

1 Basic Concept

1.1 Fundamental Operations of Matrix

If matrices A and B share the same numbers of column and row, then

$$(A \pm B)_{ij} = A_{ij} \pm B_{ij} \quad (1)$$

$$(AB)_{ij} = \sum_{r=1}^n A_{ir} B_{rj} \quad (2)$$

$$(A, I) \sim (I, A^{-1}) \quad (3)$$

$$(A_1 A_2 \cdots A_n)^T = A_n^T A_{n-1}^T \cdots A_2^T A_1^T \quad (4)$$

If A_1, A_2, \dots, A_k are all $n \times n$ matrices, then

$$(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_2^{-1} A_1^{-1} \quad (5)$$

If \mathbf{x} is the solution to equation $A\mathbf{x} = \mathbf{b}$ or $\mathbf{x}A = \mathbf{b}$, then

$$(A, \mathbf{b}) \sim (I, \mathbf{x}) \quad (6)$$

1.2 Matrix and Linear Equations

The following share the same solution:

$$\begin{cases} \text{Matrix equation } A\mathbf{x} = \mathbf{b} \\ \text{Vector equation } \sum_{k=1}^n x_k \mathbf{a}_k = \mathbf{b} \\ \text{Augmented matrix } [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n \ \mathbf{b}] \end{cases} \quad (7)$$

There are three cases for the solution:

$$\begin{cases} \text{Augmented matrix is not consistent} \rightarrow \mathbf{sol} = 0 \\ \text{No free variable} \rightarrow \mathbf{sol} = 1 \text{ (trivial solution)} \\ \text{Free variable exists} \rightarrow \mathbf{sol} = \text{infinite many} \end{cases} \quad (8)$$

1.3 Linearity

Theorem 1.3.1 If A is an $m \times n$ matrix, the following propositions are equivalent

- a. for every \mathbf{b} in \mathbb{R}^m , $A\mathbf{x} = \mathbf{b}$ has at least one solution
- b. $\mathbf{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathbb{R}^m$
- c. The columns of A generate \mathbb{R}^m
- d. A has a pivot position in every row

Theorem 1.3.2 If the columns of A are linear independent, then the matrix equation $A\mathbf{x} = \mathbf{0}$ only have trivial solution.

Theorem 1.3.3 If a mapping $x \mapsto T(x)$ is linear, then it is a closure.

2 Homogeneous Coordinate

2.1 Method

When translating a transform (compression, stretching, translation, rotation and perspective projection) of a graph into a matrix, we raise the coordinate to a higher dimension, i.e. $\mathbb{R}^n \rightarrow \mathbb{R}^{n+1}$. That is

$$(x_1, x_2, \dots, x_n) \rightarrow (x_1, x_2, \dots, x_n, 1) \quad (9)$$

Then we can use a $(n+1) \times (n+1)$ matrix A to represent the transform. **A can be considered as a set of the transform matrix applied on every basis of the homogeneous coordinate.**

2.2 Implementation

2.2.1 Compression and Stretching

$$(x_1, x_2, \dots, x_n, 1) \rightarrow (r_1 x_1, r_2 x_2, \dots, r_n x_n, 1), \quad A = \begin{bmatrix} r_1 & & & & \\ & r_2 & & & \\ & & r_3 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \quad (10)$$

2.2.2 Translation

If the translation vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$, then

$$(x_1, x_2, \dots, x_n, 1) \rightarrow (x_1 + p_1, x_2 + p_2, \dots, x_n + p_n, 1), \quad A = \begin{bmatrix} 1 & & & & p_1 \\ & 1 & & & p_2 \\ & & 1 & & p_3 \\ & & & \ddots & \vdots \\ & & & & 1 & p_n \\ & & & & & 1 \end{bmatrix} \quad (11)$$

2.2.3 Rotation

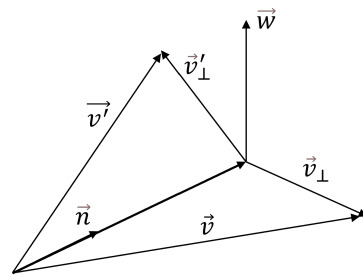
The axis around which a point (x, y, z) rotates is denoted as \mathbf{n} , where

$$\mathbf{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}, \quad n_x^2 + n_y^2 + n_z^2 = 1 \quad (12)$$

$$\begin{aligned}\omega &= \mathbf{v}_{\perp} \times \mathbf{n} \\ &= \mathbf{v} \times \mathbf{n}\end{aligned}$$

$$\mathbf{v}_{\parallel} = (\mathbf{v} \cdot \mathbf{n})\mathbf{n}, \quad \mathbf{v}_{\parallel} + \mathbf{v}_{\perp} = \mathbf{v}$$

$$\begin{aligned}\mathbf{v}'_{\perp} &= \mathbf{v}_{\perp} \cos \theta + \omega \sin \theta = (\mathbf{v} - (\mathbf{v} \cdot \mathbf{n})\mathbf{n}) \cos \theta + (\mathbf{v} \times \mathbf{n}) \sin \theta \\ \mathbf{v}' &= \mathbf{v}'_{\perp} + \mathbf{v}'_{\parallel} = (\mathbf{v} \times \mathbf{n}) \sin \theta + (\mathbf{v} - (\mathbf{v} \cdot \mathbf{n})\mathbf{n}) \cos \theta + (\mathbf{v} \cdot \mathbf{n})\mathbf{n}\end{aligned}$$



Substitute in we obtain

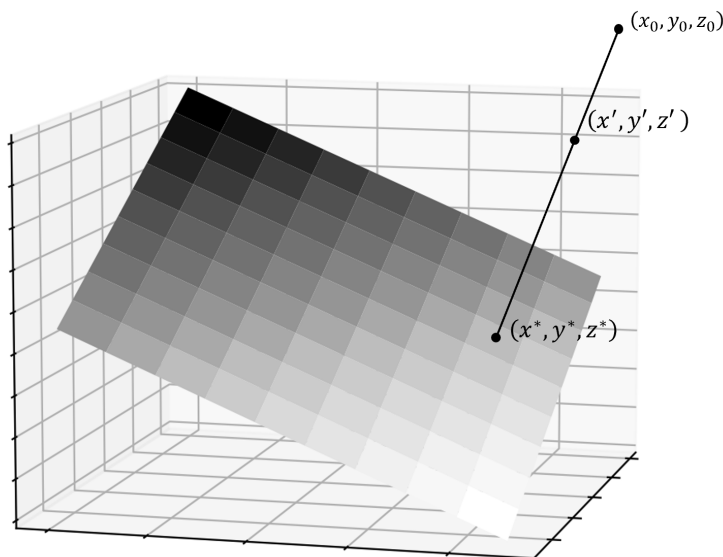
$$\begin{aligned}\mathbf{v}' &= \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \sin \theta \begin{bmatrix} n_z y - n_y z \\ n_x z - n_z x \\ n_y x - n_x y \end{bmatrix} + \cos \theta \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} (n_x x + n_y y + n_z z) \right) + \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} (n_x x + n_y y + n_z z) \\ &= \begin{bmatrix} (n_x^2(1 - \cos \theta) + \cos \theta)x + (n_x n_y(1 - \cos \theta) + n_z \sin \theta)y + (n_x n_z(1 - \cos \theta) - n_y \sin \theta)z \\ (n_y n_z(1 - \cos \theta) - n_z \sin \theta)x + (n_y^2(1 - \cos \theta) + \cos \theta)y + (n_z n_y(1 - \cos \theta) + n_x \sin \theta)z \\ (n_x n_z(1 - \cos \theta) + n_y \sin \theta)x + (n_y n_z(1 - \cos \theta) - n_x \sin \theta)y + (n_z^2(1 - \cos \theta) + \cos \theta)z \end{bmatrix}\end{aligned}$$

Then the transform matrix is

$$A = \begin{bmatrix} n_x^2(1 - \cos \theta) + \cos \theta & n_x n_y(1 - \cos \theta) + n_z \sin \theta & n_x n_z(1 - \cos \theta) - n_y \sin \theta \\ n_y n_z(1 - \cos \theta) - n_z \sin \theta & n_y^2(1 - \cos \theta) + \cos \theta & n_z n_y(1 - \cos \theta) + n_x \sin \theta \\ n_x n_z(1 - \cos \theta) + n_y \sin \theta & n_y n_z(1 - \cos \theta) - n_x \sin \theta & n_z^2(1 - \cos \theta) + \cos \theta \end{bmatrix} \quad (13)$$

2.2.4 Perspective Projection

We are interested in the matrix representing a projection with an arbitrary point (x_0, y_0, z_0) as the observation point to an arbitrary plane.



Equation of line:

$$\frac{x - x_0}{x' - x_0} = \frac{y - y_0}{y' - y_0} = \frac{z - z_0}{z' - z_0}$$

Equation of plane:

$$Ax + By + Cz + D = 0$$

$$\rightarrow \begin{cases} Ax^* + By^* + Cz^* + D = 0 \\ x^* = (x' - x_0)k + x_0 \\ y^* = (y' - y_0)k + y_0 \\ z^* = (z' - z_0)k + z_0 \end{cases}$$

(14)

Solving (14) we obtain

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} x^* \\ y^* \\ z^* \\ 1 \end{bmatrix} = \begin{bmatrix} -(By_0 + Cz_0 + D)x' + Bx_0y' + Cx_0z' + Dx_0 \\ Ay_0x' - (Ax_0 + Cz_0 + D)y' + Cy_0z' + Dy_0 \\ Az_0x' + Bz_0y' - (Ax_0 + By_0 + D)z' + Dz_0 \\ Ax' + By' + Cz' - (Ax_0 + By_0 + Cz_0) \end{bmatrix} \quad (15)$$

Hence the projection matrix is

$$A_{\text{projection}} = \begin{bmatrix} -By_0 - Cz_0 - D & Bx_0 & Cx_0 & Dx_0 \\ Ay_0 & -Ax_0 - Cz_0 - D & Cy_0 & Dy_0 \\ Az_0 & Bz_0 & -Ax_0 - By_0 - D & Dz_0 \\ A & B & C & -Ax_0 - By_0 - Cz_0 \end{bmatrix} \quad (16)$$

3 Gaussian Elimination

Definition The Gaussian Elimination is a manipulation applied on a definite matrix. It contains the following operations:

$$\left\{ \begin{array}{l} \textbf{Multiplying:} \text{ Multiplying a row by a non-zero number} \\ \textbf{Adding:} \text{ Adding the multiple of a row to another one} \\ \textbf{Swapping:} \text{ Exchanging the position of two definite rows} \end{array} \right. \quad (17)$$

The result of applying the Gaussian Elimination is generating a row echelon form, denoted as matrix U . A and U are row equivalent, i.e. $A \sim U$. For instance:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \xrightarrow{\text{Gaussian Elimination}} \begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ & a'_{22} & \cdots & a'_{2n} \\ & & \ddots & \vdots \\ & & & a'_{mn} \end{bmatrix} \quad (18)$$

If in the transformation, we only use the **Adding** operation, which can be represented as E_i at the i th step, then

$$E_p E_{p-1} \cdots E_1 A = U, \text{ or } L = (E_p E_{p-1} \cdots E_1)^{-1}, A = LU \quad (19)$$

Where L is an $m \times m$ unit lower triangular matrix, while U is an $m \times n$ upper triangular matrix equivalent to A . We can conclude that

$$(A, L) \sim (U, I) \quad (20)$$

4 Inverse of Matrix

For an $n \times n$ matrix, its inverse is defined as

$$A^{-1}: A^{-1}A = AA^{-1} = I_n \quad (21)$$

4.1 Criterion

The invertibility of an arbitrary matrix can be judged by the following features:

The columns of A are linearly independent \Leftrightarrow The columns of A is a basis of $\mathbb{R}^n \Leftrightarrow \det A \neq 0$

4.2 Proof

If the columns of A are linear dependent,

$$\exists a_i \in A, \exists \{c_1, c_2, \dots, c_{i-1}\}, a_i = \sum_{k=1}^{i-1} c_k a_k$$

$$A^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_i & \cdots & a_n \end{bmatrix}^T \sim \begin{bmatrix} a_1 & a_2 & \cdots & \mathbf{0} & \cdots & a_n \end{bmatrix}^T$$

Then

$$\det A = \det A^T = \det \begin{bmatrix} a_1 & a_2 & \cdots & \mathbf{0} & \cdots & a_n \end{bmatrix}^T = 0$$

And vice versa, QED

5 Dimensions and Rank

5.1 Basis

Definition Given that $\mathcal{B} = \{b_1, b_2, \dots, b_p\}$ is a set of basis to linear space H , then for every vector x in H , there is a unique coordinate to represent it:

$$[x]_{\mathcal{B}} = [c_1, c_2, \dots, c_p]^T, \quad x = [\mathcal{B}][x]_{\mathcal{B}} \quad (22)$$

A set of basis should be linear independent, i.e. equation $[\mathcal{B}]x = \mathbf{0}$ has only trivial solution. **Particularly, the rows of an echelon form are linear independent**

Theorem 5.1.1 Given that A is a matrix and its echelon form is U , the non-zero rows of U are the basis of **Row** A , while the corresponding columns to the pivot columns of U in A are the basis of **Col** A .

5.2 Dimension

Definition The dimension of a linear space V is the number of vectors its arbitrary basis contains, denoted as **dim** V .

$$\mathcal{B} = \{b_1, b_2, \dots, b_p\}, \quad b_1, b_2, \dots, b_p \in V \rightarrow \dim V = p$$

For finite dimensional space, p is an integer. Particularly, for an infinite dimensional space V , $\dim V = \infty$.

5.3 Rank

The rank of an $m \times n$ matrix A is the dimension of its column space, denoted as $\text{rank}A$.

$$\text{rank}A + \dim \text{Nul}A = n \quad (23)$$

6 Inverse of Block Matrix

We are interested in the inverse of a 2×2 block matrix T , denoted as

$$T = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad (24)$$

Assume that the left-upper block \mathbf{A}_{11} is invertible, then denote the inverse matrix as T^{-1} , where

$$T^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \quad (25)$$

From the definition $TT^{-1} = \mathbf{I}$, we have

$$\begin{cases} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} = \mathbf{I} \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I} \end{cases} \quad (26)$$

We take steps to solve the equations system.

$$\begin{cases} \mathbf{B}_{12} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I} \end{cases} \rightarrow \mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} = \mathbf{F}, \quad \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{22}\mathbf{F}$$

$$\begin{cases} \mathbf{B}_{11} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{21} = \mathbf{A}_{11}^{-1} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} = \mathbf{0} \end{cases} \rightarrow \begin{aligned} \mathbf{B}_{21} &= -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} = -\mathbf{F}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\ \mathbf{B}_{11} &= \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{21} = \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}) \end{aligned}$$

Hence we can conclude that the inverse matrix is

$$T^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{22}\mathbf{F} \\ -\mathbf{F}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F} \end{bmatrix} \quad \mathbf{F} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \quad (27)$$

7 Cramer's Rule

7.1 Content

If the linear equation $Ax = c$ has solution vector $x = (x_1, x_2, \dots, x_n)^T$. Then the solution x_i is

$$x_i = \frac{\det A_i(c)}{\det A} \quad (28)$$

where A_i is the resultant matrix by replacing the i th column of the original matrix by column vector c .

7.2 Proof

Given that $Ax = c$, then

$$A^{-1}Ax = A^{-1}c = Ix = x \quad (29)$$

We are interested in $\det A_i$, that is

$$\begin{aligned}
 \det A_i(c) &= \det(a_1, a_2, \dots, a_{i-1}, c, a_{i+1}, \dots, a_{n-1}, a_n) \\
 &= \det(a_1, a_2, \dots, a_{i-1}, \sum_{k=1}^n x_k a_k, a_{i+1}, \dots, a_{n-1}, a_n) \\
 &= \sum_{k=1}^n x_k \det(a_1, a_2, \dots, a_{i-1}, a_k, a_{i+1}, \dots, a_{n-1}, a_n)
 \end{aligned} \tag{30}$$

With the alternative feature of determinant, we have

$$\det(u_1, u_2, \dots, u_i, \dots, u_j, \dots, u_n) = -\det(u_1, u_2, \dots, u_j, \dots, u_i, \dots, u_n) \tag{31}$$

$$\exists i, j, i \neq j, u_i = u_j, u_i, u_j \in A \rightarrow \det A = 0 \tag{32}$$

Then the equation in (6) gives

$$\begin{aligned}
 \sum_{k=1}^n x_k \det(a_1, a_2, \dots, a_{i-1}, a_k, a_{i+1}, \dots, a_{n-1}, a_n) &= x_i \det(a_1, a_2, \dots, a_n) \\
 \det A_i(c) &= x_i \det A \quad \text{QED}
 \end{aligned} \tag{33}$$

7.3 Application

A significant application of Cramer's Rule is to give a simplified calculation for the inverse to a square matrix. To find an inverse, we may consider the following matrix equation.

$$AA^{-1} = I, \quad \text{or} \quad \bigoplus_{j=1}^n AA_j^{-1} = \bigoplus_{j=1}^n e_j \tag{34}$$

Where A_j^{-1} and I_j represents the j th column of A^{-1} and I respectively. Then we treat every column of A^{-1} as the solution to equation

$$Ax = e_j \tag{35}$$

Applying Cramer's rule,

$$\begin{aligned}
 x_i &= \frac{\det A_i(e_j)}{\det A} \\
 &= \frac{1}{\det A} \cdot \det \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,i-1} & 0 & a_{1,i+1} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{j,1} & a_{j,2} & \cdots & a_{j,i-1} & 1 & a_{j,i+1} & \cdots & a_{j,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,i-1} & 0 & a_{n,i+1} & \cdots & a_{n,n} \end{bmatrix} \\
 &= \frac{1}{\det A} C_{j,i}
 \end{aligned} \tag{36}$$

$$x = \frac{1}{\det A} [C_{j,1} \ C_{j,2} \ \cdots C_{j,n}]^T \tag{37}$$

Then we can combine all the solutions to obtain the general solution

$$\begin{aligned}
 A^{-1} &= \bigoplus_{j=1}^n \mathbf{x} = \frac{1}{\det A} \begin{bmatrix} C_{1,1} & C_{2,1} & \cdots & C_{n,1} \\ C_{1,2} & C_{2,2} & \cdots & C_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1,n} & C_{2,n} & \cdots & C_{n,n} \end{bmatrix} \\
 &= \frac{1}{\det A} \operatorname{adj} A
 \end{aligned} \tag{38}$$

The matrix $\operatorname{adj} A$ is called the adjugate matrix of A .

8 Vandermonde Matrix

Definition A Vandermonde matrix is an $m \times n$ matrix with geometric terms in each row, i.e.

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}, \quad \text{or } V_{i,j} = x_i^{j-1} \tag{39}$$

The determinant of the matrix can be calculated as follows when $m = n$. A new matrix V' can be defined such that

$$V'_{i,j} = V_{i,j} - x_1 V_{i,j-1}, \quad 1 \leq i, j \leq n \tag{40}$$

$$V' = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ 1 & x_3 - x_1 & x_3(x_3 - x_1) & \cdots & x_3^{n-2}(x_3 - x_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_1 & x_n(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{bmatrix} \tag{41}$$

As V'^T and V^T are row equivalent, $\det V' = \det V$. Applying Laplace expansion on V' , we obtain

$$\begin{aligned}
 \det V' &= \sum_{k=1}^n v'_{1,k} C_{1,k} = v'_{1,1} \cdot \det B = \begin{vmatrix} x_2 - x_1 & x_2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) & \cdots & x_3^{n-2}(x_3 - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ x_n - x_1 & x_n(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{vmatrix} \\
 &= \prod_{k=2}^n (x_k - x_1) \begin{vmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ 1 & x_3 & \cdots & x_3^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} \end{vmatrix}
 \end{aligned} \tag{42}$$

The determinant of an $m \times n$ has reduced to a $(n-1) \times (n-1)$ one. Recursively, we can conclude that

$$\begin{aligned}
 \det V &= \prod_{k=2}^n (x_k - x_1) \prod_{k=3}^n (x_k - x_2) \cdots \prod_{k=n}^n (x_k - x_{n-1}) \\
 &= \prod_{1 \leq i < j \leq n} (x_j - x_i)
 \end{aligned} \tag{43}$$

9 Eigenvalue and Eigenvector

9.1 Definition

For an $n \times n$ matrix A , if

$$\exists \lambda, \quad A\mathbf{x} = \lambda\mathbf{x} \text{ has non-trivial solution} \quad (44)$$

In other words, $\text{Nul}(A - I\lambda) \neq \{\mathbf{0}\}$. λ is defined as the **eigenvalue** of matrix A , while \mathbf{x} is the eigenvector corresponds to λ .

9.2 Linearity of Eigenvectors

Theorem 8.2.1 if $\lambda_1, \lambda_2, \dots, \lambda_p$ are the mutually different eigenvalues of matrix A , and corresponds to eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ respectively, then

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\} \quad (45)$$

is a set of linearly independent vectors.

Proof If the set is linearly dependent, we obtain

$$\exists \mathbf{v}_r \in \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}, \quad \sum_{k=1}^r c_k \mathbf{v}_k = \mathbf{v}_{r+1} \quad (46)$$

Then

$$\begin{aligned} A \sum_{k=1}^r c_k \mathbf{v}_k - \lambda_{r+1} \sum_{k=1}^r c_k \mathbf{v}_k &= A\mathbf{v}_{r+1} - \lambda_{r+1} \mathbf{v}_{r+1} \\ \sum_{k=1}^r c_k (\lambda_k - \lambda_{r+1}) \mathbf{v}_k &= \mathbf{0} \end{aligned} \quad (47)$$

As $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ are linearly independent, the equation above has only trivial solution $\lambda_k \equiv \lambda_{r+1}$, which contradicts with the condition that the eigenvalues are mutually different. The assumption is invalid. ■

Corollary 8.2.1 If λ is an eigenvalue of matrix A , then λ^m is an eigenvalue of matrix A^m .

Proof We have

$$A\mathbf{x} = \lambda\mathbf{x} \quad (48)$$

has non-trivial solution for $m = 1$. Assume that the corollary holds for $m = n$, then for $m = n + 1$

$$A^{n+1}\mathbf{x} = A(A^n\mathbf{x}) = A(\lambda^n\mathbf{x}) = \lambda^n(A\mathbf{x}) = \lambda^{n+1}\mathbf{x} \quad (49)$$

Then we can conclude that the corollary holds for arbitrary integer m . ■

9.3 Eigenvalue Decomposition (EVD)

If $n \times n$ matrix A has linearly independent eigenvectors of number n , then A is diagonalizable.

$$\begin{aligned} [A\mathbf{x}_1 \quad A\mathbf{x}_2 \quad \cdots \quad A\mathbf{x}_n] &= [\lambda_1\mathbf{x}_1 \quad \lambda_2\mathbf{x}_2 \quad \cdots \quad \lambda_n\mathbf{x}_n] \\ \Leftrightarrow A[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] &= [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \end{aligned} \quad (50)$$

Let $\text{Col } P = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, and the eigenvalues be the diagonal entries of diagonal matrix Λ . Evidently, P is invertible. Thus

$$\begin{aligned} AP &= P\Lambda \\ \boxed{A &= P\Lambda P^{-1}} \end{aligned} \quad (51)$$

The exponent of A can thus be interpreted as

$$\boxed{A^k = P\Lambda^k P^{-1}} \quad (52)$$

9.4 Criterion of Diagonalizability

Theorem 8.4.1 If λ_k is an eigenvalue of A , with multiplicity m_k , then

$$1 \leq \dim E_{\lambda_k} \leq m_k \quad (53)$$

Proof Assume that A is a linear operator on vector space V , then

$$E_{\lambda_k} = \text{Span}\{v_1, v_2, \dots, v_p\}, \quad E_{\lambda_k} \in V \quad (54)$$

If $n \times n$ matrix A has mutually different eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$, $p \leq n$, and the multiplicities are m_1, m_2, \dots, m_p respectively. Using the lemma that any ordered basis of a subspace can be expanded to a basis of the total space, we have

$$\{v_1, v_2, \dots, v_p\} \rightarrow \{v_1, v_2, \dots, v_p, \dots, v_n\} = \beta \quad (55)$$

is an ordered basis of V . Let $T = [A]_\beta$, then

$$\begin{aligned} T &= PAP^{-1} \\ &= \begin{bmatrix} \lambda_k I_p & B \\ \mathbf{0} & C \end{bmatrix} \end{aligned} \quad (56)$$

The characteristic polynomial of A is

$$\begin{aligned} \det(T - tI_n) &= \det\left(\begin{bmatrix} \lambda_k I_p & B \\ \mathbf{0} & C \end{bmatrix} - t \begin{bmatrix} I_p & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix}\right) \\ &= \det\begin{bmatrix} (\lambda_k - t)I_p & B \\ \mathbf{0} & C - tI_{n-p} \end{bmatrix} \\ &= (\lambda_k - t)^p \cdot \det(C - tI_{n-p}) = 0 \end{aligned} \quad (57)$$

The factor $\det(C - tI_{n-p})$ generates a non-negative multiplicity for λ_k . Thus we can conclude that

$$m_k \geq p = \dim E_{\lambda_k} \quad (58)$$

Theorem 8.4.2 $n \times n$ matrix A is diagonalizable if and only if

$$\dim E_{\lambda_i} = m_i \quad (59)$$

for any eigenvalue λ_i of A .

Proof

① **Sufficiency** Assume that A is diagonalizable, then the eigenvector space of linear operator A is V , its eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_p$, and its basis is $\beta = \{v_1, v_2, \dots, v_n\}$. For each eigenvalue λ_i , the set of its eigenvectors are β_{λ_i} . Denote its multiplicity as m_i , dimension of E_{λ_i} as d_i , and the dimension of space $\text{Span}(\beta \cap \beta_{\lambda_i})$ as n_i . Then from 8.4.1 we have

$$d_i \leq m_i \quad (60)$$

Also we can conclude that β_i is a sub-set of eigenvector space E_{λ_i} , then

$$n_i \leq d_i \quad (61)$$

Thus summing up both sides we obtain

$$\sum_{i=1}^n n_i \leq \sum_{i=1}^n d_i \leq \sum_{i=1}^n m_i \quad (62)$$

For the LHS,

$$\sum_{i=1}^n n_i = \dim \text{Span} \left(\beta \cap \left(\bigcup_{i=1}^p \beta_{\lambda_i} \right) \right) = n \quad (63)$$

while obviously for the RHS,

$$\sum_{i=1}^n m_i = n \quad (64)$$

Thus

$$0 \leq \sum_{i=1}^n (d_i - m_i) \leq 0 \quad (65)$$

Then $m_i = d_i$ holds for all eigenvalues. ■

② **Necessity** Given that

$$\dim E_{\lambda_i} = d_i = m_i \quad (66)$$

then denote the set of basis for E_{λ_i} as β_{λ_i} . As all the eigenvectors are linearly independent, we have

$$\beta_{\lambda_i} \cap \beta_{\lambda_j} = \emptyset \quad \text{for } i \neq j \quad (67)$$

Hence

$$\dim \text{Span} \bigcup_{i=1}^p \beta_{\lambda_i} = \sum_{i=1}^p m_i = n \quad (68)$$

Thus, the eigenvectors are sufficient to generate V , or A is diagonalizable. ■

10 Eigenvectors from Eigenvalues - Terence Tao

It is proved that

$$|\mathbf{v}_{ij}|^2 \prod_{k=1; k \neq i}^n (\lambda_i(A) - \lambda_k(A)) = \prod_{k=1}^{n-1} (\lambda_i(A) - \lambda_k(M_j)) \quad (69)$$

Where A is an $n \times n$ matrix, \mathbf{v}_{ij} is the j^{th} element of the i^{th} eigenvector of A , M_{jj} is the jj cofactor of A .

10.1 Proof

Consider the adjugate matrix of A , i.e.

$$\text{adj}(A) = \sum_{1 \leq i, j \leq n} (-1)^{i+j} \det M_{ji} \quad (70)$$

Lemma 1

$$\text{adj}(A)A = A\text{adj}(A) = \det(A)I_n \quad (71)$$

Proof We consider replace the j^{th} column of A with its k^{th} column, which generates A' , then we have

$$\det(A') = \sum_{i=1}^n a'_{ij} C'_{ij} = \sum_{i=1}^n a_{ik} C_{ik} = \begin{cases} 0 & j \neq k \\ \det(A) & j = k \end{cases} \quad (72)$$

Applying the identity to $\det(A)A$, we obtain

$$\begin{aligned} \text{adj}(A)A &= \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n a_{i1} C_{i1} & \sum_{i=1}^n a_{i2} C_{i1} & \cdots & \sum_{i=1}^n a_{in} C_{i1} \\ \sum_{i=1}^n a_{i1} C_{i2} & \sum_{i=1}^n a_{i2} C_{i2} & \cdots & \sum_{i=1}^n a_{in} C_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{i1} C_{in} & \sum_{i=1}^n a_{i2} C_{in} & \cdots & \sum_{i=1}^n a_{in} C_{in} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n a_{i1} C_{i1} & & & \\ & \sum_{i=1}^n a_{i2} C_{i2} & & \\ & & \ddots & \\ & & & \sum_{i=1}^n a_{in} C_{in} \end{bmatrix} = \det(A)I_n \end{aligned} \quad (73)$$

The proof is exactly the same for $A\text{adj}(A)$. ■

The spectral decomposition for A gives $A = \sum_{i=1}^n \lambda_i(A) \mathbf{v}_i \mathbf{v}_i^T$, where \mathbf{v}_i are orthonormal vectors. Then

$$A\mathbf{v}_k = \sum_{i=1}^n \lambda_i(A) \mathbf{v}_i \mathbf{v}_i^T \mathbf{v}_k = \lambda_k(A) \mathbf{v}_k \mathbf{v}_k^T \mathbf{v}_k = \lambda_k(A) \mathbf{v}_k \quad (74)$$

Notice that

$$\mathbf{adj}(A)A\mathbf{v}_k = \mathbf{det}(A)I_n\mathbf{v}_k = \lambda_k(A)\mathbf{adj}(A)\mathbf{v}_k \rightarrow \mathbf{adj}(A)\mathbf{v}_k = \frac{\mathbf{det}(A)}{\lambda_k(A)}\mathbf{v}_k \quad (75)$$

where

$$\mathbf{det}(A) = \mathbf{det}(P\Lambda P^{-1}) = \mathbf{det}(\Lambda)\mathbf{det}(P)\mathbf{det}(P^{-1}) = \prod_{i=1}^n \lambda_i(A) \quad (76)$$

Substitute in we obtain

$$\mathbf{adj}(A)\mathbf{v}_k = \prod_{i=1; i \neq k}^n \lambda_i(A)\mathbf{v}_k \quad (77)$$

Hence, $\prod_{i=1; i \neq k} \lambda_i$ is the eigenvalue that corresponds to \mathbf{v}_k of $\mathbf{adj}(A)$. Likewise, we apply spectral decomposition to $\mathbf{adj}(A)$, we obtain

$$\mathbf{adj}(A) = \sum_{i=1}^n \prod_{k=1; k \neq i}^n \lambda_k(A) \mathbf{v}_i \mathbf{v}_i^T \quad (78)$$

As $\lambda_1, \lambda_2, \dots, \lambda_n$ are the roots to A 's characteristic equation, then $\mathbf{det}(\lambda I_n - A) = \prod_{i=1}^n (\lambda - \lambda_i(A))$. We have

$$\mathbf{adj}(\lambda I_n - A) = \sum_{i=1}^n \prod_{k=1; k \neq i}^n (\lambda - \lambda_k(A)) \mathbf{v}_i \mathbf{v}_i^T \quad (79)$$

Take $\lambda = \lambda_i(A)$, then all the term including $\lambda - \lambda_i(A)$ will be eliminated. That is

$$\mathbf{adj}(\lambda_i(A)I_n - A) = \prod_{k=1; k \neq i}^n (\lambda_i(A) - \lambda_k(A)) \mathbf{v}_i \mathbf{v}_i^T \quad (80)$$

Then we extract the j^{th} column and j^{th} row on the both hand sides, then we obtain

$$\mathbf{adj}(\lambda_i(A)I_n - A)_{jj} = (-1)^{j+j} \mathbf{det}(\lambda_i(A)I_{n-1} - M_j) = \prod_{k=1; k \neq i}^n (\lambda_i(A) - \lambda_k(A)) (\mathbf{v}_i \mathbf{v}_i^T)_{jj} \quad (81)$$

The simplified form yields ■

$$\mathbf{det}(\lambda_i(A)I_{n-1} - M_j) = \prod_{k=1}^n (\lambda_i(A) - \lambda_k(M_j)) = |\mathbf{v}_{ij}|^2 \prod_{k=1; k \neq i}^n (\lambda_i(A) - \lambda_k(A)) \quad (82)$$

11 Markov Chain

11.1 Concept

Definition ① A probability vector is defined as

$$\{[x_1 \ x_2 \ \cdots \ x_n]^T \mid \sum_{i=1}^n x_i = 1\} \quad (83)$$

② A stochastic matrix is defined as

$$\{M | M = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_n]\} \quad (84)$$

with the vectors $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ all being probability vectors.

③ A Markov Chain is a set of probability vector sequence $\{u_0, u_1, u_2, \dots\}$, with the state shifting equation

$$\mathbf{u}_{k+1} = M\mathbf{u}_k, \quad \mathbf{u}_k = M^k\mathbf{u}_0 \quad (85)$$

M is a **stochastic matrix**. The element $M_{i,j}$ depicts the probability to shift from state i to state j .

11.2 Kolmogorov Equation

Mathematically, we can prove that in every state shifting process, the action of a definite probability will be

$$m_{i,j} = \sum_{k=1}^n m_{i,k} \cdot m_{k,j} \quad (86)$$

This is the Kolmogorov equation in Markov chain. **Intuitively, it tells that the probability of shifting from status i to j is the weighted summation of the probability to shift by any possible intermediate state.**

11.3 Equilibrium State

Theorem 9.3.1 1 is an eigenvalue of the state shifting operator M .

Proof If 1 is an eigenvalue, then

$$A - \lambda I \Big|_{\lambda=1} = [\mathbf{p}_1 - \mathbf{e}_1 \ \mathbf{p}_2 - \mathbf{e}_2 \ \cdots \ \mathbf{p}_n - \mathbf{e}_n] \quad (87)$$

As

$$\sum_{i=1}^n (\mathbf{p}_k - \mathbf{e}_k)_i = \sum_{i=1}^n (\mathbf{p}_k)_i - 1 = 0 \quad (88)$$

Thus,

$$\sum_{i=1}^n (\mathbf{p}_i - \mathbf{e}_i)^T = \mathbf{0} \quad (89)$$

The rows of $A - I$ are linearly dependent, hence

$$\det(A - I) = 0, \quad \text{QED} \quad (90)$$

Theorem 9.3.3 A Markov chain will eventually converge to an equilibrium state for an arbitrary initial state.

Proof Assume that an arbitrary initial state is

$$\mathbf{u}_0 = [x_1 \ x_2 \ \cdots \ x_n]. \quad (91)$$

Using the set of eigenvector space of M , denoted as \mathcal{M} as a basis for \mathbb{R}^n , and all the eigenvectors form the columns of P , such that

$$\mathbf{u}_0 = P[\mathbf{u}_0]_{\mathcal{M}} \quad (92)$$

Hence

$$\begin{aligned} \mathbf{u}_k &= P\Lambda^k P^{-1}\mathbf{u}_0 \\ &= P\Lambda^k P^{-1}P[\mathbf{u}_0]_{\mathcal{M}} = P\Lambda^k[\mathbf{u}_0]_{\mathcal{M}} \\ &= [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\ &= \sum_{i=1}^n c_i \lambda_i^k \mathbf{v}_i \end{aligned} \quad (93)$$

The equilibrium state $\boldsymbol{\pi}$ can be calculated as

$$\boldsymbol{\pi} = \lim_{k \rightarrow \infty} \mathbf{u}_k = c_e \mathbf{v}_e \quad (94)$$

where the corresponding eigenvalue $\lambda_e = 1$.

12 Orthogonality

12.1 Inner Product Space

Definition The inner product on a vector space V is a function. For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ on V , if a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ follows the axioms below

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= \langle \mathbf{v}, \mathbf{u} \rangle \\ \langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle \\ \langle c\mathbf{u}, \mathbf{v} \rangle &= c\langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{u}, \mathbf{u} \rangle &\geq 0 \\ \langle \mathbf{u}, \mathbf{u} \rangle &= 0 \text{ iff } \mathbf{u} = \mathbf{0} \end{aligned} \quad (95)$$

Then we call V an inner product space.

Definition The norm of vector is defined as

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle \quad (96)$$

Inner Product in \mathbb{R}^n

For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the inner product is defined as

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \rightarrow \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{u} = \sum_{k=1}^n u_k v_k \quad (97)$$

Inner Product in \mathbb{P}_n

In the polynomial space \mathbb{P}_n , the basis is

$$\mathcal{P} = \{1, t, t^2, \dots, t^n\} \quad (98)$$

An element p in \mathbb{P}_n is defined as

$$p(t) = \mathcal{P}[p]_{\mathcal{P}} = \sum_{k=1}^n c_k t^k \quad (99)$$

When t takes different values t_0, t_1, \dots, t_n , the polynomial element p can be interpreted as

$$\mathbf{p} = \begin{bmatrix} p(t_0) \\ p(t_1) \\ \vdots \\ p(t_n) \end{bmatrix} \quad (100)$$

Likewise, the inner product in \mathbb{P}_n is

$$\langle \mathbf{p}, \mathbf{q} \rangle = \sum_{k=1}^n p(t_k) q(t_k) \quad (101)$$

Inner Product on $C[a, b]$

A widely used inner space on interval $C[a, b]$ is the set of all continuous functions defined on it. We generalized the inner product in polynomial space. For functions \mathbf{f}, \mathbf{g} on $C[a, b]$, let $\Delta t = b - a/n + 1$, thus $t_j = j(b - a)/n + 1$. The interval is split into n sections, where

$$\max\{t_1, t_2, \dots, t_n\} = t_k \rightarrow 0 \quad (102)$$

Thus the inner product is

$$\begin{aligned} \langle \mathbf{f}, \mathbf{g} \rangle &= \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{j=0}^n f(t_j) g(t_j) \\ &= \frac{1}{b-a} \lim_{n \rightarrow \infty} \frac{b-a}{n+1} \sum_{j=0}^n f\left(\frac{j(b-a)}{n+1}\right) g\left(\frac{j(b-a)}{n+1}\right) \\ &= \frac{1}{b-a} \int_a^b f(t) g(t) dt \end{aligned} \quad (103)$$

As the coefficient does not affect the completeness of inner product space, we normalize it and obtain

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(t) g(t) dt \quad (104)$$

12.2 Orthogonality

Definition

In an inner product space, two vectors are orthogonal if and only if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad (105)$$

Definition

A set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is called an orthogonal set if and only if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle \Big|_{i \neq j} = 0 \quad (106)$$

An orthogonal set that generates a space V is called an **orthogonal basis** of V .

Lemma 10.2.1

An orthogonal set constructed by non-zero vectors in an inner product space V is a basis of the subspace P it generates.

Proof For an orthogonal set $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$, if it is linear dependent, then

$$\begin{aligned} \exists \mathbf{u}_k \in \beta, \mathbf{u}_k &= \sum_{i=1}^{k-1} c_i \mathbf{u}_i \\ \text{For } 1 \leq j \leq k-1, \quad \langle \mathbf{u}_k, \mathbf{u}_j \rangle &= \left\langle \sum_{i=1}^{k-1} c_i \mathbf{u}_i, \mathbf{u}_j \right\rangle = c_j \langle \mathbf{u}_j, \mathbf{u}_j \rangle = 0 \end{aligned}$$

As \mathbf{u}_j is a non-zero vector, thus

$$\langle \mathbf{u}_j, \mathbf{u}_j \rangle \geq 0 \rightarrow c_j = 0 \quad (107)$$

which draws the conclusion that

$$\mathbf{u}_k = \mathbf{0} \quad (108)$$

It is evident that the requisite holds for no vector in set β . The assumption is thus invalid. ■

Orthogonal Matrix

Assume that β is an orthogonal basis of an inner product space, and

$$\beta = \left\{ \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\} \mid \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0, \|\mathbf{u}_k\| = 1 \text{ for } 1 \leq i, j, k \leq p, i \neq j \right\} \quad (109)$$

If a matrix U satisfies $\text{Col}U = \text{Span}\beta$, then U is called an **orthogonal matrix**. A matrix U is an orthogonal matrix if and only if

$$\begin{aligned} U^T U &= [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_p]^T [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_p] \\ &= \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \cdots & \mathbf{u}_1^T \mathbf{u}_p \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \cdots & \mathbf{u}_2^T \mathbf{u}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_p^T \mathbf{u}_1 & \mathbf{u}_p^T \mathbf{u}_2 & \cdots & \mathbf{u}_p^T \mathbf{u}_p \end{bmatrix} = \begin{bmatrix} \|\mathbf{u}_1\|^2 & & & \\ & \|\mathbf{u}_2\|^2 & & \\ & & \ddots & \\ & & & \|\mathbf{u}_p\|^2 \end{bmatrix} = I \end{aligned} \quad (110)$$

12.3 Orthogonal Complement**Definition**

For an inner product space V , if another inner product space W follows

$$\{\mathbf{w} \mid \mathbf{w} \in W, \forall \mathbf{v} \in V, \langle \mathbf{w}, \mathbf{v} \rangle = 0\} \quad (111)$$

then W is called the orthogonal complement of space V , denoted as V^\perp .

Subspace Defined by Linear Operator A

The two typical subspace of an inner product space V defined by operator A is

$$\text{Col}A = \{\mathbf{b} | A\mathbf{x} = \mathbf{b}\}, \quad \text{Nul}A = \{\mathbf{x} | A\mathbf{x} = \mathbf{0}\} \quad (112)$$

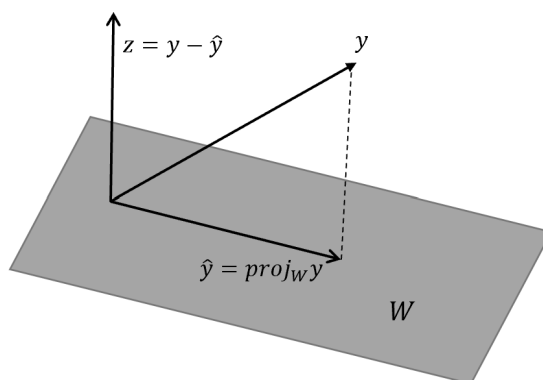
We can establish a relationship between this two subspace,

$$\begin{aligned} \forall \mathbf{u} \in \text{Nul}A, \forall \mathbf{v} \in \text{Row}A = \text{Col}A^T, \exists \mathbf{w} \in V, A^T \mathbf{w} = \mathbf{v} \\ \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A^T \mathbf{w} \rangle = (A^T \mathbf{w})^T \mathbf{u} = \mathbf{w}^T (A^T)^T \mathbf{u} = \mathbf{w}^T (A\mathbf{u}) = 0 \end{aligned}$$

Thus we can conclude that ■

$$\begin{aligned} (\text{Row}A)^\perp &= \text{Nul}A \\ (\text{Row}A^T)^\perp &= (\text{Col}A)^\perp = \text{Nul}A^T \end{aligned}$$

12.4 Orthogonal Projection



Definition Two vectors \mathbf{u}, \mathbf{v} are defined in an inner vector space, where $\mathbf{u} \cap \mathbf{v} = \{\mathbf{0}\}$. Then \mathbf{u} can be uniquely decomposed into

$$\mathbf{u} = \hat{\mathbf{u}} + \mathbf{z} = \alpha \mathbf{v} + \mathbf{z} = \text{proj}_{\text{Span}\{\mathbf{v}\}} \mathbf{u} + \mathbf{z} \quad (113)$$

where $\mathbf{z} \in (\text{Span}\{\mathbf{v}\})^\perp$, and $\alpha \mathbf{v}$ is called the **orthogonal projection** of \mathbf{u} on $\text{Span}\{\mathbf{v}\}$. The coefficient α can be calculated as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{z}, \mathbf{v} \rangle = \alpha \langle \mathbf{v}, \mathbf{v} \rangle \rightarrow \alpha = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}, \quad \text{proj}_{\text{Span}\{\mathbf{v}\}} \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v} \quad (114)$$

Corollary 10.4.1 For an inner product space W , if an orthogonal basis $w = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}$ sufficiently generates it, then if every single orthogonal basis corresponds to the columns of matrix P , then

$$\begin{aligned} \forall \mathbf{y} \in W, \quad \mathbf{y} = P[\mathbf{y}]_w = \sum_{i=1}^p c_i \mathbf{w}_i \\ \text{For } 1 \leq j \leq p, \quad \langle \mathbf{w}_j, \mathbf{y} \rangle = \langle \mathbf{w}_j, \sum_{i=1}^p c_i \mathbf{w}_i \rangle = c_j \langle \mathbf{w}_j, \mathbf{w}_j \rangle \end{aligned} \quad (115)$$

Thus

$$c_j = \frac{\langle \mathbf{w}_j, \mathbf{y} \rangle}{\langle \mathbf{w}_j, \mathbf{w}_j \rangle} \quad (116)$$

Without loss of generality, we can conclude that any vector \mathbf{y} in an inner product space W with an arbitrary orthogonal basis β can be interpreted as

$$\beta = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}, \mathbf{y} \notin W \rightarrow \hat{\mathbf{y}} = \sum_{i=1}^p \text{proj}_{\text{Span}\{\mathbf{w}_i\}} \mathbf{y} \quad (117)$$

A more generalized expression is:

$$\text{proj}_{\text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}} \mathbf{y} = \sum_{i=1}^p \text{proj}_{\text{Span}\{\mathbf{w}_i\}} \mathbf{y} \quad (118)$$

12.5 Gram-Schmidt Orthogonalization

For an inner product space W , assume that $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is a basis of W . Then we generate an orthogonal basis $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$. Then the orthogonalization for a basis vector \mathbf{u}_k is

$$\begin{aligned} \mathbf{v}_k &= \mathbf{u}_k - \text{proj}_{\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k-1}\}} \mathbf{u}_k \\ &= \mathbf{u}_k - \sum_{i=1}^{k-1} \text{proj}_{\text{Span}\{\mathbf{v}_i\}} \mathbf{u}_k \\ &= \mathbf{u}_k - \sum_{i=1}^n \frac{\langle \mathbf{u}_k, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \mathbf{v}_i \end{aligned} \quad (119)$$

13 Method of Least Square

13.1 Principle

The method of least square is a standard approach to approximate the solution for **overdetermined system**. Given a set of collected data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, we tend to obtain a generalized form of the solution

$$y_j = \sum_{i=0}^k \beta_i f_i(x_j) \quad (120)$$

Or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_k(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad (121)$$

Which can be written as

$$\mathbf{y} = X\beta \quad (122)$$

\mathbf{y} is the **sample vector**. X is prediction value vector based on every x , called the **design matrix**. β is the **parameter vector**.

13.2 Solution of Least Square Equation

The equation (102) is not always consistent (as the system is over-determined). There will be some bias between $\text{Col}X$ and y . Then we write (102) as

$$y = X\beta + \epsilon \Leftrightarrow \epsilon = y - X\beta \quad (123)$$

We tend to minimize $\|\epsilon\|$. Thus we take

$$X\beta = \hat{y} = \text{proj}_{\text{Col}X} y \quad (124)$$

$\hat{y} \in \text{proj}_{\text{Col}X} y$, hence the equation is now consistent. As $z = y - \hat{y} \in (\text{Col}X)^T = \text{Nul}X^T$

$$X^T(y - \hat{y}) = X^T(y - X\beta) = 0 \rightarrow \boxed{X^T X\beta = X^T y} \quad (125)$$

13.3 Applications

13.3.1 Weighted Least Square Method

Sometimes data pairs have different significance to fitting the curve. Thus we give every pair of data a weight w_i when maximizing the residual.

$$\epsilon' = \begin{bmatrix} w_1(y_1 - \hat{y}_1) \\ w_2(y_2 - \hat{y}_2) \\ \vdots \\ w_n(y_n - \hat{y}_n) \end{bmatrix} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} = W\epsilon \quad (126)$$

Then to minimize the norm of $\epsilon' = \sqrt{\langle \epsilon', \epsilon' \rangle}$, we solve the equation

$$WX\beta = Wy \rightarrow (WX)^T(WX)\beta = (WX)^T y \quad (127)$$

13.3.2 Lagrange Interpolation

To approximate a curve, an intuitive choice is to choose polynomials as the basis. For points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, we use a homogeneous polynomial

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_{n-1} x^{n-1} \quad (128)$$

Then we build up the matrix equation $X\beta = y$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (129)$$

Applying Cramer's rule,

$$\begin{aligned} \beta_i &= \frac{\det X_i(y)}{\det X} = \frac{1}{\prod_{1 \leq i < j \leq n} (x_j - x_i)} \begin{vmatrix} 1 & x_1 & \cdots & x_1^{i-2} & y_1 & x_1^i & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{i-2} & y_2 & x_2^i & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{i-2} & y_n & x_n^i & \cdots & x_n^{n-1} \end{vmatrix} \\ &= \frac{1}{\prod_{1 \leq i < j \leq n} (x_j - x_i)} \sum_{j=1}^n y_j C_{j,i} = \frac{1}{\prod_{1 \leq i < j \leq n} (x_j - x_i)} \sum_{j=1}^n (-1)^{i+j} y_j A_{j,i} \end{aligned} \quad (130)$$

Where $A_{j,i}$ equals to the coefficient of t^{i-1} in

$$\begin{vmatrix} 1 & x_1 & \cdots & x_1^{i-2} & x_1^i & x_1^{i+1} & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{i-2} & x_2^i & x_2^{i+1} & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t & \cdots & t^{i-2} & t^{i-1} & t^i & \cdots & t^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{i-2} & x_n^{i-1} & x_n^i & \cdots & x_n^{n-1} \end{vmatrix} = (-1)^{n-j} \prod_{\substack{1 \leq j-1 \\ s < k \leq n}} (x_k - x_s) \prod_{\substack{1 \leq k \leq n \\ k \neq j}} (t - x_k) \prod_{j+1 \leq s < k \leq n} (x_k - x_s)$$

$$\text{Coef}_t = (-1)^{n-j} \prod_{\substack{1 \leq j-1 \\ s < k \leq n}} (x_k - x_s) \prod_{j+1 \leq s < k \leq n} (x_k - x_s) \cdot (-1)^{n-i+1} \sum_{1 \leq \alpha_1 \leq \cdots \leq \alpha_{i-1} \leq n} x_{\alpha_1} x_{\alpha_2} \cdots x_{\alpha_{i-1}}$$

Actually the polynomial can be written into a simplified form that

$$f(x) = \sum_{i=0}^{n-1} \beta_i x^i = \sum_{i=1}^n y_i \prod_{\substack{1 \leq j \leq n \\ j \neq i}} \frac{(x - x_j)}{(x_i - x_j)} \quad (131)$$

13.3.3 Fourier Series

Lemma For $n \geq 1$, the following set

$$\mathcal{F} = \{1, \cos t, \cos 2t, \dots, \cos nt, \sin t, \sin 2t, \dots, \sin nt\}$$

is an orthogonal set under inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt$$

Proof

$$\begin{aligned} \langle \cos at, \cos bt \rangle &= \int_0^{2\pi} \frac{\cos(a-b)t + \cos(a+b)t}{2} dt = 0 \\ \langle \cos at, \sin bt \rangle &= \int_0^{2\pi} \frac{\sin(b+a)t + \sin(b-a)t}{2} dt = 0 \end{aligned}$$

for $a \neq b$ ■

Then \mathcal{F} is an orthogonal basis of a subspace W in $C[0, 2\pi]$. The **optimal approximation** of an arbitrary function f is

$$f(t) = a_0 + \sum_{m=1}^n (a_m \cos mt + b_m \sin mt) \quad (132)$$

where

$$a_m = \frac{\langle f, \cos mt \rangle}{\langle \cos mt, \cos mt \rangle}, \quad b_m = \frac{\langle f, \sin mt \rangle}{\langle \sin mt, \sin mt \rangle} \quad (133)$$

(108) is called the **Fourier series** of $f(t)$.

14 Jacobian Matrix

If there are two coordinate systems with mapping relationship between basis

$$\alpha = \{x_1, x_2, \dots, x_n\} \xrightarrow{f: \mathbb{R}^n \rightarrow \mathbb{R}^n} \beta = \{y_1, y_2, \dots, y_n\} \quad (134)$$

$$\text{or } y_i = f_i(x_1, x_2, \dots, x_n) \quad (135)$$

Then the relationship between the infinitesimals correspond to the basis are

$$dy_i = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} dx_j \quad (136)$$

or

$$\begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \frac{\partial f_1}{\partial x_i} dx_i \\ \sum_{i=1}^n \frac{\partial f_2}{\partial x_i} dx_i \\ \vdots \\ \sum_{i=1}^n \frac{\partial f_n}{\partial x_i} dx_i \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} \quad (137)$$

Lemma 1 Assume that a geometry body in the space spanned by basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, and the linear transformation depicted by $n \times n$ matrix A transforms the basis into $\{y_1, y_2, \dots, y_n\}$, i.e.

$$A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} \quad (138)$$

determines new coordinates y_1, y_2, \dots, y_n . The magnitude of geometry body determined by new coordinates is

$$V_Y = |\det A| V_X \quad (139)$$

Proof For each vector in the original space, we have

$$\mathbf{u} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_n \mathbf{x}_n \quad (140)$$

Then the magnitude to the geometry body depicted by vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ is

$$\det [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \quad (141)$$

For the transformed vector, we have

$$\mathbf{u}' = c_1 \mathbf{y}_1 + c_2 \mathbf{y}_2 + \cdots + c_n \mathbf{y}_n = c_1 A \mathbf{x}_1 + c_2 A \mathbf{x}_2 + \cdots + c_n A \mathbf{x}_n = A(c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_n \mathbf{x}_n) \quad (142)$$

Then the new geometry body has the magnitude

$$\begin{aligned} \det [\mathbf{u}'_1 \quad \mathbf{u}'_2 \quad \cdots \quad \mathbf{u}'_n] &= \det [A \mathbf{u}_1 \quad A \mathbf{u}_2 \quad \cdots \quad A \mathbf{u}_n] = \det A [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \\ &= (\det A)(\det [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n]) \end{aligned} \quad \blacksquare$$

Then the relationship between the magnitude of the infinitesimal geometry body is

$$\Omega_Y = \left\| \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \right\| \Omega_X = |\det \mathbb{J}| \Omega_X \quad (143)$$

Where \mathbb{J} is called the **Jacobian Matrix**. It depicts the scaling coefficient between two infinitesimals in different coordinates.

15 Symmetric Matrix

15.1 Concept

If a matrix A satisfies

$$A = A^T \quad (144)$$

then it is called a **symmetric matrix**.

15.2 Orthogonality of Eigenvectors

The set of the eigenvectors of a symmetric matrix is an orthogonal set.

Proof Assume that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are eigenvectors of symmetric matrix A , and correspond to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Then

$$\begin{aligned} \forall \lambda_i, \lambda_j, \quad A\mathbf{v}_i &= \lambda_i\mathbf{v}_i, \quad A\mathbf{v}_j = \lambda_j\mathbf{v}_j \\ \lambda_i\mathbf{v}_i^T &= \mathbf{v}_j^T A^T \rightarrow \lambda_i\mathbf{v}_i^T \mathbf{v}_j = \mathbf{v}_i^T A^T \mathbf{v}_j = \mathbf{v}_i^T A \mathbf{v}_j = \lambda_j\mathbf{v}_i^T \mathbf{v}_j \end{aligned} \quad (145)$$

As $\lambda_i \neq \lambda_j$, we have

$$\mathbf{v}_i^T \mathbf{v}_j = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \quad (146)$$

which indicates that any two eigenvectors in the set are mutually orthogonal. Also we can conclude that the eigenvectors form an orthogonal basis of the column space determined by matrix A . ■

15.3 Orthogonal Diagonalization

We tend to find the diagonalization of matrix A , i.e.

$$A = P\Lambda P^{-1} \quad (147)$$

With the eigenvectors forming the columns of matrix P . If we unitize each eigenvector to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, $\mathbf{u}_i = \mathbf{v}_i / \sqrt{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}$, then we denote the new matrix as Q , where $QQ^T = 1$. We have

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T \rightarrow A^T = (Q\Lambda Q^T)^T = Q\Lambda Q^T = A \quad (148)$$

Thus the diagonalization is well-defined.

15.4 Spectral Decomposition

The diagonalization above gives

$$A = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T \quad (149)$$

15.5 Singular Value Decomposition (SVD)

15.5.1 Derivation

We are interested in building up a generalized decomposition for an arbitrary $m \times n$ matrix A . Notice that $(A^T A)^T = A^T (A^T)^T = A^T A$, which indicates that $A^T A$ itself is a symmetric matrix. Then we apply EVD to $A^T A$.

Assume that $A^T A = Q \Lambda Q^T$, with $\text{rank}(A) = k$, so the eigenvectors set $q = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ forms an orthonormal basis for Q , then

$$\forall \mathbf{v}_1, \mathbf{v}_2 \in q, \quad \langle A\mathbf{v}_i, A\mathbf{v}_j \rangle = (A\mathbf{v}_j)^T A\mathbf{v}_i = \mathbf{v}_j^T A^T A \mathbf{v}_i = \mathbf{v}_j^T \lambda_i \mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i = \lambda_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$$

So the set $q' = \{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_k\}$ is also orthogonal. Assume that the vectors in q forms the columns of matrix V , then

$$AV = A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_k \end{bmatrix} \quad (150)$$

If we normalize q' , i.e. $\mathbf{u}_i = A\mathbf{v}_i / \sqrt{\langle A\mathbf{v}_i, A\mathbf{v}_i \rangle} = A\mathbf{v}_i / \sqrt{\lambda_i}$, then we denote the singular value $\sigma_i = \sqrt{\lambda_i}$ and obtain

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \Leftrightarrow A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix} \quad (151)$$

We use U and Σ to denote the two matrices on the RHS respectively, and thus we conclude that

$$AV = U\Sigma \rightarrow \boxed{A = U\Sigma V^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T} \quad (152)$$

The product on the RHS shows the **singular value decomposition** of an arbitrary matrix A . If we expand the set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ to form a basis of \mathbb{R}^n and \mathbb{R}^n respectively, we then obtain

$$A = \left[\begin{array}{cccc|cccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_n \end{array} \right] \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & & \mathbf{0} \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ \hline & & & \mathbf{0} & & & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_m^T \end{array} \right] \quad (153)$$

15.5.2 Application