1 Basic Concept

1.1 Fundamental Operations of Matrix

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

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

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

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

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

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

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

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

$$(A,\ \boldsymbol{b}) \sim (I,\ \boldsymbol{x}) \tag{6}$$

1.2 Matrix and Linear Equations

The following share the same solution:

$$\begin{cases} \text{Matrix equation} \quad A\boldsymbol{x} = \boldsymbol{b} \\ \text{Vector equation} \quad \displaystyle\sum_{k=1}^{n} x_k \boldsymbol{a_k} = \boldsymbol{b} \\ \text{Augmented matrix} \quad [\boldsymbol{a_1}\ \boldsymbol{a_2}\ \ldots\ \boldsymbol{a_n}\ \boldsymbol{b}] \end{cases} \tag{7}$$

There are three cases for the solution:

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

1.3 Linearity

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

a. for every $\boldsymbol{b}$ in $\mathbb{R}^m$, $A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution
b. $\boldsymbol{Span}\{a_1, \ldots, a_n\} = \mathbb{R}^m$
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 $A\boldsymbol{x} = \boldsymbol{0}$ 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. $\mathbb{R}^n \to \mathbb{R}^{n+1}$. That is

$$(x_1, x_2, \ldots, x_n) \to (x_1, x_2, \ldots, x_n, 1) \tag{9}$$

Then we can use a $(n+1) \times (n+1)$ matrix A to represent the transform. **$\boldsymbol{A}$ 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

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

2.2.2 Translation

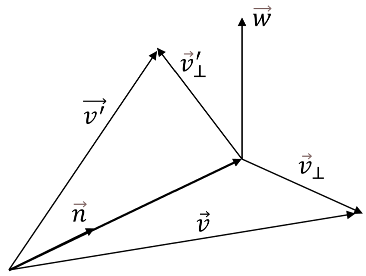
If the translation vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$, then

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

2.2.3 Rotation

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

$$\boldsymbol{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}, \quad n_x^2 + n_y^2 + n_z^2 = 1 \tag{12}$$



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

$$\boldsymbol{v}_{||} = (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}, \quad \boldsymbol{v}_{||} + \boldsymbol{v}_\perp = \boldsymbol{v}$$

$$\boldsymbol{v}'_\perp = \boldsymbol{v}_\perp \cos\theta + \boldsymbol{\omega}\sin\theta = (\boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n})\cos\theta + (\boldsymbol{v}\times\boldsymbol{n})\sin\theta$$

$$\boldsymbol{v}' = \boldsymbol{v}'_\perp + \boldsymbol{v}'_{||} = (\boldsymbol{v}\times\boldsymbol{n})\sin\theta + (\boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n})\cos\theta + (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}$$

Substitute in we obtain

$$\boldsymbol{v}' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \sin\theta \begin{bmatrix} n_z y - n_y z \\ n_x z - n_z x \\ n_y x - n_x y \end{bmatrix} + \cos\theta \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} (n_x x + n_y y + n_z z) \right) + \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} (n_x x + n_y y + n_z z)$$

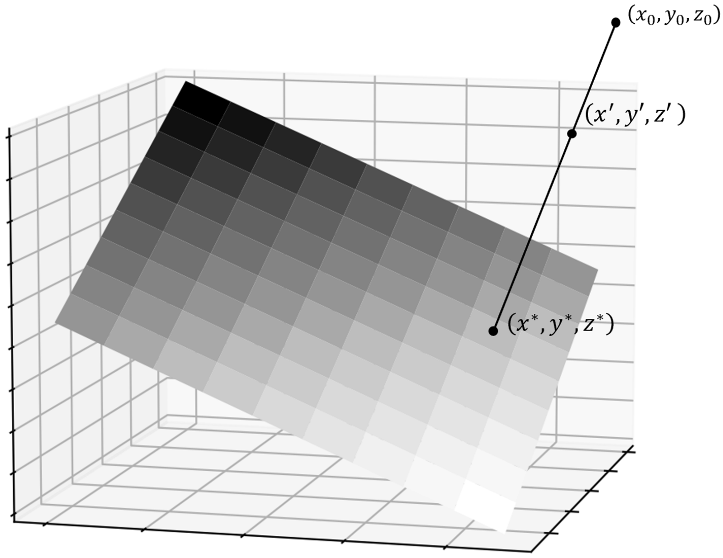
$$= \begin{bmatrix} (n_x^2(1-\cos\theta)+\cos\theta)x + (n_x n_y(1-\cos\theta)+n_z\sin\theta)y + (n_x n_z(1-\cos\theta)-n_y\sin\theta)z \\ (n_y n_z(1-\cos\theta)-n_z\sin\theta)x + (n_y^2(1-\cos\theta)+\cos\theta)y + (n_z n_y(1-\cos\theta)+n_x\sin\theta)z \\ (n_x n_z(1-\cos\theta)+n_y\sin\theta)x + (n_y n_z(1-\cos\theta)-n_x\sin\theta)y + (n_z^2(1-\cos\theta)+\cos\theta)z \end{bmatrix}$$

Then the transform matrix is

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

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:

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

$$\rightarrow \begin{cases} Ax^* + By^* + Cz^* + D = 0 \\ x^* = (x' - x_0)k + x_0 \\ y^* = (y' - y_0)k + y_0 \\ z^* = (z' - z_0)k + z_0 \end{cases} \tag{14}$$

Solving (14) we obtain

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

Hence the projection matrix is

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

3 Gaussian Elimination

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

$$\begin{cases} \textbf{Multiplying}\text{: Multiplying a row by a non-zero number} \\ \textbf{Adding}\text{: Adding the multiple of a row to another one} \\ \textbf{Swapping}\text{: Exchanging the position of two definite rows} \end{cases} \tag{17}$$

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:

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

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

$$E_pE_{p-1}\cdots E_1A = U, \quad \text{or } L = (E_pE_{p-1}\cdots E_1)^{-1}, \ A = LU \tag{19}$$

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

$$(A,\ L) \sim (U,\ I) \tag{20}$$

4 Inverse of Matrix

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

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

4.1 Criterion

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

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

4.2 Proof

If the columns of A are linear dependent,

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

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

Then

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

And vice versa, QED

5 Dimensions and Rank

5.1 Basis

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

$$[\boldsymbol{x}]_{\boldsymbol{\mathcal{B}}} = [c_1, c_2, \ldots, c_p]^T, \quad \boldsymbol{x} = [\mathcal{B}][\boldsymbol{x}]_{\boldsymbol{\mathcal{B}}} \tag{22}$$

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 <u>echelon form</u> 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$\boldsymbol{A}$**, while the corresponding columns to the pivot columns of U in A are the basis of **Col$\boldsymbol{A}$** .

5.2 Dimension

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

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

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

5.3 Rank

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

$$\mathrm{rank}A + \mathrm{dim\ Nul}A = n \tag{23}$$

6 Inverse of Block Matrix

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

$$T = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \tag{24}$$

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

$$T^{-1} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} \tag{25}$$

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

$$\begin{cases} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} = \boldsymbol{I} \\ \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} = \boldsymbol{0} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} = \boldsymbol{0} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} = \boldsymbol{I} \end{cases} \tag{26}$$

We take steps to solve the equations system.

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

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

Hence we can conclude that the inverse matrix is

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

7 Cramer's Rule

7.1 Content

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

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

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

7.2 Proof

Given that $Ax = c$, then

$$A^{-1}Ax = A^{-1}c = Ix = x \tag{29}$$

We are interested in $\boldsymbol{det}A_i$, that is

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

With the alternative feature of determinant, we have

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

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

Then the equation in (6) gives

$$\sum_{k=1}^{n} x_k \ \boldsymbol{det}(a_1, a_2, \dots, a_{i-1}, a_k, a_{i+1}, \dots, a_{n-1}, a_n) = x_i \ \boldsymbol{det}(a_1, a_2, \dots, a_n)$$

$$\boldsymbol{det}A_i(c) = x_i \ \boldsymbol{det}A \quad \text{QED} \tag{33}$$

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 jth column of A^{-1} and I respectively. Then we treat every column of A^{-1} as the solution to equation

$$A\boldsymbol{x} = e_j \tag{35}$$

Applying Cramer's rule,

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

$$\boldsymbol{x} = \frac{1}{\boldsymbol{det}A} \begin{bmatrix} C_{j,1} & C_{j,2} & \cdots C_{j,n} \end{bmatrix}^T \tag{37}$$

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

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

The matrix $\mathbf{adj}\boldsymbol{A}$ 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.

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

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

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

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

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

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

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

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

9 Eigenvalue and Eigenvector

9.1 Definition

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

$$\exists\ \lambda, \quad A\boldsymbol{x} = \lambda\boldsymbol{x} \text{ has non-trivial solution} \tag{44}$$

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

9.2 Linearity of Eigenvectors

Theorem 8.2.1 if $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the mutually different eigenvalues of matrix A, and corresponds to eigenvectors $v_1, v_2, \ldots, v_p$ respectively, then

$$\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\} \tag{45}$$

is a set of linearly independent vectors.
Proof If the set is linearly dependent, we obtain

$$\exists\ \boldsymbol{v_r} \in \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}, \quad \sum_{k=1}^{r} c_k \boldsymbol{v_k} = \boldsymbol{v_{r+1}} \tag{46}$$

Then

$$\begin{aligned} A\sum_{k=1}^{r} c_k \boldsymbol{v_k} - \lambda_{r+1}\sum_{k=1}^{r} c_k \boldsymbol{v_k} &= A\boldsymbol{v_{r+1}} - \lambda_{r+1}\boldsymbol{v_{r+1}} \\ \sum_{k=1}^{r} c_k(\lambda_k - \lambda_{r+1})\boldsymbol{v_k} &= \mathbf{0} \end{aligned} \tag{47}$$

As $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}$ 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

$$A\boldsymbol{x} = \lambda\boldsymbol{x} \tag{48}$$

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

$$A^{n+1}\boldsymbol{x} = A(A^n\boldsymbol{x}) = A(\lambda^n\boldsymbol{x}) = \lambda^n(A\boldsymbol{x}) = \lambda^{n+1}\boldsymbol{x} \tag{49}$$

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.

$$\begin{aligned} &\begin{bmatrix} A\boldsymbol{x_1} & A\boldsymbol{x_2} & \cdots & A\boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{x_1} & \lambda_2\boldsymbol{x_2} & \cdots & \lambda_n\boldsymbol{x_n} \end{bmatrix} \\ \Leftrightarrow\ &A\begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \cdots & \boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \cdots & \boldsymbol{x_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \end{aligned} \tag{50}$$

Let Col $P = \boldsymbol{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$, and the eigenvalues be the diagonal entries of diagonal matrix Λ. Evidently, P is invertible. Thus

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

The exponent of A can thus be interpreted as

$$\boxed{A^k = P\Lambda^k P^{-1}} \tag{52}$$

9.4 Criterion of Diagonalizability

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

$$1 \leq \text{dim}E_{\lambda_k} \leq m_k \tag{53}$$

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

$$E_{\lambda_k} = \boldsymbol{Span}\{v_1, v_2, \dots, v_p\}, \quad E_{\lambda_k} \in V \tag{54}$$

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

$$\{v_1, v_2, \dots, v_p\} \to \{v_1, v_2, \dots, v_p, \dots v_n\} = \beta \tag{55}$$

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

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

The characteristic polynomial of A is

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

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

$$m_k \geq p = \text{dim}E_{\lambda_k} \tag{58}$$

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

$$\text{dim}E_{\lambda_i} = m_i \tag{59}$$

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 $\boldsymbol{Span}(\beta \cap \beta_{\lambda_i})$ as n_i. Then from **8.4.1** we have

$$d_i \leq m_i \tag{60}$$

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

$$n_i \leq d_i \tag{61}$$

Thus summing up both sides we obtain

$$\sum_{i=1}^{n} n_i \leq \sum_{i=1}^{n} d_i \leq \sum_{i=1}^{n} m_i \tag{62}$$

For the LHS,

$$\sum_{i=1}^{n} n_i = \dim \boldsymbol{Span}\Big(\beta \cap (\bigcup_{i=1}^{p} \beta_{\lambda_i})\Big) = n \tag{63}$$

while obviously for the RHS,

$$\sum_{i=1}^{n} m_i = n \tag{64}$$

Thus

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

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

② **Necessity** Given that

$$\dim E_{\lambda_i} = d_i = m_i \tag{66}$$

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

$$\beta_{\lambda_i} \cap \beta_{\lambda_j} = \varnothing \quad \text{for} \quad i \neq j \tag{67}$$

Hence

$$\dim \boldsymbol{Span} \bigcup_{i=1}^{p} \beta_{\lambda_i} = \sum_{i=1}^{p} m_i = n \tag{68}$$

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

10 Eigenvectors from Eigenvalues - Terence Tao

It is proved that

$$|\boldsymbol{v_{ij}}|^2 \prod_{k=1;k\neq i}^{n} (\lambda_i(A) - \lambda_k(A)) = \prod_{k=1}^{n-1} (\lambda_i(A) - \lambda_k(M_j)) \tag{69}$$

Where A is an $n \times n$ matrix, $\boldsymbol{v_{ij}}$ 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.

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

Lemma 1

$$\boldsymbol{adj}(A)A = A\boldsymbol{adj}(A) = \boldsymbol{det}(A)I_n \tag{71}$$

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

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

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

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

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

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

$$A\boldsymbol{v_k} = \sum_{i=1}^{n} \lambda_i(A)\boldsymbol{v_i}\boldsymbol{v_i}^T\boldsymbol{v_k} = \lambda_k\boldsymbol{v_k}\boldsymbol{v_k}^T\boldsymbol{v_k} = \lambda_k\boldsymbol{v_k} \tag{74}$$

Notice that

$$\boldsymbol{adj}(A)A\boldsymbol{v_k} = \boldsymbol{det}(A)I_n\boldsymbol{v_k} = \lambda_k\boldsymbol{adj}(A)\boldsymbol{v_k} \rightarrow \boldsymbol{adj}(A)\boldsymbol{v_k} = \frac{\boldsymbol{det}(A)}{\lambda_k}\boldsymbol{v_k} \tag{75}$$

where

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

Substitute in we obtain

$$\boldsymbol{adj}(A)\boldsymbol{v_k} = \prod_{i=1;i\neq k}\lambda_i\boldsymbol{v_k} \tag{77}$$

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

$$\boldsymbol{adj}(A) = \sum_{i=1}^{n}\prod_{k=1;k\neq i}^{n}\lambda_k\boldsymbol{v_i}\boldsymbol{v_i}^T \tag{78}$$

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

$$\boldsymbol{adj}(\lambda_p I_n - A) = \sum_{i=1}^{n}\prod_{k=1;k\neq i}^{n}(\lambda_p - \lambda_k(A))\boldsymbol{v_i}\boldsymbol{v_i}^T \tag{79}$$

11 Markov Chain

11.1 Concept

Definition ① A probability vector is defined as

$$\{\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^T | \sum_{i=1}^{n}x_i = 1\} \tag{80}$$

② A stochastic matrix is defined as

$$\{M | M = \begin{bmatrix}\boldsymbol{p_1} & \boldsymbol{p_2} & \cdots & \boldsymbol{p_n}\end{bmatrix}\} \tag{81}$$

with the vectors $\boldsymbol{p_1}, \boldsymbol{p_2}, \ldots, \boldsymbol{p_n}$ all being probability vectors.

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

$$\boldsymbol{u_{k+1}} = M\boldsymbol{u_k}, \quad \boldsymbol{u_k} = M^k\boldsymbol{u_0} \tag{82}$$

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

$$m_{i,j} = \sum_{k=1}^{n} m_{i,k} \cdot m_{k,j} \tag{83}$$

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

$$A - \lambda I \Big|_{\lambda=1} = \begin{bmatrix} \boldsymbol{p_1} - \boldsymbol{e_1} & \boldsymbol{p_2} - \boldsymbol{e_2} & \cdots & \boldsymbol{p_n} - \boldsymbol{e_n} \end{bmatrix} \tag{84}$$

As

$$\sum_{i=1}^{n} (\boldsymbol{p_k} - \boldsymbol{e_k})_i = \sum_{i=1}^{n} (\boldsymbol{p_k})_i - 1 = 0 \tag{85}$$

Thus,

$$\sum_{i=1}^{n} (\boldsymbol{p_i} - \boldsymbol{e_i})^T = \mathbf{0} \tag{86}$$

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

$$\boldsymbol{det}(A - I) = 0, \quad \text{QED} \tag{87}$$

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

$$\boldsymbol{u_0} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}. \tag{88}$$

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

$$\boldsymbol{u_0} = P[\boldsymbol{u_0}]_{\mathcal{M}} \tag{89}$$

Hence

$$\begin{aligned} \boldsymbol{u_k} &= P\Lambda^k P^{-1} \boldsymbol{u_0} \\ &= P\Lambda^k P^{-1} P[\boldsymbol{u_0}]_{\mathcal{M}} = P\Lambda^k [\boldsymbol{u_0}]_{\mathcal{M}} \\ &= \begin{bmatrix} \boldsymbol{v_1} & \boldsymbol{v_2} & \cdots & \boldsymbol{v_n} \end{bmatrix} \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\ &= \sum_{i=1}^{n} c_i \lambda_i^k \boldsymbol{v_i} \end{aligned} \tag{90}$$

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

$$\boldsymbol{\pi} = \lim_{k\to\infty} \boldsymbol{u_k} = c_e \boldsymbol{v_e} \tag{91}$$

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 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ on v, if a real number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ follows the axioms below

$$\begin{aligned}
&\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle \\
&\langle \boldsymbol{u} + \boldsymbol{w}, \boldsymbol{v} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{w}, \boldsymbol{v} \rangle \\
&\langle c\boldsymbol{u}, \boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle \\
&\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0 \\
&\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0 \ \text{ iff } \boldsymbol{u} = \boldsymbol{0}
\end{aligned} \tag{92}$$

Then we call V an inner product space.
Definition The norm of vector is defined as

$$\|\boldsymbol{v}\|^2 = \langle \boldsymbol{v}, \boldsymbol{v} \rangle \tag{93}$$

Inner Product in $\mathbb{R}^n$

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

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \rightarrow \quad \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{v}^T \boldsymbol{u} = \sum_{k=1}^{n} u_k v_k \tag{94}$$

Inner Product in $\mathbb{P}_n$

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

$$\mathcal{P} = \{1, t, t^2, \ldots, t^n\} \tag{95}$$

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

$$p(t) = \mathcal{P}[p]_{\mathcal{P}} = \sum_{k=1}^{n} c_k t^k \tag{96}$$

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

$$\boldsymbol{p} = \begin{bmatrix} p(t_0) \\ p(t_1) \\ \vdots \\ p(t_n) \end{bmatrix} \tag{97}$$

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

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \sum_{k=1}^{n} p(t_k)q(t_k) \tag{98}$$

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 $\boldsymbol{f}, \boldsymbol{g}$ 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

$$\max\{t_1, t_2, \ldots, t_n\} = t_k \to 0 \tag{99}$$

Thus the inner product is

$$\begin{aligned} \langle \boldsymbol{f}, \boldsymbol{g} \rangle &= \lim_{n\to\infty} \frac{1}{n+1} \sum_{j=0}^{n} f(t_j)g(t_j) \\ &= \frac{1}{b-a} \lim_{n\to\infty} \frac{b-a}{n+1} \sum_{j=0}^{n} f\left(\frac{j(b-a)}{n+1}\right) g\left(\frac{j(b-a)}{n+1}\right) \\ &= \frac{1}{b-a} \int_a^b f(t)g(t)\mathrm{d}t \end{aligned} \tag{100}$$

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

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_a^b f(t)g(t)\mathrm{d}t \tag{101}$$

12.2 Orthogonality

Definition

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

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \tag{102}$$

Definition

A set $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ is called an orthogonal set if and only if

$$\langle \boldsymbol{u_i}, \boldsymbol{u_j} \rangle \Big|_{i \neq j} = 0 \tag{103}$$

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 $\beta = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_p}\}$, if it is linear dependent, then

$$\exists\ \boldsymbol{u_k} \in \beta, \boldsymbol{u_k} = \sum_{i=1}^{k-1} c_i \boldsymbol{u_i}$$

$$\text{For } 1 \le j \le k-1,\quad \langle \boldsymbol{u_k}, \boldsymbol{u_j} \rangle = \langle \sum_{i=1}^{k-1} c_i \boldsymbol{u_i}, \boldsymbol{u_j} \rangle = c_j \langle \boldsymbol{u_j}, \boldsymbol{u_j} \rangle = 0$$

As $\boldsymbol{u_j}$ is a non-zero vector, thus

$$\langle \boldsymbol{u_j}, \boldsymbol{u_j} \rangle \ge 0 \quad \to c_j = 0 \tag{104}$$

which draws the conclusion that

$$\boldsymbol{u_k} = \boldsymbol{0} \tag{105}$$

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

$$\beta = \left\{ \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_p}\} \middle| \langle \boldsymbol{u_i}, \boldsymbol{u_j} \rangle = 0,\ \|\boldsymbol{u_k}\| = 1 \text{ for } 1 \le i, j, k \le p,\ i \ne j \right\} \tag{106}$$

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

$$\begin{aligned} U^T U &= \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_p} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_p} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u_1^T u_1} & \boldsymbol{u_1^T u_2} & \cdots & \boldsymbol{u_1^T u_p} \\ \boldsymbol{u_2^T u_1} & \boldsymbol{u_2^T u_2} & \cdots & \boldsymbol{u_2^T u_p} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u_p^T u_1} & \boldsymbol{u_p^T u_2} & \cdots & \boldsymbol{u_p^T u_p} \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{u_1}\|^2 & & & \\ & \|\boldsymbol{u_2}\|^2 & & \\ & & \ddots & \\ & & & \|\boldsymbol{u_p}\|^2 \end{bmatrix} = I \end{aligned} \tag{107}$$

12.3 Orthogonal Complement

Definition

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

$$\{\boldsymbol{w} | \boldsymbol{w} \in W, \forall \boldsymbol{v} \in V, \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0\} \tag{108}$$

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

$$\text{Col}A = \{\boldsymbol{b} | A\boldsymbol{x} = \boldsymbol{b}\}, \quad \text{Nul}A = \{\boldsymbol{x} | A\boldsymbol{x} = \boldsymbol{0}\} \tag{109}$$

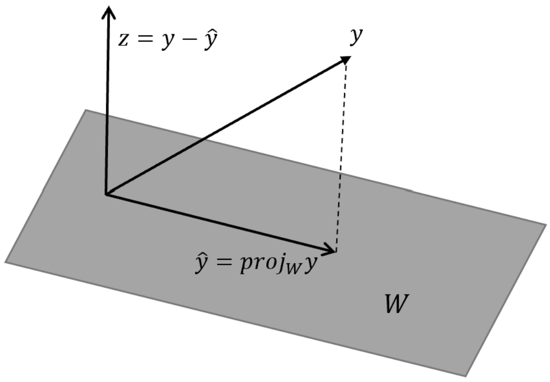
We can establish a relationship between this two subspace,

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

Thus we can conclude that ■

$$(\text{Row}A)^\perp = \text{Nul}A$$
$$(\text{Row}A^T)^\perp = (\text{Col}A)^\perp = \text{Nul}A^T$$

12.4 Orthogonal Projection



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

$$\boldsymbol{u} = \hat{\boldsymbol{u}} + \boldsymbol{z} = \alpha\boldsymbol{v} + \boldsymbol{z} = \text{proj}_{\boldsymbol{Span}\{\boldsymbol{v}\}}\boldsymbol{u} + \boldsymbol{z} \tag{110}$$

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

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \alpha\langle \boldsymbol{v}, \boldsymbol{v} \rangle + \langle \boldsymbol{z}, \boldsymbol{v} \rangle = \alpha\langle \boldsymbol{v}, \boldsymbol{v} \rangle \quad \to \quad \alpha = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v}.\boldsymbol{v} \rangle},\ \text{proj}_{\boldsymbol{Span}\{\boldsymbol{v}\}}\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v}.\boldsymbol{v} \rangle}\boldsymbol{v} \tag{111}$$

Corollary 10.4.1 For an inner product space W, if an orthogonal basis $w = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_p}\}$ sufficiently generates it, then if every single orthogonal basis corresponds to the columns of matrix P, then

$$\forall \boldsymbol{y} \in W, \quad \boldsymbol{y} = P[\boldsymbol{y}]_w = \sum_{i=1}^{p} c_i\boldsymbol{w_i}$$
$$\text{For } 1 \leq j \leq p,\ \langle \boldsymbol{w_j}, \boldsymbol{y} \rangle = \langle \boldsymbol{w_j}, \sum_{i=1}^{p} c_i\boldsymbol{w_i} \rangle = c_j\langle \boldsymbol{w_j}, \boldsymbol{w_j} \rangle \tag{112}$$

Thus

$$c_j = \frac{\langle \boldsymbol{w_j}, \boldsymbol{y} \rangle}{\langle \boldsymbol{w_j}, \boldsymbol{w_j} \rangle} \tag{113}$$

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

$$\beta = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_p}\},\ \boldsymbol{y} \notin W \ \rightarrow\ \hat{\boldsymbol{y}} = \sum_{i=1}^{p} \text{proj}_{\boldsymbol{Span\{w_i\}}} \boldsymbol{y} \tag{114}$$

A more generalized expression is:

$$\text{proj}_{\boldsymbol{Span\{w_1, w_2, \ldots, w_p\}}} \boldsymbol{y} = \sum_{i=1}^{p} \text{proj}_{\boldsymbol{Span\{w_i\}}} \boldsymbol{y} \tag{115}$$

12.5 Gram-Schmidt Orthogonalization

For an inner product space W, assume that $\alpha = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_p}\}$ is a basis of W. Then we generate an orthogonal basis $\beta = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$. Then the orthogonalization for a basis vector $\boldsymbol{u_k}$ is

$$\begin{aligned} \boldsymbol{v_k} &= \boldsymbol{u_k} - \text{proj}_{\boldsymbol{Span\{v_1, v_2, \ldots, v_{k-1}\}}} \boldsymbol{u_k} \\ &= \boldsymbol{u_k} - \sum_{i=1}^{k-1} \text{proj}_{\boldsymbol{Span\{v_i\}}} \boldsymbol{u_k} \\ &= \boldsymbol{u_k} - \sum_{i=1}^{n} \frac{\langle \boldsymbol{u_k}, \boldsymbol{v_i} \rangle}{\langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle} \boldsymbol{v_i} \end{aligned} \tag{116}$$

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 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we tend to obtain a generalized form of the solution

$$y_j = \sum_{i=0}^{k} \beta_i f_i(x_j) \tag{117}$$

Or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_2) & \cdots & f_k(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \tag{118}$$

Which can be written as

$$\boldsymbol{y} = X\boldsymbol{\beta} \tag{119}$$

$\boldsymbol{y}$ is the **sample vector**. X is prediction value vector based on every x, called the **design matrix**. $\boldsymbol{\beta}$ 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 Col$X$ and y. Then we write (102) as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \;\Leftrightarrow\; \boldsymbol{\epsilon} = \boldsymbol{y} - X\boldsymbol{\beta} \tag{120}$$

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

$$X\boldsymbol{\beta} = \hat{\boldsymbol{y}} = \text{proj}_{\text{Col}X}\boldsymbol{y} \tag{121}$$

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

$$X^T(\boldsymbol{y} - \hat{\boldsymbol{y}}) = X^T(\boldsymbol{y} - X\boldsymbol{\beta}) = \boldsymbol{0} \quad \rightarrow \quad \boxed{X^T X\boldsymbol{\beta} = X^T\boldsymbol{y}} \tag{122}$$

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.

$$\boldsymbol{\epsilon}' = \begin{bmatrix} w_1(y_1 - \hat{y_1}) \\ w_2(y_2 - \hat{y_2}) \\ \vdots \\ w_n(y_n - \hat{y_n}) \end{bmatrix} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix} \begin{bmatrix} y_1 - \hat{y_1} \\ y_2 - \hat{y_2} \\ \vdots \\ y_n - \hat{y_n} \end{bmatrix} = W\boldsymbol{\epsilon} \tag{123}$$

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

$$WX\boldsymbol{\beta} = W\boldsymbol{y} \quad \rightarrow \quad (WX)^T(WX)\boldsymbol{\beta} = (WX)^T\boldsymbol{y} \tag{124}$$

13.3.2 Lagrange Interpolation

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

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_{n-1}x^{n-1} \tag{125}$$

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

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{126}$$

Applying Cramer's rule,

$$\begin{aligned} \beta_i = \frac{\boldsymbol{det}X_i(\boldsymbol{y})}{\boldsymbol{det}X} &= \frac{1}{\prod_{1\le i<j\le n}(x_j - x_i)} \begin{vmatrix} 1 & x_1 & \cdots & x_1^{i-2} & y_1 & x_1^i & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{i-2} & y_2 & x_2^i & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{i-2} & y_n & x_n^i & \cdots & x_n^{n-1} \end{vmatrix} \\ &= \frac{1}{\prod_{1\le i<j\le n}(x_j - x_i)} \sum_{j=1}^{n} y_j C_{j,i} = \frac{1}{\prod_{1\le i<j\le n}(x_j - x_i)} \sum_{j=1}^{n} (-1)^{i+j} y_j A_{j,i} \end{aligned} \tag{127}$$

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

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

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

Actually the polynomial can be written into a simplified form that

$$f(x) = \sum_{i=0}^{n-1} \beta_i x^i = \sum_{i=1}^{n} y_i \prod_{\substack{1\le j\le n \\ j\neq i}} \frac{(x - x_j)}{(x_i - x_j)} \tag{128}$$

13.3.3 Fourier Series

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

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

is an orthogonal set under inner product

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

Proof

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

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

$$f(t) = a_0 + \sum_{m=1}^{n} (a_m \text{cos}mt + b_m \text{sin}mt) \tag{129}$$

where

$$a_m = \frac{\langle f, \text{cos}mt\rangle}{\langle \text{cos}mt, \text{cos}mt\rangle}, \quad b_m = \frac{\langle f, \text{sin}mt\rangle}{\langle \text{sin}mt, \text{sin}mt\rangle} \tag{130}$$

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

14 Jacobian Matrix

If there are two coordinate systems with mapping relationship between basis

$$\alpha = \{x_1, x_2, \ldots, x_n\} \xrightarrow{f:\ \mathbb{R}^n \mapsto \mathbb{R}^n} \beta = \{y_1, y_2, \ldots, y_n\} \tag{131}$$

$$\text{or} \quad y_i = f_i(x_1, x_2, \ldots, x_n) \tag{132}$$

Then the relationship between the infinitesimals correspond to the basis are

$$\mathrm{d}\boldsymbol{y_i} = \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j} \mathrm{d}\boldsymbol{x_j} \tag{133}$$

or

$$\begin{bmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \vdots \\ \mathrm{d}y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial f_1}{\partial x_i} \mathrm{d}x_i \\ \sum_{i=1}^{n} \frac{\partial f_2}{\partial x_i} \mathrm{d}x_i \\ \ddots \\ \sum_{i=1}^{n} \frac{\partial f_n}{\partial x_i} \mathrm{d}x_i \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \\ \vdots \\ \mathrm{d}x_n \end{bmatrix} \tag{134}$$

Lemma 1 Assume that a geometry body in the space spanned by basis $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$, and the linear transformation depicted by $n \times n$ matrix A transforms the basis into $\{y_1, y_2, \ldots, y_n\}$, i.e.

$$A \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \cdots & \boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y_1} & \boldsymbol{y_2} & \cdots & \boldsymbol{y_n} \end{bmatrix} \tag{135}$$

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

$$V_{\mathcal{Y}} = |\boldsymbol{det}A| V_{\mathcal{X}} \tag{136}$$

Proof For each vector in the original space, we have

$$\boldsymbol{u} = c_1 \boldsymbol{x_1} + c_2 \boldsymbol{x_2} + \cdots + c_n \boldsymbol{x_n} \tag{137}$$

Then the magnitude to the geometry body depicted by vectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ is

$$\boldsymbol{det} \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix} \tag{138}$$

For the transformed vector, we have

$$\boldsymbol{u}' = c_1 \boldsymbol{y_1} + c_2 \boldsymbol{y_2} + \cdots + c_n \boldsymbol{y_n} = c_1 A \boldsymbol{x_1} + c_2 A \boldsymbol{x_2} + \cdots + c_n A \boldsymbol{x_n} = A(c_1 \boldsymbol{x_1} + c_2 \boldsymbol{x_2} + \cdots + c_n \boldsymbol{x_n}) \tag{139}$$

Then the new geometry body has the magnitude

$$\begin{aligned} \boldsymbol{det} \begin{bmatrix} \boldsymbol{u_1'} & \boldsymbol{u_2'} & \cdots & \boldsymbol{u_n'} \end{bmatrix} &= \boldsymbol{det} \begin{bmatrix} A\boldsymbol{u_1} & A\boldsymbol{u_2} & \cdots & A\boldsymbol{u_n} \end{bmatrix} = \boldsymbol{det}A \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix} \\ &= (\boldsymbol{det}A)(\boldsymbol{det} \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix}) \end{aligned}$$

■

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

$$\Omega_y = \left| \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix} \right| \Omega_x = |\boldsymbol{det}\mathbb{J}| \Omega_x \tag{140}$$

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

$$A = A^T \tag{141}$$

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 $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}$ are eigenvectors of symmetric matrix A, and correspond to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then

$$\begin{gathered} \forall\ \lambda_i, \lambda_j, \quad A\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i},\ A\boldsymbol{v_j} = \lambda_j \boldsymbol{v_j} \\ \lambda_i \boldsymbol{v_i}^T = \boldsymbol{v_j}^T A^T \ \rightarrow\ \lambda_i \boldsymbol{v_i}^T \boldsymbol{v_j} = \boldsymbol{v_i}^T A^T \boldsymbol{v_j} = \boldsymbol{v_i}^T A \boldsymbol{v_j} = \lambda_j \boldsymbol{v_i}^T \boldsymbol{v_j} \end{gathered} \tag{142}$$

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

$$\boldsymbol{v_i}^T \boldsymbol{v_j} = \langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = 0 \tag{143}$$

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.

$$A = P\Lambda P^{-1} \tag{144}$$

With the eigenvectors forming the columns of matrix P. If we unitize each eigenvector to obtain an orthonormal basis $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$, $\boldsymbol{u_i} = \boldsymbol{v_i}/\sqrt{\langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle}$, then we denote the new matrix as Q, where $QQ^T = 1$. We have

$$\boxed{A = Q\Lambda Q^{-1} = Q\Lambda Q^T} \ \rightarrow\ A^T = (Q\lambda Q^T)^T = Q\Lambda Q^T = A \tag{145}$$

Thus the diagonalization is well-defined.

15.4 Spectral Decomposition

The diagonalization above gives

$$A = \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \vdots \\ \boldsymbol{u_n}^T \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u_i} \boldsymbol{u_i}^T \tag{146}$$

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 $(A^TA)^T = A^T(A^T)^T = A^TA$, which indicates that A^TA itself is a symmetric matrix. Then we apply EVD to A^TA.

Assume that $A^TA = Q\Lambda Q^T$, with $\boldsymbol{rank}(A) = k$, so the eigenvectors set $q = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ forms an orthonormal basis for Q, then

$$\forall\ \boldsymbol{v_1}, \boldsymbol{v_2}\ \in\ q,\ \ \langle A\boldsymbol{v_i}, A\boldsymbol{v_j}\rangle = (A\boldsymbol{v_j})^T A\boldsymbol{v_i} = \boldsymbol{v_j}^T A^T A\boldsymbol{v_i} = \boldsymbol{v_j}^T \lambda_i \boldsymbol{v_i} = \lambda_i \boldsymbol{v_j}^T \boldsymbol{v_i} = \lambda_i \langle \boldsymbol{v_i}, \boldsymbol{v_j}\rangle = 0$$

So the set $q' = \{A\boldsymbol{v_1}, A\boldsymbol{v_2}, \ldots, A\boldsymbol{v_k}\}$ is also orthogonal. Assume that the vectors in q forms the columns of matrix V, then

$$AV = A\begin{bmatrix}\boldsymbol{v_1} & \boldsymbol{v_2} & \cdots & \boldsymbol{v_k}\end{bmatrix} = \begin{bmatrix}A\boldsymbol{v_1} & A\boldsymbol{v_2} & \cdots & A\boldsymbol{v_k}\end{bmatrix} \tag{147}$$

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

$$A\boldsymbol{v_i} = \sigma_i \boldsymbol{u_i}\ \leftrightarrow\ A\begin{bmatrix}\boldsymbol{v_1} & \boldsymbol{v_2} & \cdots & \boldsymbol{v_k}\end{bmatrix} = \begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k\end{bmatrix}\begin{bmatrix}\boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_k}\end{bmatrix} \tag{148}$$

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

$$AV = U\Sigma\ \rightarrow\ \boxed{A = U\Sigma V^T = \sum_{i=1}^{k} \sigma_i \boldsymbol{u_i}\boldsymbol{v_i}^T} \tag{149}$$

The product on the RHS shows the **singular value decomposition** of an arbitrary matrix A. If we expand the set $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ and $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ to form a basis of $\mathbb{R}^n$ and $\mathbb{R}^n$ respectively, we then obtain

$$A = \left[\begin{array}{cccc|ccc}\boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_k} & \boldsymbol{u_{k+1}} & \cdot & \boldsymbol{u_n}\end{array}\right]\left[\begin{array}{cccc|c}\sigma_1 & & & & \\ & \sigma_2 & & & \mathbf{0} \\ & & \ddots & & \\ & & & \sigma_k & \\ \hline & & \mathbf{0} & & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\boldsymbol{v_1}^T \\ \boldsymbol{v_2}^T \\ \vdots \\ \boldsymbol{v_k}^T \\ \hline \boldsymbol{v_{k+1}}^T \\ \vdots \\ \boldsymbol{v_m}^T\end{array}\right] \tag{150}$$

15.5.2 Application