

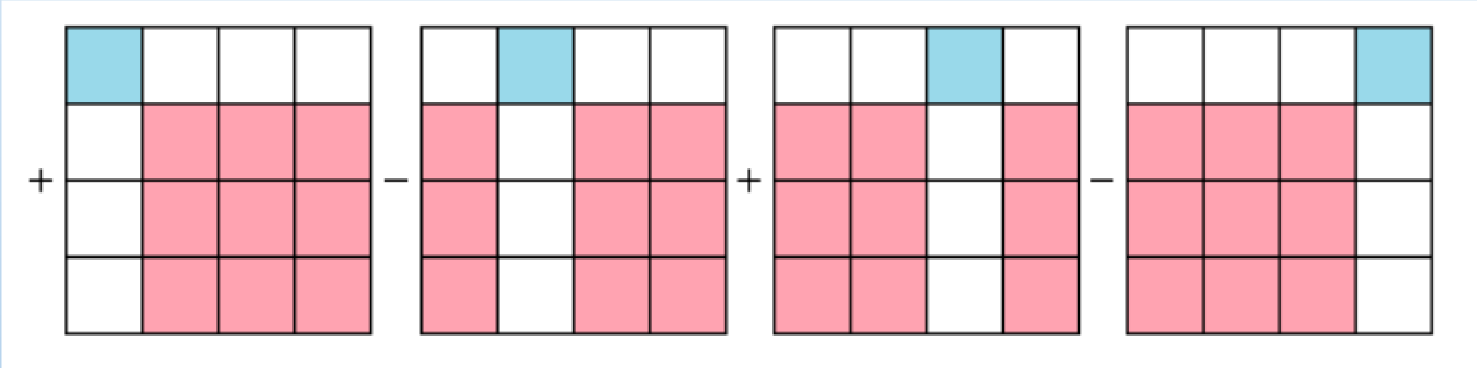
$$|\boldsymbol{A}| = -a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{23} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

Determinant and Inverse

For four by four, we use

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

$$A = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix}$$



$$\begin{aligned} \det(A) = & + \det[a_{00}] \cdot \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} - \det[a_{01}] \cdot \det \begin{bmatrix} a_{10} & a_{12} & a_{13} \\ a_{20} & a_{22} & a_{23} \\ a_{30} & a_{32} & a_{33} \end{bmatrix} \\ & + \det[a_{02}] \cdot \det \begin{bmatrix} a_{10} & a_{11} & a_{13} \\ a_{20} & a_{21} & a_{23} \\ a_{30} & a_{31} & a_{33} \end{bmatrix} - \det[a_{03}] \cdot \det \begin{bmatrix} a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \\ a_{30} & a_{31} & a_{32} \end{bmatrix} \end{aligned}$$

Determinant and Inverse

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 3 & 0 & 0 & 5 \\ 2 & 2 & 4 & -3 \\ 1 & 0 & 5 & 0 \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 3 & 0 & 0 & 5 \\ 2 & 2 & 4 & -3 \\ 1 & 0 & 5 & 0 \end{pmatrix}$$

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

$$\begin{vmatrix} 1 & 0 & 0 & -1 \\ 3 & 0 & 0 & 5 \\ 2 & 2 & 4 & -3 \\ 1 & 0 & 5 & 0 \end{vmatrix} = -2 \begin{vmatrix} 1 & 0 & -1 \\ 3 & 0 & 5 \\ 1 & 5 & 0 \end{vmatrix}$$

$$-2 \begin{vmatrix} 1 & 0 & -1 \\ 3 & 0 & 5 \\ 1 & 5 & 0 \end{vmatrix} = 10 \begin{vmatrix} 1 & -1 \\ 3 & 5 \end{vmatrix} = 80$$

Determinant and Inverse

$|\boldsymbol{I}| = 1$

$|\boldsymbol{A}^T| = |\boldsymbol{A}|$

$|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$

$|\alpha\boldsymbol{A}| = \alpha^n|\boldsymbol{A}|$

Example:

Let $|\boldsymbol{A}| = -\frac{\mathbf{1}}{\mathbf{5}}$, what is $|5\boldsymbol{A}|$? I do not know because I do not know the size of $\boldsymbol{A}$. For 3-by-3, $|5\boldsymbol{A}| = \mathbf{125}\left(-\frac{\mathbf{1}}{\mathbf{5}}\right)$

Determinant and Inverse

Only square matrix may have inverse. Only **square matrix $\boldsymbol{A}$ with $|\boldsymbol{A}|$ not equal to zero has inverse**.

Given a $\boldsymbol{A}$ if we have $\boldsymbol{B}$ such that

$\boldsymbol{AB} = \boldsymbol{I}$

$\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$. Also, $\boldsymbol{A}$ is the inverse of $\boldsymbol{B}$

The inverse of $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^{-1}$

Some identity

- $\boldsymbol{AA}^{-1} = \boldsymbol{I}$, $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$
- $(\boldsymbol{AC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}^{T})^{-1} = (\boldsymbol{A}^{T})^{-1}$
- $|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$

If inverse does not exist , we call singular

If inverse exists, we call non-singular

Determinant and Inverse

Only square matrix $\boldsymbol{A}$ with $|\boldsymbol{A}|$ not equal to zero has inverse.

Given a $\boldsymbol{A}$ if we have $\boldsymbol{B}$ such that

$$\boldsymbol{AB} = \boldsymbol{I}$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 4 \\ 0 & -2 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} -4 & 1 & -4 \\ 1/2 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$. Also, $\boldsymbol{A}$ is the inverse of $\boldsymbol{B}$

The inverse of $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^{-1}$

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Some identity

- $\boldsymbol{AA}^{-1} = \boldsymbol{I}$, $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$
- $(\boldsymbol{AC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}^{T})^{-1} = (\boldsymbol{A}^{T})^{-1}$

$$\begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 4 \\ 2 & 4 & 8 \end{bmatrix}$$

No inverse

Determinant and Inverse

How to find $\boldsymbol{A}^{-1}$

For 2-by-2,

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

$$\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}^{-1} = \frac{1}{4\times6-7\times2}\begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

$$= \frac{1}{10}\begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

For other, we use **Gaussian elimination or other methods**

Determinant and Inverse

Let $\boldsymbol{A} = \begin{pmatrix} 0 & 2 & 0 \\ -1 & 1 & 1 \\ 0 & -3 & 2 \end{pmatrix}$

(a) Evaluate $\boldsymbol{P}^3 - 3\boldsymbol{P}^2 + \boldsymbol{7P}$

(b) Find $\boldsymbol{P}^{-1}$

(a) $\boldsymbol{P}^3 - 3\boldsymbol{P}^2 + \boldsymbol{7P} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

(b) $\boldsymbol{P}(\boldsymbol{P}^2 - 3\boldsymbol{P} + \boldsymbol{7I}) = 4\boldsymbol{I}$

$\frac{1}{4}(\boldsymbol{P}^2 - 3\boldsymbol{P} + \boldsymbol{7I})$ is inverse of $\boldsymbol{P}$

Determinant and Inverse

Given $\boldsymbol{A}$, $\boldsymbol{B}$ non-singular matrix such that $\boldsymbol{A^2 = B}$,

Show that

(a) $\boldsymbol{AB = BA}$

(b) $\boldsymbol{AB^{-1} = A^{-1}}$

(c) $\boldsymbol{\left(AB^{-1} + BA^{-1}\right)^2 = B + B^{-1} + 2I}$

(a) $\boldsymbol{AB = A^3 = AAA = BA}$

(b) $\boldsymbol{AB^{-1} = AA^{-1}A^{-1} = A^{-1}}$

(c)

$$\boldsymbol{\left(AB^{-1} + BA^{-1}\right)\left(AB^{-1} + BA^{-1}\right)}$$

$$\boldsymbol{= AB^{-1}AB^{-1} + AB^{-1}BA^{-1} + BA^{-1}AB^{-1} + BA^{-1}BA^{-1}}$$

$$\boldsymbol{= B^{-1} + I + I + BA^{-1}BA^{-1}}$$

$$\boldsymbol{= B^{-1} + 2I + BA^{-1}BA^{-1} = B^{-1} + 2I + BAB^{-1}BAB^{-1}}$$

$$\boldsymbol{= B^{-1} + 2I + ABB^{-1}ABB^{-1} = B^{-1} + 2I + A^2 = B^{-1} + 2I + B}$$

Determinant and Inverse

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be an 2-by2 matrix,

(a) Show that if $\boldsymbol{A}^3 = \boldsymbol{I}$, then $|\boldsymbol{A}| = 1$

(b) Let $\boldsymbol{B}$ be an 2-by2 matrix with $\boldsymbol{B}^2 + \boldsymbol{B} + \boldsymbol{I} = \emptyset$

(i) Show that $\boldsymbol{B}^3 = \boldsymbol{I}$ and $\boldsymbol{B}^{-1} = -(\boldsymbol{B} + \boldsymbol{I})$

(ii) Simplify $\boldsymbol{I} + \boldsymbol{B} + \boldsymbol{B}^2 + \ \ldots + \boldsymbol{B}^{100}$

(a) $|\boldsymbol{A}^3| = |\boldsymbol{A}||\boldsymbol{A}||\boldsymbol{A}| = 1 \Rightarrow |\boldsymbol{A}| = 1$.

$\boldsymbol{A}^{-1}$

(b) $\boldsymbol{B}^2 + \boldsymbol{B} + \boldsymbol{I} = \emptyset \Rightarrow -\boldsymbol{B}(\boldsymbol{B} + \boldsymbol{I}) = \boldsymbol{I}$

$\Rightarrow -(\boldsymbol{B} + \boldsymbol{I})$ is an inverse of $\boldsymbol{B}$.

Also, $(\boldsymbol{B}^3 - \boldsymbol{I}) = (\boldsymbol{B} - \boldsymbol{I})(\boldsymbol{B}^2 + \boldsymbol{B} + \boldsymbol{I})$. $\boldsymbol{B}^2 + \boldsymbol{B} + \boldsymbol{I} = \emptyset \Rightarrow \boldsymbol{B}^3 = \boldsymbol{I}$

$$\boldsymbol{I} + \boldsymbol{B} + \boldsymbol{B}^2 + \ \ldots + \boldsymbol{B}^{100} = \emptyset + \boldsymbol{B}^3 + \boldsymbol{B}^4 + \ \ldots + \boldsymbol{B}^{100} = \boldsymbol{B}^3(\boldsymbol{I} + \boldsymbol{B} + \boldsymbol{B}^2) + \boldsymbol{B}^6 + \cdots + \boldsymbol{B}^{99} + \boldsymbol{B}^{100},$$

$$\ldots\ldots\ldots = \boldsymbol{B}^{99} + \boldsymbol{B}^{100} = \boldsymbol{B}^{33}(\boldsymbol{B} + \boldsymbol{I}) = (\boldsymbol{B} + \boldsymbol{I})$$

Linear System equations and Gaussian elimination

Consider a set (also called a system) of m linear equations in n variables or unknowns $x_1, x_2, \ldots, x_n$

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2, \ldots, a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2, \ldots, a_{2n}x_n = b_2 \\ \vdots \qquad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2, \ldots, a_{mn}x_n = b_m \end{array}$$

a_{ij}'s are coefficients, b_i can be right-hand sides. And they are known. Our aim is to find $x_1, x_2, \ldots, x_n$.

In matrix-vector form: $\boldsymbol{Ax} = \boldsymbol{b}$

A set of linear equations can have no solutions,

one solution, or multiple solutions.

Linear System equations and Gaussian elimination

Examples:

$x_1 + x_2 = 1, x_1 = -1, x_1 - x_2 = 0$

$$\boldsymbol{Ax} = \boldsymbol{b}$$

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

Why no solutions? When you put $x_1 = -1$ to other equations => contradiction. (Over-determined)

$x_1 + x_2 = 1, x_2 + x_3 = 1$

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

It has multiple solutions (over-determined), some solutions are

$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Note that some authors use $m > n, m < n, m = n$ to investigate and no solutions, one solution, or multiple solutions. It is not totally correct. We should use independent and rank to make the conclusion.

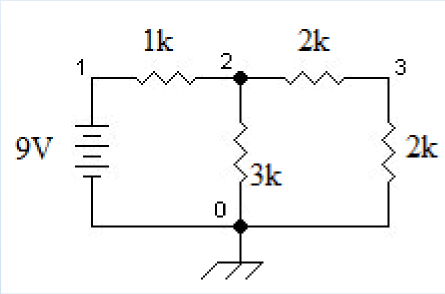
Linear System equations

Circuits:

$$9 = 1000 \times i_1 + 4000 \times i_2$$

$$9 = 1000 \times i_1 + 3000 \times i_3$$

$$0 = i_1 - i_2 - i_3$$



$$\begin{pmatrix} 1000 & 4000 & 0 \\ 1000 & 0 & 3000 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ i_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 9 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} i_1 \\ i_2 \\ i_3 \end{pmatrix} = \begin{pmatrix} 0.0033 \\ 0.0014 \\ 0.0019 \end{pmatrix}$$

Gaussian elimination (one solution)

Idea : make the matrix to be an upper triangular form.

If our system is

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$0x_1 \quad + a_{22}x_2 + a_{23}x_3 = b_2$$
$$0x_1 \quad + \quad 0x_2 + a_{33}x_3 = b_m$$

Then by back-substitution, we solve the problem.

How ? To make system to be an upper triangular form.

Linear System equations and Gaussian elimination

Consider a set (also called a system) of m linear equations in n variables or unknowns $x_1, x_2, \dots, x_n$

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2, \dots, a_{1n}x_1 = b_1 \\ a_{21}x_1 + a_{22}x_2, \dots, a_{2n}x_2 = b_2 \\ \vdots \qquad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2, \dots, a_{mn}x_2 = b_2 \end{array}$$

a_{ij}'s are coefficients, b_i can be right-hand sides. And they are known. Our aim is to find $x_1, x_2, \dots, x_n$.

In matrix-vector form: $\boldsymbol{Ax = b}$

Gaussian elimination (one solution)

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

If $a_{11} \neq 0$, we multiply the first equation with $-a_{21}/a_{11}$ add it to the second. (If zero, exchange the rows)

We multiply the first equation with $-a_{31}/a_{11}$ add it to the third.

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$0x_1 + \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right)x_2 + \left(a_{23} - \frac{a_{21}a_{13}}{a_{11}}\right)x_3 = b_2 - a_{21}b_1/a_{11}$$
$$0x_1 + \left(a_{32} - \frac{a_{31}a_{12}}{a_{11}}\right)x_2 + \left(a_{33} - \frac{a_{31}a_{13}}{a_{11}}\right)x_3 = b_3 - a_{31}b_1/a_{11}$$

Gaussian elimination (one solution)

Instead of performing the algorithm in equation format, we have use matrix format:

$$-3x_1 + 2x_2 - 1x_3 = -1$$
$$6\,x_1 \;-\; 6x_2 + 7x_3 = -7$$
$$3\,x_1 \;- 4x_2 + 4\,x_3 = -6$$

$$\begin{pmatrix} -3 & 2 & -1 & -1 \\ 6 & -6 & 7 & -7 \\ 3 & -4 & 4 & -6 \end{pmatrix}$$

we multiply the first row with 2 and add it to the second.

We multiply the first row with 1 and add it to the third.

$$\begin{pmatrix} -3 & 2 & -1 & -1 \\ 0 & -2 & 5 & -9 \\ 0 & -2 & 3 & -7 \end{pmatrix}$$

We multiply the second row with -1 and add it to the third.

$$\begin{pmatrix} -3 & 2 & -1 & -1 \\ 0 & -2 & 5 & -9 \\ 0 & 0 & -2 & 2 \end{pmatrix}.$$

Gaussian elimination (one solution)

$$\begin{pmatrix} -3 & 2 & -1 & -1 \\ 0 & -2 & 5 & -9 \\ 0 & 0 & -2 & 2 \end{pmatrix}.$$

$$\begin{aligned} -3x_1 + 2x_2 - x_3 &= -1, \\ -2x_2 + 5x_3 &= -9, \\ -2x_3 &= 2. \end{aligned}$$

Using back substitution, we get

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}$$

Gaussian elimination (one solution)

Example:

$$\begin{bmatrix} 25 & 5 & 1 \\ 64 & 8 & 1 \\ 144 & 12 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 106.8 \\ 177.2 \\ 279.2 \end{bmatrix}$$

$$\begin{bmatrix} 25 & 5 & 1 \\ 0 & -4.8 & -1.56 \\ 144 & 12 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 106.8 \\ -96.208 \\ 279.2 \end{bmatrix}$$

$$\begin{bmatrix} 25 & 5 & 1 \\ 0 & -4.8 & -1.56 \\ 0 & -16.8 & -4.76 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 106.8 \\ -96.208 \\ -335.968 \end{bmatrix}$$

$$\begin{bmatrix} 25 & 5 & 1 \\ 0 & -4.8 & -1.56 \\ 0 & 0 & 0.7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 106.8 \\ -96.208 \\ 0.76 \end{bmatrix}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0.290472 \\ 19.6905 \\ 1.08571 \end{bmatrix}$$

Gaussian elimination (one solution)

Sometimes, we have precision problem. There are great difference in the magnitudes of coefficients.

$$\begin{bmatrix} 20 & 15 & 10 \\ -3 & -2.249 & 7 \\ 5 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 45 \\ 1.751 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 20 & 15 & 10 \\ 0 & 0.001 & 8.5 \\ 0 & -2.75 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 45 \\ 8.501 \\ -2.25 \end{bmatrix}$$

$$\begin{bmatrix} 20 & 15 & 10 \\ 0 & 0.001 & 8.5 \\ 0 & 0 & 23375.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 45 \\ 8.501 \\ 23375.4 \end{bmatrix}$$

We need some special methods to solve it.

Gaussian elimination (for matrix inverse)

Due to time limit, we only discuss the following basic method,

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$$

We want to find $\boldsymbol{A}^{-1}$. How ?

Or saying find $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{I}$

Let $\boldsymbol{b}_j$ be the j-th column of $\boldsymbol{B}$.

The problem becomes n systems of equations

$$\boldsymbol{Ab}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \end{pmatrix}, \boldsymbol{Ab}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \dots, \boldsymbol{Ab}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Gaussian elimination (for matrix inverse)

$\boldsymbol{AB} = \boldsymbol{I}$

Write the equation in block matrix form :

$$\boldsymbol{A}[\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \dots \quad \boldsymbol{b}_n] = \left[\begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \quad \dots \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}\right]$$

The problem becomes n systems of equations

$$\boldsymbol{Ab}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \end{pmatrix}, \boldsymbol{Ab}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \dots, \boldsymbol{Ab}_n = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$$

Gaussian elimination (for matrix inverse)

$$\begin{pmatrix} -3 & 2 & -1 \\ 6 & -6 & 7 \\ 3 & -4 & 4 \end{pmatrix} \boldsymbol{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -3 & 2 & -1 \\ 6 & -6 & 7 \\ 3 & -4 & 4 \end{pmatrix} \boldsymbol{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -3 & 2 & -1 \\ 6 & -6 & 7 \\ 3 & -4 & 4 \end{pmatrix} \boldsymbol{b}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Gaussian elimination (for matrix inverse)

$$\left[\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \qquad (\tfrac{1}{2}\ \textit{row}\ 1 + \textit{row}\ 2)$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \qquad (\tfrac{2}{3}\ \textit{row}\ 2 + \textit{row}\ 3)$$

$$\begin{pmatrix} \text{Zero above} \\ \text{third pivot} \end{pmatrix} \to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \qquad (\tfrac{3}{4}\ \textit{row}\ 3 + \textit{row}\ 2)$$

$$\begin{pmatrix} \text{Zero above} \\ \text{second pivot} \end{pmatrix} \to \begin{bmatrix} 2 & 0 & 0 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \qquad (\tfrac{2}{3}\ \textit{row}\ 2 + \textit{row}\ 1)$$

$$\begin{array}{l} (\text{divide by } 2) \\ (\text{divide by } \frac{3}{2}) \\ (\text{divide by } \frac{4}{3}) \end{array} \quad \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right]$$

Gaussian elimination (for check linearly independent)

Given a set of vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$, they are **linearly dependent**, if

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 \dots + a_k\boldsymbol{x}_k = \boldsymbol{0}$$

for some $a_1, \dots, a_k$ where at least one of $a_1, \dots, a_k$ is non–zero.

Otherwise, independent.

Independent means that for all $\boldsymbol{x}_i$ cannot be a linear combination of others.

Note that "a few $\boldsymbol{x}_i$'s are linear combination of others" does not imply Independent.

Gaussian elimination (for check linearly independent)

$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 4 \\ 8 \end{bmatrix}$ **linearly independent**?

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + a_3\boldsymbol{x}_k = \boldsymbol{0}$$

$\boldsymbol{X}\boldsymbol{a} = \boldsymbol{0}$ has non zero solution ?

$$\begin{bmatrix} 1 & 3 & 0 \\ 2 & 2 & 4 \\ 3 & 1 & 8 \end{bmatrix} \boldsymbol{a} = \boldsymbol{0}$$

$$\left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 2 & 2 & 4 & 0 \\ 3 & 1 & 8 & 0 \end{array}\right] \xrightarrow[R_3:=R_3-3R_1]{R_2:=R_2-2R_1} \left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 0 & -4 & 4 & 0 \\ 0 & -8 & 8 & 0 \end{array}\right]$$

$$\xrightarrow{R_3:=R_3-2R_2} \left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The solution set is $-a_2 + a_3 = 0$

$$a_1 + 3a_2 = 0$$

For example, we can set $a_2 = a_3 = 0$, **and** $a_1 = 3$. **=> dependent**

$$3\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Gaussian elimination (for check linearly independent)

$\boldsymbol{x}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix}3\\2\\9\end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix}5\\2\\-1\end{bmatrix}$ **linearly independent**?

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + a_3\boldsymbol{x}_k = \boldsymbol{0}$$

$\boldsymbol{X}\boldsymbol{a} = \boldsymbol{0}$ has non zero solution ?

$$\begin{bmatrix}1 & 3 & 5\\2 & 2 & 2\\3 & 9 & -1\end{bmatrix}\boldsymbol{a} = \boldsymbol{0}$$

$$\left[\begin{array}{ccc|c}1 & 3 & 5 & 0\\2 & 2 & 2 & 0\\3 & 9 & -1 & 0\end{array}\right] \xrightarrow[R_3:=R_3-3R_1]{R_2:=R_2-2R_1} \left[\begin{array}{ccc|c}1 & 3 & 5 & 0\\0 & -4 & -8 & 0\\0 & 0 & -16 & 0\end{array}\right]$$

$$\xrightarrow[R_3:=-\frac{1}{16}R_3]{R_2:=-\frac{1}{4}R_2} \left[\begin{array}{ccc|c}1 & 3 & 5 & 0\\0 & 1 & 2 & 0\\0 & 0 & 1 & 0\end{array}\right]$$

Clearly, only zero vector is the solution. => independent

Example

Example:

Suppose we have a polynomial,

$$p(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$$

And we have four pairs of $p(-1.1) = b_1, p(-0.4) = b_2, p(0.2) = b_3,$ $p(0.8) = b_4$.

We need to find c_i's

These four pairs create the problem $\boldsymbol{Ac} = \boldsymbol{b}$

$$A = \begin{bmatrix} 1 & -1.1 & (-1.1)^2 & (-1.1)^3 \\ 1 & -0.4 & (-0.4)^2 & (-0.4)^3 \\ 1 & 0.2 & (0.2)^2 & (0.2)^3 \\ 1 & 0.8 & (0.8)^2 & (0.8)^3 \end{bmatrix}$$

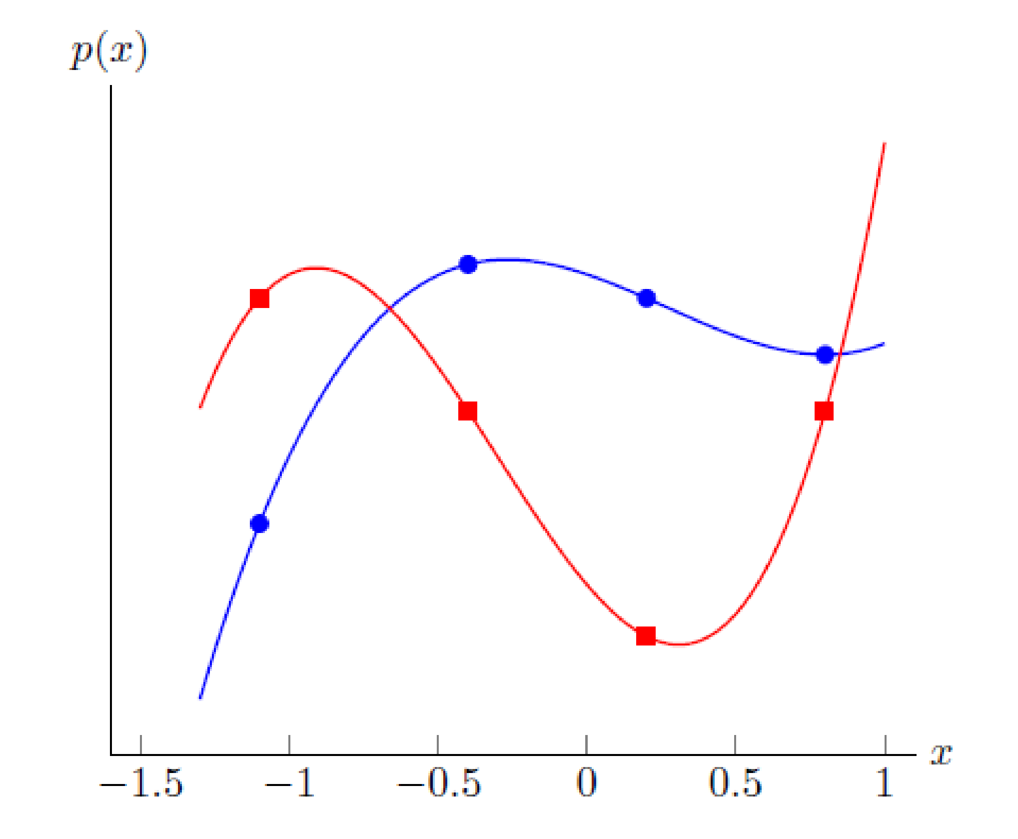
In a technical term, $x = -1.1$, , $x = -0.4$, …. … are inputs of an unknown system, and $b_i's$ are measurement outputs of the system.

Example

For a given $\boldsymbol{b}$, we can solve $\boldsymbol{c} = \boldsymbol{A}^{-1}\boldsymbol{b}$, and then create the polynomial.

For another $\boldsymbol{b}$, we can solve $\boldsymbol{c} = \boldsymbol{A}^{-1}\boldsymbol{b}$, and then create another polynomial.

$$A^{-1} = \begin{bmatrix} -0.5784 & 1.9841 & -2.1368 & 0.7310 \\ 0.3470 & 0.1984 & -1.4957 & 0.9503 \\ 0.1388 & -1.8651 & 1.6239 & 0.1023 \\ -0.0370 & 0.3492 & 0.7521 & -0.0643 \end{bmatrix}$$



Eigenvalues & eigenvectors

- Given a square matrix $\boldsymbol{A}$, a non-zero vector $\boldsymbol{x}$ is called an *eigenvector* of $\boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, *where* λ *is a scalar.*
- *The corresponding* λ is called *eigenvalue*
- *The normalized version* $\boldsymbol{v}$ *of the corresponding* $\boldsymbol{x}$ is

$$\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\boldsymbol{2}}}$$

- Exampe:
- $\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$
- $\lambda = 4$ and the corresponding normlaized eigenvector is $\frac{1}{\sqrt{13}}\begin{pmatrix} 3 \\ 2 \end{pmatrix}$

Eigenvalues & eigenvectors

$$\boldsymbol{Ax} = \lambda\boldsymbol{x} \Leftrightarrow (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$$

How to calculate $\boldsymbol{x}$ and λ:

Calculate the determinant $\det(\boldsymbol{A} - \lambda\boldsymbol{I})$ of $(\boldsymbol{A} - \lambda\boldsymbol{I})$, yields a polynomial of λ (degree n)

Determine roots to the polynomial, roots are eigenvalues λ

Solve $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$ for each λ to obtain eigenvectors $\boldsymbol{x}$ (or $\boldsymbol{v}$)

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

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0 \Rightarrow \lambda^2 = 1 \Rightarrow \lambda = 1 \; or \; -1.$$

For $\lambda_1 = 1$, $(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

For $\lambda_2 = -1$, $(\boldsymbol{A} + \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Eigenvalues & eigenvectors

- Eigenvalues

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \quad \begin{matrix} \lambda_1 = -5 \\ \lambda_2 = 2 \end{matrix}$$

- Determine eigenvectors: $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$

$$\begin{matrix} x_1 + 2x_2 = \lambda x_1 \\ 3x_1 - 4x_2 = \lambda x_2 \end{matrix} \quad \Rightarrow \quad \begin{matrix} (1-\lambda)x_1 + 2x_2 = 0 \\ 3x_1 - (4+\lambda)x_2 = 0 \end{matrix}$$

- Eigenvector for λ_1 = -5

$$\begin{matrix} 6x_1 + 2x_2 = 0 \\ 3x_1 + x_2 = 0 \end{matrix} \quad \Rightarrow \quad \mathbf{x}_1 = \begin{bmatrix} -0.3162 \\ 0.9487 \end{bmatrix} \text{or } \mathbf{x}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

- Eigenvector for λ_1 = 2

$$\begin{matrix} -x_1 + 2x_2 = 0 \\ 3x_1 - 6x_2 = 0 \end{matrix} \quad \Rightarrow \quad \mathbf{x}_2 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix} \text{or } \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Eigenvalues & eigenvectors

Given $\boldsymbol{A} = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}$:

(a) Find the eigenvalues λ_1, λ_2, where $\lambda_1 > \lambda_2$

4, and -1

Let the eigenvectors be $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

and $\boldsymbol{P} = (\boldsymbol{x}_1 \quad \boldsymbol{x}_2)$.

(b) What is $\boldsymbol{P^{-1}AP}$?

$$\boldsymbol{AP} = \boldsymbol{P}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\Rightarrow \boldsymbol{P^{-1}AP} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Eigenvalues & eigenvectors

Given $\boldsymbol{A} = \begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 2 & -9 \\ 0 & 5 \end{pmatrix}$

(a) Verify that the column vectors of $\boldsymbol{A}$ is eigenvector $\boldsymbol{B}$.

(b) Verify that $\boldsymbol{Y} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}$ is diagonal matrix

(a) $\boldsymbol{B}\boldsymbol{A} = (\boldsymbol{a}_1 \quad 5\boldsymbol{a}_2)$

(b) $\begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$

For this example, very interesting

$\boldsymbol{A}^{-1} = \boldsymbol{A}$

Eigenvector and orthogonal

For square matrix

$\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$

we call orthogonal matrix. Another way =>

$\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$

$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$.

The column vectors $\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$

$$\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n]$$

Since $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$,

$$\boldsymbol{Q}^T\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{q}_1 & \boldsymbol{q}_1^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_1^T\boldsymbol{q}_n \\ \boldsymbol{q}_2^T\boldsymbol{q}_1 & \boldsymbol{q}_2^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_2^T\boldsymbol{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_{\mathrm{n}}^T\boldsymbol{q}_1 & \boldsymbol{q}_{\mathrm{n}}^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_{\mathrm{n}}^T\boldsymbol{q}_n \end{bmatrix} = \boldsymbol{I}$$

$\Rightarrow \quad \boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$ are orthonormal.

Since $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$, the row vectors of $\boldsymbol{Q}$ are orthonormal

Eigenvectors of and orthogonal

For symmetric matrix $\boldsymbol{A}$, let $\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$ be the eigenvectors (normalized).

we have

$$\boldsymbol{A} = \boldsymbol{Q\lambda Q}^T$$

$$\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n]$$

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \vdots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

$\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_n$ are eigenvalues.

Proof

$$\boldsymbol{Q}^T\boldsymbol{AQ} = \boldsymbol{Q}^T\boldsymbol{A}[\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n] = \boldsymbol{Q}^T[\lambda_1\boldsymbol{q}_1 \quad \lambda_2\boldsymbol{q}_2 \quad \cdots \quad \lambda_n\boldsymbol{q}_n]$$

$$\boldsymbol{Q}^T\boldsymbol{AQ} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \vdots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \boldsymbol{\lambda} \Rightarrow \boldsymbol{QQ}^T\boldsymbol{AQQ}^T = \boldsymbol{Q\lambda Q}^T \Rightarrow \boldsymbol{A} = \boldsymbol{Q\lambda Q}^T$$

Eigenvectors of and orthogonal

Example

$$\boldsymbol{A} = \begin{pmatrix} 19 & 20 & -16 \\ 20 & 13 & 4 \\ -16 & 4 & 31 \end{pmatrix}$$

$$\boldsymbol{Q} = \tfrac{1}{9}\begin{pmatrix} -6 & 3 & -6 \\ -3 & 6 & -6 \\ 6 & 6 & 3 \end{pmatrix}$$

$$\boldsymbol{\lambda} = \begin{pmatrix} 45 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & -9 \end{pmatrix}$$

$$\boldsymbol{A} = \frac{1}{9}\begin{pmatrix} -6 & 3 & -6 \\ -3 & 6 & -6 \\ 6 & 6 & 3 \end{pmatrix}\begin{pmatrix} 45 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & -9 \end{pmatrix}\frac{1}{9}\begin{pmatrix} -6 & 3 & -6 \\ -3 & 6 & -6 \\ 6 & 6 & 3 \end{pmatrix}^T$$

$$\boldsymbol{A} = \frac{1}{81}\begin{pmatrix} -6 & 3 & -6 \\ -3 & 6 & -6 \\ 6 & 6 & 3 \end{pmatrix}\begin{pmatrix} 45 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & -9 \end{pmatrix}\begin{pmatrix} -6 & -3 & 6 \\ 3 & 6 & 6 \\ -6 & -6 & 3 \end{pmatrix}^T$$

Eigenvector and orthogonal

Also, given orthonormal column vectors $\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$

$$\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n]$$

$$\boldsymbol{Q}\boldsymbol{Q}^T = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{q}_1 & \boldsymbol{q}_1^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_1^T\boldsymbol{q}_n \\ \boldsymbol{q}_2^T\boldsymbol{q}_1 & \boldsymbol{q}_2^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_2^T\boldsymbol{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_\mathrm{n}^T\boldsymbol{q}_1 & \boldsymbol{q}_\mathrm{n}^T\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_\mathrm{n}^T\boldsymbol{q}_n \end{bmatrix} = \boldsymbol{I}$$

$\Rightarrow \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I} \Rightarrow (\boldsymbol{Q}^T\boldsymbol{Q})^T = \boldsymbol{I}^T \Rightarrow \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$

We can use orthonormal column vectors $\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$ to form a orthogonal matrix.

- $\boldsymbol{q}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\boldsymbol{q}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\boldsymbol{Q} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
- $\boldsymbol{q}_1 = \begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix}$, $\boldsymbol{q}_2 = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix}$, $\boldsymbol{Q} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$,

$$\boldsymbol{Q}\boldsymbol{Q}^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Eigenvector and orthogonal

There are many interpretations

Rotation a vector:

$\boldsymbol{A} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, given a vector $\boldsymbol{a}$

If we want to rotation $\boldsymbol{a}$ with counter clockwise

$\boldsymbol{b} = \boldsymbol{A}^T \boldsymbol{a}$

Or saying $\boldsymbol{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

If we want to rotation $\boldsymbol{a}$ with counter clockwise

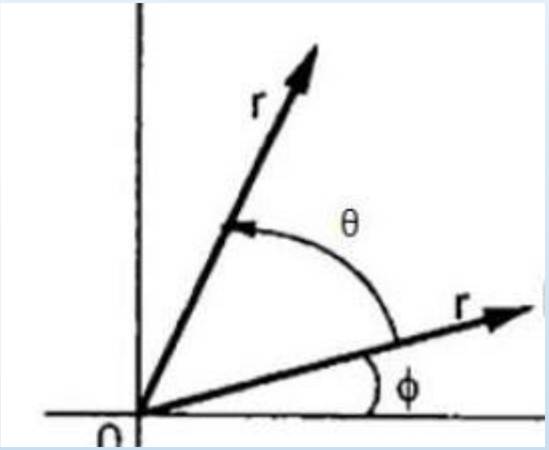
$\boldsymbol{b} = \boldsymbol{R}\boldsymbol{a}$

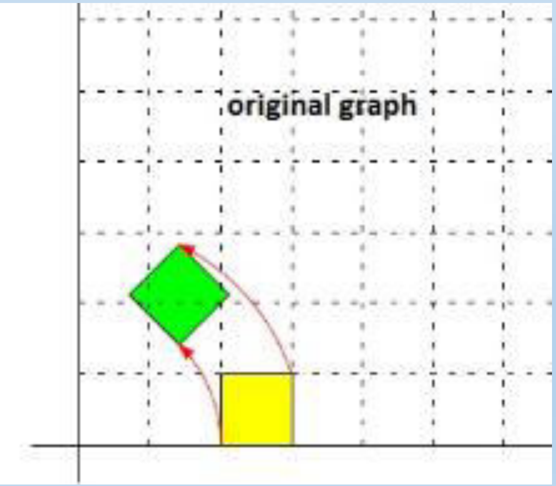
Note that "clockwise" use $\boldsymbol{A}$

Very useful for computer graphics

$\theta = \frac{\pi}{4}, \boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{b} = \begin{pmatrix} 0.7071 \\ 0.7071 \end{pmatrix}$

$\theta = \frac{\pi}{3}, \boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow \boldsymbol{b} = \begin{pmatrix} -1.231 \\ 1.8660 \end{pmatrix}$





Eigenvector and orthogonal

Coordinate Transform

Given a point p, its coordinate is $\boldsymbol{x}$ in the original coordinate system defined by the standard unit vectors, $\boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \cdots \quad \boldsymbol{e}_n$,

We can define a new coordinate system based on a set of orthonormal vectors $\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n$ to represent the point p.

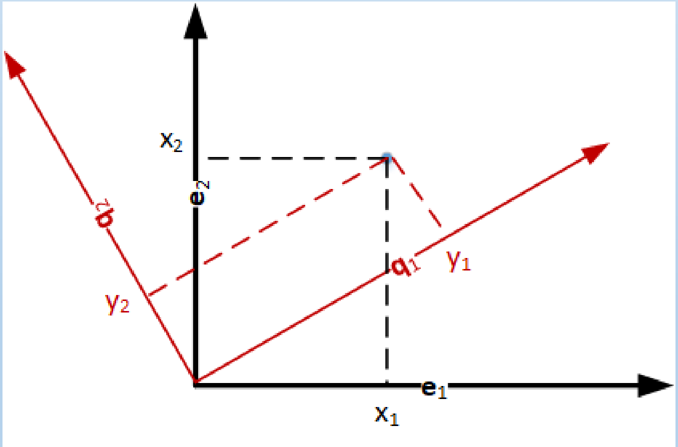
The coordinate of point p in new coordinate system is

$$\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x}$$

$\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_n]$.

Back the original coordinate system

$\mathbf{x} = \boldsymbol{Q}\boldsymbol{y}$



Eigenvector and orthogonal

Coordinate Transform

Example: $\boldsymbol{Q} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x} = \begin{pmatrix} 0.7071 \\ 0.7071 \end{pmatrix}$$

$$\boldsymbol{x} = \begin{pmatrix} -0.5 \\ 0.6 \end{pmatrix} \Rightarrow \boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x} = \begin{pmatrix} 0.0707 \\ -0.7778 \end{pmatrix}$$

Eigenvector and orthogonal

Let $\boldsymbol{Q} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \boldsymbol{P} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

(a) $\boldsymbol{P}^{-1}$? $= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

(b) Show that $\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P} = \begin{bmatrix} 3+\sin 2\theta & \cos 2\theta \\ \cos 2\theta & 3-\sin 2\theta \end{bmatrix}$

(c) If $\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P} = \begin{bmatrix} a & 0 \\ \mathbf{0} & b \end{bmatrix}$

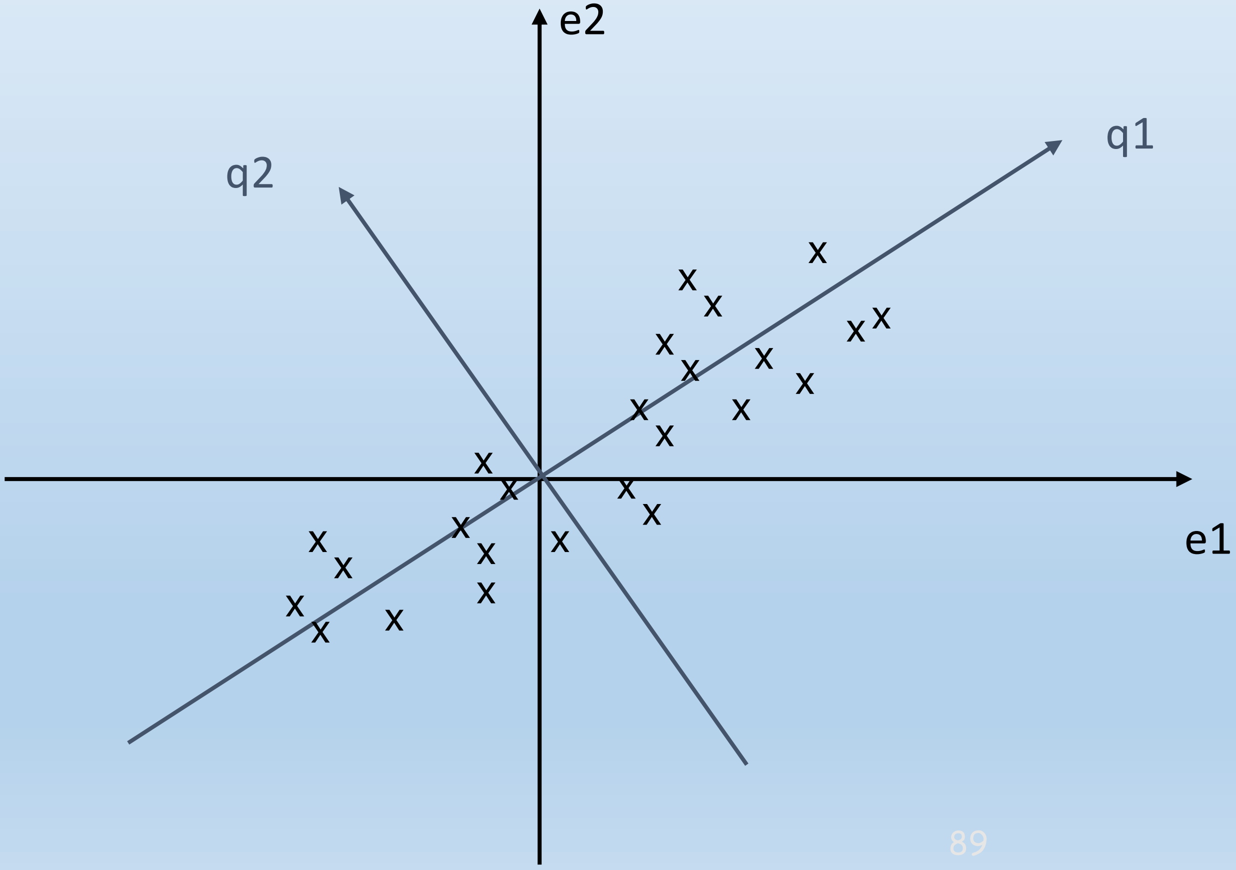
$What\,\theta, a, b\ ?$

$\theta = \frac{\pi}{4}$, **then easy**

Principal Component Analysis

May not teach

Summarization of high dimension data in a low dimension space.

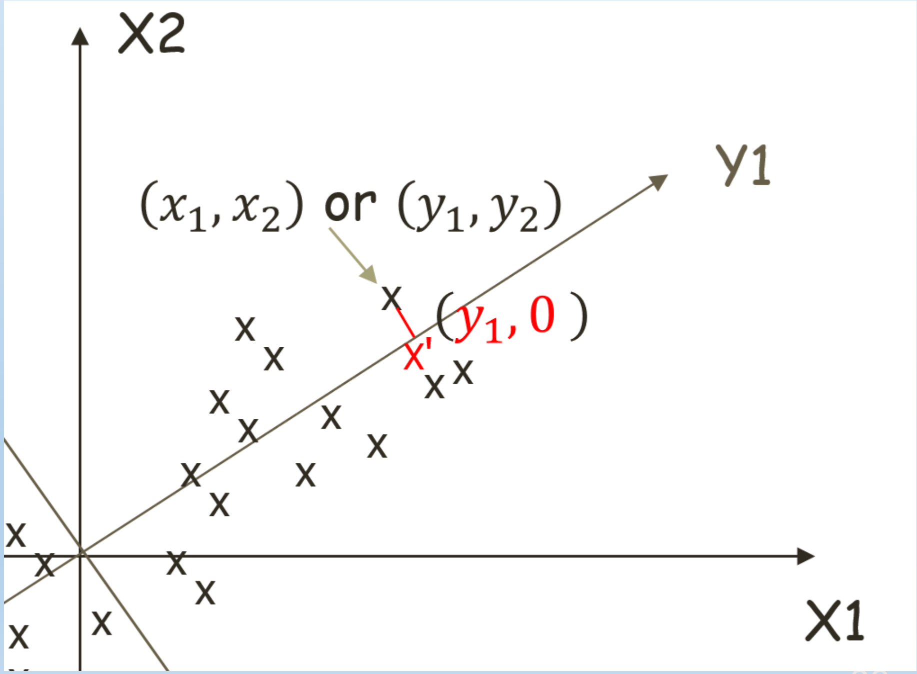


89

Principal Component Analysis

May not teach

- In the figure, a data point can be represented by two coordinate systems: (x_1, x_2) or (y_1, y_2).
- If we use y_1-axis only (store , there is small distortion, but we save the storage space.
- How to find the important axes (y_1**-axis and** y_2**-axis)** to represent the data?



Principal components

May not teach

- 1. principal component (PC1)
  - The eigenvalue with the largest absolute value will indicate that the data have the largest variance along its eigenvector, the direction along which there is greatest variation
- 2. principal component (PC2)
  - the direction with maximum variation left in data, orthogonal to the PC1
- In general, only few directions manage to capture most of the variability in the data.

Steps of PCA

May not teach

Given a data matrix $\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \dots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nN} \end{pmatrix} = (\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \dots \quad \boldsymbol{x}_N)$

n features and N sample vectors.

Let $\overline{\boldsymbol{x}}$ be the mean vector (taking the mean across the columns)
Adjust the original data by the mean vector

$$\overline{\overline{\boldsymbol{X}}} = (\boldsymbol{x}_1 - \overline{\boldsymbol{x}} \quad \boldsymbol{x}_2 - \overline{\boldsymbol{x}} \quad \dots \quad \boldsymbol{x}_N - \overline{\boldsymbol{x}})$$

compute the covariance matrix $\boldsymbol{C} = \frac{1}{N-1}\overline{\overline{\boldsymbol{X}}}\,\overline{\overline{\boldsymbol{X}}}^T$. Why $N-1$? Prof Wong will tell you later
Find the eigenvectors and eigenvalues of $\boldsymbol{C}$.

Afterwards, we have n eigenvalues $\lambda_1, \dots, \lambda_n$ and n eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$. Note that $\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = 0$ for $i \neq j$. $\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i = 1$.
The eigenvectors form an orthonormal set.

Transformed Data

May not teach

- Eigenvalues λ_i corresponds to variance on each component i.
- *Thus, sort by* λ_i
- Take the first p eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$ where $p < n$.
- These are the directions with the largest variances
- The transform matrix is given by $\boldsymbol{T} = \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_p^{\mathrm{T}} \end{pmatrix}$
- $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_p^{\mathrm{T}} \end{pmatrix} (\boldsymbol{x} - \overline{\boldsymbol{x}})$ for each data vector.
- Note that there are p elements in $\boldsymbol{y}$
- Now we store $\boldsymbol{y}$ (need p real numbers) rather than $\boldsymbol{x}$ (need n real numbers) If we have many data vectors, we save storage space

Inverse Transform

May not teach

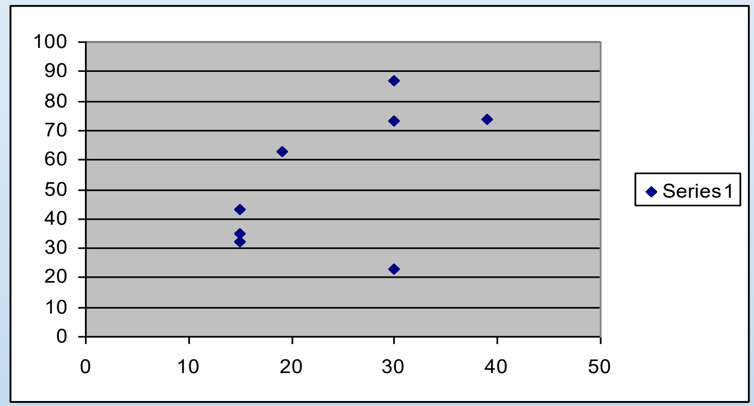
- How to restore the data,
- *Pick a stored* $\boldsymbol{y}$
- The reconstruction $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}$ is
- $\widehat{\boldsymbol{x}} = (\boldsymbol{e}_1 \quad \cdots \quad \boldsymbol{e}_p)\boldsymbol{y}+\overline{\boldsymbol{x}}$

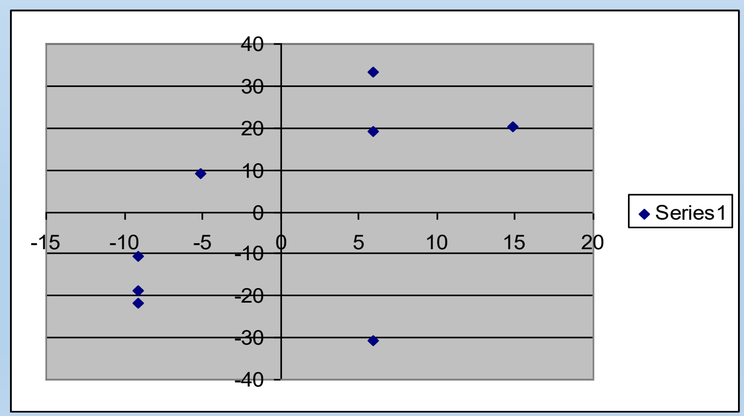
An Example

May not teach

Mean1=24.1
Mean2=53.8

k	x_{1k}	x_{2k}	x'_{1k}	x'_{2k}
1	19	63	-5.1	9.25
2	39	74	14.9	20.25
3	30	87	5.9	33.25
4	30	23	5.9	-30.75
5	15	35	-9.1	-18.75
6	15	43	-9.1	-10.75
7	15	32	-9.1	-21.75
8	30	73	5.9	19.25





95

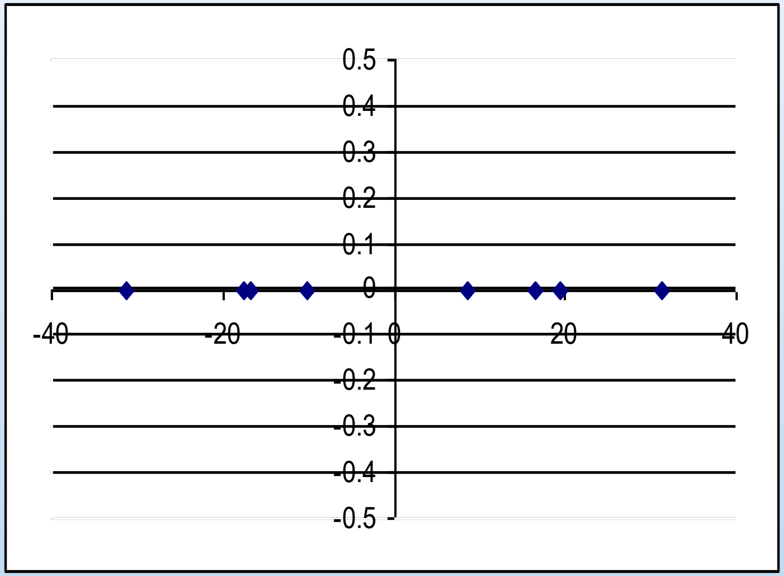
Covariance Matrix

May not teach

- We find out:
  - Eigenvectors:
  - $\boldsymbol{e}_1 = (0.21 \quad -0.98)^{\mathrm{T}}, \lambda_1 = 560.2$
  - $\boldsymbol{e}_2 = (-0.98 \quad -0.21)^{\mathrm{T}}, \lambda_2 = 51.8$

If we only keep one dimension: e_1

May not teach



- We keep the dimension of $\boldsymbol{e}_1 = (0.21 \quad -0.98)^{\mathrm{T}}$
- We can obtain the final data as
  - $y_k = (0.21 \quad -0.98)(\boldsymbol{x}_k - \overline{\boldsymbol{x}})$

y_k
-10.14
-16.72
-31.35
31.374
16.464
8.624
19.404
-17.63

Noise removal

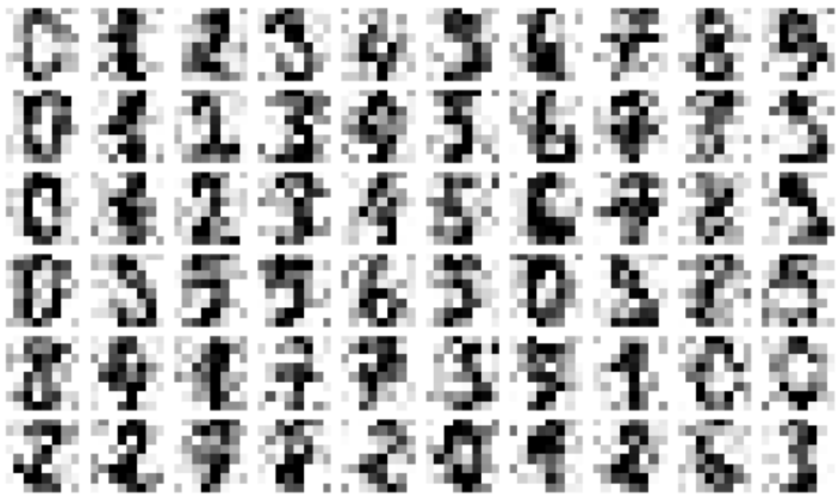
May not teach

PCA as Noise Filtering PCA can also be used as a filtering approach for noisy data. The idea is this: any components with variance much larger than the effect of the noise should be relatively unaffected by the noise. So if you reconstruct the data using just the largest subset of principal components, you should be preferentially keeping the signal and throwing out the noise. Let's see how this looks with the digits data. First we will plot several of the input noise-free data:

Noise removal

May not teach

```
np.random.seed(42)
noisy = np.random.normal(digits.data, 4)
plot_digits(noisy)
```



```
pca = PCA(0.70).fit(noisy)
pca.n_components_
```

2]: 26

Noise removal

May not teach

Now we compute these 26 components, and then use the inverse of the transform to reconstruct the filtered digits:

```
components = pca.transform(noisy)
filtered = pca.inverse_transform(components)
plot_digits(filtered)
```

This signal preserving/noise filtering property makes PCA a very useful feature selection routine—for example, rather than training a classifier on very high-dimensional data, you might instead train the classifier on the lower-dimensional representation, which will automatically serve to filter out random noise in the inputs.