

# Linear Algebra – Part II

A summary for MIT 18.06SC

Jiaxin Hu

July 8, 2020

## 1 Orthogonality

### 1.1 Orthogonal vectors and subspaces

**Definition 1** (*Orthogonal vectors*). Suppose two vectors  $x, y \in \mathbb{R}^n$ . The vectors  $x$  and  $y$  are orthogonal **iff**  $x^T y = y^T x = 0$ , denoted  $x \perp y$ .

**Definition 2** (*Orthogonal subspaces*). Suppose two subspaces  $S, T$ . The subspaces  $S$  and  $T$  are orthogonal **iff** for any  $s \in S$  and for any  $t \in T$ ,  $s^T t = t^T s = 0$ , denoted  $S \perp T$ .

Given a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , there are four subspaces related to  $\mathbf{A}$ : column space  $C(\mathbf{A})$ , row space  $C(\mathbf{A}^T)$ , nullspace  $N(\mathbf{A})$ , and left nullspace  $N(\mathbf{A}^T)$ . Suppose the matrix rank of  $\mathbf{A}$  is  $\text{rank}(\mathbf{A}) = r$ , the dimensions of these subspaces are:

$$\dim(C(\mathbf{A})) = \dim(C(\mathbf{A}^T)) = r, \quad \dim(N(\mathbf{A})) = n - r, \quad \dim(N(\mathbf{A}^T)) = m - r.$$

**Theorem 1.1** (Orthogonality of matrix subspaces). Suppose a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . The row space  $C(\mathbf{A}^T)$  and the nullspace  $N(\mathbf{A})$  are orthogonal; the column space  $C(\mathbf{A})$  and the left nullspace  $N(\mathbf{A}^T)$  are orthogonal i.e.

$$C(\mathbf{A}^T) \perp N(\mathbf{A}); \quad C(\mathbf{A}) \perp N(\mathbf{A}^T).$$

*Proof.* Consider the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . For any vector  $x \in N(\mathbf{A})$ , we have  $\mathbf{A}x = 0$ .

$$\mathbf{A}x = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

where  $a_i \in \mathbb{R}^n, i \in [m]$  is the  $i$ -th row of  $\mathbf{A}$ . By the definition 1,  $x$  is orthogonal with the rows in matrix  $\mathbf{A}$ . For any vector  $v \in C(\mathbf{A}^T)$ ,  $v$  is a linear combination of the rows, i.e.  $v = c_1 a_1 + \dots + c_m a_m$ , where  $c_i, i \in [m]$  are constants. Multiplying vectors  $v$  and  $x$ ,

$$v^T x = c_1 a_1^T x + \dots + c_m a_m^T x = 0.$$

Therefore,  $v \perp x$ , and  $N(\mathbf{A}) \perp C(\mathbf{A}^T)$ .

Similarly, for any  $x \in N(\mathbf{A}^T)$ , we have  $\mathbf{A}^T x = 0$ , which implies  $N(\mathbf{A}^T) \perp C(\mathbf{A})$ . □

**Theorem 1.2** (Relationship between  $\mathbf{A}^T \mathbf{A}$  and  $\mathbf{A}$ ). Consider a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . We have

$$N(\mathbf{A}^T \mathbf{A}) = N(\mathbf{A}); \quad \text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}).$$

*Proof.* Consider the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .

If  $x \in N(\mathbf{A})$ , then  $\mathbf{A}x = 0 \Rightarrow \mathbf{A}^T 0 = 0$ , which implies that for any  $x \in N(\mathbf{A})$ , the vector  $x \in N(\mathbf{A}^T \mathbf{A})$ . Therefore, we only need to prove that for any  $x \in N(\mathbf{A}^T \mathbf{A})$ , the vector  $x \in N(\mathbf{A})$ . We prove this by contradiction.

Suppose a vector  $x \in N(\mathbf{A}^T \mathbf{A})$  but  $x \notin N(\mathbf{A})$ . We have

$$\mathbf{A}x = b \neq 0, \quad \mathbf{A}^T \mathbf{A}x = 0 \quad \Rightarrow \quad \mathbf{A}^T b = 0.$$

By the first equation  $b \in C(\mathbf{A})$ , and by the third equation  $b \in N(\mathbf{A}^T)$ . This contradicts the theorem 1.1, i.e.  $C(\mathbf{A}) \perp N(\mathbf{A}^T)$ .

Next, given  $N(\mathbf{A}^T \mathbf{A}) = N(\mathbf{A})$ , the rank of matrix  $\text{rank}(\mathbf{A}^T \mathbf{A}) = n - \dim(N(\mathbf{A}^T \mathbf{A})) = n - \dim(N(\mathbf{A})) = \text{rank}(\mathbf{A})$ .  $\square$

**Corollary 1** (Invertibility of  $\mathbf{A}^T \mathbf{A}$ ). If  $\mathbf{A}$  has independent columns, then  $\mathbf{A}^T \mathbf{A}$  is invertible.

*Proof.* If  $\mathbf{A} \in \mathbb{R}^{m \times n}$  has independent columns, then  $\text{rank}(\mathbf{A}) = n$ . By the theorem 1.2,  $\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}) = n$ . Since  $\mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$  is a square matrix,  $\mathbf{A}^T \mathbf{A}$  is invertible.  $\square$

## 1.2 Projections onto subspaces

**Definition 3** (*Projection and projection matrix*). Consider a vector  $x \in \mathbb{R}^m$  and a matrix  $\mathbf{A}^{m \times n}$  that has independent columns. Suppose a vector  $p \in C(\mathbf{A})$ , such that

$$(x - p) \perp C(\mathbf{A}). \quad (1)$$

The vector  $p$  is the projection of vector  $x$  onto the space  $C(\mathbf{A})$ . Since  $p \in C(\mathbf{A})$ , there exists a vector  $\hat{x}$  such that  $p = \mathbf{A}\hat{x}$ . By equation (1), we have

$$\mathbf{A}^T(x - p) = \mathbf{A}^T(x - \mathbf{A}\hat{x}) \Rightarrow \hat{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x, \quad p = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x.$$

The matrix  $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  is called projection matrix.

**Theorem 1.3** (Properties of projection matrix). Consider a projection matrix  $\mathbf{P}$  to the column space  $C(\mathbf{A})$ , where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is a matrix. Then,

$$\mathbf{P}^T = \mathbf{P}; \quad \mathbf{P}^2 = \mathbf{P}.$$

*Proof.* Since the projection matrix  $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  and  $\mathbf{A}^T \mathbf{A}$  is symmetric, then

$$\mathbf{P}^T = (\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T)^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{P}.$$

$$\mathbf{P}^2 = \mathbf{P}^T \mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{P}.$$

$\square$

**Corollary 2** (Projection onto  $N(\mathbf{A}^T)$ ). Suppose  $\mathbf{P}$  is a projection matrix in theorem 1.3, then  $I - \mathbf{P}$  is also a projection matrix to the left nullspace  $N(\mathbf{A}^T)$ .

*Proof.* For any  $x \in \mathbb{R}^m$ , we have  $x - \mathbf{P}x \perp C(\mathbf{A})$

$$x - \mathbf{P}x \perp C(\mathbf{A}) \Rightarrow (I - \mathbf{P})x \perp C(\mathbf{A}) \Rightarrow (I - \mathbf{P})x \in N(\mathbf{A}^T) \text{ and } (x - (I - \mathbf{P})x) \perp N(\mathbf{A}^T).$$

Therefore,  $I - \mathbf{P}$  is a projection matrix to the left nullspace  $N(\mathbf{A}^T)$ .  $\square$

### 1.3 Projection matrices and least squares

Given observation vector  $y \in \mathbb{R}^n$  and the design matrix  $\mathbf{X} \in \mathbb{R}^{n \times (k+1)}$ , we propose the linear regression model

$$y = \mathbf{X}\beta + \epsilon,$$

where  $\beta = (\beta_0, \beta_1, \dots, \beta_k)$  are coefficients of our interests and  $\epsilon$  is the noise. The least square estimate of  $\beta$  minimizes the loss

$$\hat{\beta}_{LS} = \arg \min_{\beta \in \mathbb{R}^{k+1}} \|y - \mathbf{X}\beta\|^2,$$

where  $\|\cdot\|$  is the euclidean norm. The vector  $\mathbf{X}\hat{\beta}_{LS}$  can be considered as a the projection of  $y$  onto the column space of  $\mathbf{X}$ . Therefore, we may use projection tools to solve the least square estimate. The projection  $\mathbf{X}\hat{\beta}_{LS}$  satisfies

$$\mathbf{X}^T(y - \mathbf{X}\hat{\beta}_{LS}) = 0 \quad \Rightarrow \quad \hat{\beta}_{LS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y.$$

The estimate  $\hat{\beta}_{LS}$  is identical to the estimates solved by other methods using the derivative.

### 1.4 Orthogonal matrices and Gram-Schmidt

**Definition 4** (*Orthonormal vectors*). The vectors  $q_1, \dots, q_n$  are orthonormal if

$$q_i^T q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

Orthonormal vectors are always independent.

**Definition 5** (*Orthonormal matrix and orthogonal matrix*). Consider a matrix  $\mathbf{Q} \in \mathbb{R}^{m \times n}$ . If the columns of  $\mathbf{Q}$  are orthonormal, the matrix  $\mathbf{Q}$  is an orthonormal matrix. If  $m = n$ , the square matrix  $\mathbf{Q}$  is a orthogonal matrix.

If  $\mathbf{Q} \in \mathbb{R}^{m \times n}$  is an orthonormal matrix,  $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n$ . If  $\mathbf{Q}$  is an orthogonal matrix,  $\mathbf{Q}^T = \mathbf{Q}^{-1}$ . For orthonormal matrix  $\mathbf{Q}$ , the projection matrix to  $C(\mathbf{P})$  becomes  $\mathbf{P} = \mathbf{I}_m$ .

**Definition 6** (*Gram-Schmidt Process and QR decomposition*). Consider a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  with  $\text{rank}(\mathbf{A}) = n$ . Gram-Schmidt process finds the orthonormal basis for  $C(\mathbf{A})$ . Let  $a_i \in \mathbb{R}^m, i \in [m]$  be the columns of matrix  $\mathbf{A}$ .

$$\begin{aligned} u_1 &= a_1, & e_1 &= \frac{u_1}{\|u_1\|} \\ u_2 &= a_2 - \frac{u_1^T a_2}{u_1^T u_1} u_1, & e_2 &= \frac{u_2}{\|u_2\|} \\ u_3 &= a_3 - \frac{u_1^T a_3}{u_1^T u_1} u_1 - \frac{u_2^T a_3}{u_2^T u_2} u_2, & e_3 &= \frac{u_3}{\|u_3\|} \\ &\vdots & & \end{aligned}$$

The vectors  $e_1, \dots, e_n$  are orthonormal basis of the  $C(\mathbf{A})$ . By matrix operations, we obtain a decomposition of matrix  $\mathbf{A}$

$$\mathbf{A} = [a_1, \dots, a_n] = [e_1, \dots, e_n] \begin{bmatrix} e_1^T a_1 & e_1^T a_2 & \dots & e_1^T a_n \\ 0 & e_2^T a_2 & \dots & e_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e_n^T a_n \end{bmatrix} = \mathbf{Q} \mathbf{R} \quad (2)$$

where  $\mathbf{Q} \in \mathbb{R}^{m \times n}$  is an orthonormal matrix and  $\mathbf{R} \in \mathbb{R}^{n \times n}$  is a upper triangular matrix. We call the matrix decomposition as equation (2) QR decomposition.

## 2 Determinants

The *determinant* is a number associated with any square matrix. For a square matrix  $\mathbf{A}$ , let  $\det(\mathbf{A})$  or  $|\mathbf{A}|$  denote the determinant of matrix  $\mathbf{A}$ .

### 2.1 Properties of determinants

We give ten properties of determinants. The last seven properties are deduced by first three basic properties

1. The determinant of identity matrix is equal to 1, i.e.  $\det(\mathbf{I}) = 1$ .
2. Exchanging two rows of a matrix reverses the sign of the matrix's determinant. Hence, an odd number of row exchanges reverse the sign of the determinant while an even number of row exchanges maintain the sign of the determinant.
3. (a) If one row of the matrix is multiplied by a constant  $t$ , the determinant of the matrix is multiplied by  $t$ .

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

- (b) The determinant is linearly additive on the rows of the matrix.

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

4. If two rows of a matrix are equal, the determinant of the matrix is 0.

*Proof.* Suppose  $\mathbf{A}$  has two identical rows  $a_i, a_j$ . Exchanging  $a_i, a_j$ , the exchanged matrix  $\mathbf{A}' = \mathbf{A}$ . By property 2,  $\det(\mathbf{A}) = -\det(\mathbf{A}') = -\det(\mathbf{A})$ . Therefore,  $\det(\mathbf{A}) = 0$ .  $\square$

5. For a square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , subtracting  $t$  times of the  $i$ -th row from the  $j$ -th row does not change the determinant of  $\mathbf{A}$ , where  $i \neq j$ .

*Proof.* By property 3, property 3, and property 4,

$$\begin{vmatrix} a & b \\ c - ta & d - tb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} - t \begin{vmatrix} a & b \\ a & b \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

The proof for higher-dimension matrix is similar.  $\square$

6. If the matrix  $\mathbf{A}$  has a row that is all zeros, the determinant of  $\mathbf{A}$  is 0.

*Proof.* By property 3, letting  $t = 0$  leads to property 6.  $\square$

7. Suppose  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is a triangular matrix with diagonal elements  $d_1, \dots, d_n$ . The determinant of  $\mathbf{A}$  is equal to the product  $d_1 \dots d_n$ .

*Proof.* By property 5, eliminating the triangular matrix  $\mathbf{A}$  to a diagonal matrix does not change the determinant of  $\mathbf{A}$ . By property 3 and 1,  $\det(\mathbf{A}) = d_1 \dots d_n \det(\mathbf{I}_n) = d_1 \dots d_n$ .  $\square$

8. If the square matrix  $\mathbf{A}$  is singular, the determinant of  $\mathbf{A}$  is 0.

*Proof.* If  $\mathbf{A}$  is a singular matrix, the reduced row echelon form of  $\mathbf{A}$ ,  $\text{RREF}(\mathbf{A})$ , has a row with all 0 elements. By property 5 and 6, the determinant of  $\mathbf{A}$  is 0.  $\square$

9. For square matrices  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ , the determinant of the product satisfies  $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ .

*Proof.* If at least one of  $\mathbf{A}, \mathbf{B}$  is singular, the product  $\mathbf{AB}$  is also singular. Then,  $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}) = 0$ . If both  $\mathbf{A}, \mathbf{B}$  are invertible,  $\mathbf{A}, \mathbf{B}$  can be eliminated to diagonal matrices  $\mathbf{Q}_A, \mathbf{Q}_B$  by QR decomposition. Suppose  $a_1, \dots, a_n$  and  $b_1, \dots, b_n$  are diagonal elements for  $\mathbf{Q}_A$  and  $\mathbf{Q}_B$  respectively. By property 7,  $\det(\mathbf{AB}) = \prod_{i=1}^n a_i b_i = \det(\mathbf{A}) \det(\mathbf{B})$ .  $\square$

10. For a square matrix  $\mathbf{A}$ , the determinant  $\det(\mathbf{A}^T) = \det(\mathbf{A})$ .

*Proof.* By LU decomposition, the matrix  $\mathbf{A} = \mathbf{LU}$ , where  $\mathbf{L}$  is a lower-triangular matrix and  $\mathbf{U}$  is an upper-triangular matrix. The transport  $\mathbf{L}^T, \mathbf{U}^T$  have the same diagonal elements with  $\mathbf{L}, \mathbf{U}$  respectively. By property 7,  $\det(\mathbf{L}^T) = \det(\mathbf{L})$ ,  $\det(\mathbf{U}^T) = \det(\mathbf{U})$ . Therefore, by property 9,  $\det(\mathbf{A}) = \det(\mathbf{L}) \det(\mathbf{U}) = \det(\mathbf{U}^T) \det(\mathbf{L}^T) = \det(\mathbf{U}^T \mathbf{L}^T) = \det(\mathbf{A}^T)$ .  $\square$

## 2.2 Determinant computation

**Proposition 1** (Big formula for computing determinant). *Consider a square matrix  $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ . The big formula for computing the determinant of  $\mathbf{A}$  states below.*

$$\det(\mathbf{A}) = \sum_{(\alpha_1, \dots, \alpha_n) \in \mathcal{P}} (-1)^{N(\alpha_1, \dots, \alpha_n)} a_{1, \alpha_1} \dots a_{n, \alpha_n},$$

where  $(\alpha_1, \alpha_2, \dots, \alpha_n)$  is the permutation of  $(1, 2, \dots, n)$  and  $\mathcal{P}$  is the collection of all possible  $(\alpha_1, \dots, \alpha_n)$ , and  $N(\alpha_1, \dots, \alpha_n)$  is the number of element exchanges from  $(\alpha_1, \alpha_2, \dots, \alpha_n)$  to  $(1, 2, \dots, n)$ .

*Proof.* For a square matrix  $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ , we decompose the matrix  $\mathbf{A}$  by the summation of  $n^n$  matrices, in which there is only one entry comes from  $\mathbf{A}$  in each row and other entries are 0. By property 3, the determinant of  $\mathbf{A}$  is equal to the summation of the determinants of these  $n^n$  matrices. Take a 3-by-3 square matrix as an example.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & 0 & 0 \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} + \dots + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} \triangleq \mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_{27}.$$

For the square matrix  $\mathbf{A}$ , there are  $n!$  non-singular matrices among the  $n^n$  matrices. The number of non-singular matrices coincides with the size of  $\mathcal{P}$ . Intuitively, there are  $n$  ways to choose an element from the first row, after which there are only  $n - 1$  ways to choose an element from the second row to avoid the zero determinant. Therefore, we have  $n \times (n - 1) \times (n - 2) \times \dots \times 2 = n!$  non-singular matrices.

By property 7 and 2, the determinant for the non-singular matrix follows the formula

$$(-1)^{N(\alpha_1, \dots, \alpha_n)} a_{1, \alpha_1} \dots a_{n, \alpha_n},$$

where  $(\alpha_1, \alpha_2, \dots, \alpha_n)$  is the permutation of  $(1, 2, \dots, n)$  and  $\mathcal{P}$  is the collection of all possible  $(\alpha_1, \dots, \alpha_n)$ , and  $N(\alpha_1, \dots, \alpha_n)$  is the number of element exchanges from  $(\alpha_1, \alpha_2, \dots, \alpha_n)$  to  $(1, 2, \dots, n)$ .  $\square$

**Definition 7** (Cofactors, cofactor matrix, and cofactor formula). Consider a square matrix  $\mathbf{A} = \llbracket a_{ij} \rrbracket \in \mathbb{R}^{n \times n}$ . Let  $\mathbf{A}_{-i, -j}$  be the submatrix of  $\mathbf{A}$  after removing the  $i$ -th row and  $j$ -th column. The cofactor associated with  $a_{ij}$  is

$$C_{ij} = (-1)^{i+j} \det(\mathbf{A}_{-i, -j}).$$

The matrix  $\mathbf{C} = \llbracket C_{ij} \rrbracket \in \mathbb{R}^{n \times n}$  is called cofactor matrix. The cofactor formula of  $\det(\mathbf{A})$  is

$$\det(\mathbf{A}) = \sum_j^n a_{ij} C_{ij} = \sum_j^n a_{ji} C_{ji}, \quad \forall i \in [n].$$

**Example 1** (Tridiagonal matrix). One example usage of cofactor formula is computing the determinant of *tridiagonal matrix*. A tridiagonal matrix is a matrix in which only non-zero elements lie on or adjacent to the diagonal. Let  $\mathbf{T}_n \in \mathbb{R}^{n \times n}$  denote the tridiagonal matrix of 1's, i.e.

$$\mathbf{T}_n = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}_{n \times n}$$

Let  $\mathbf{T}_{n, -i, -j}$  be the submatrix of  $\mathbf{T}_n$  after removing the  $i$ -th row and the  $j$ -th column. By cofactor formula,

$$\det(\mathbf{T}_n) = 1 \times \det(\mathbf{T}_{n, -1, -1}) - 1 \times \det(\mathbf{T}_{n, -1, -2}).$$

By the definition of tridiagonal matrix,  $\mathbf{T}_{n, -1, -1} = \mathbf{T}_{n-1}$  and  $\det(\mathbf{T}_{n, -1, -2}) = 1 \times \det(\mathbf{T}_{n-2})$ . Therefore,

$$\det(\mathbf{T}_n) = \det(\mathbf{T}_{n-1}) - \det(\mathbf{T}_{n-2}).$$

## 2.3 Inverse matrices

Previously, we use Gauss-Jordan elimination to obtain the inverse matrix of the invertible matrix. Here, we apply cofactor formula for computing the inverse matrix.

**Theorem 2.1** (Inverse matrix by cofactors). Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  be an invertible matrix and  $\mathbf{C}$  be the cofactor matrix of  $\mathbf{A}$ . The inverse matrix of  $\mathbf{A}$  satisfies

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T.$$

*Proof.* It is equivalent to show that  $\mathbf{A} \mathbf{C}^T = \det(\mathbf{A}) \mathbf{I}_n$ . By the definition of cofactor matrix,

$$\mathbf{A} \mathbf{C}^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

Therefore, the diagonal element of  $\mathbf{AC}^T$  satisfies  $(\mathbf{AC}^T)_{ii} = \sum_k^n a_{ik}C_{ik} = \det(\mathbf{A})$ . Consider the off-diagonal entries  $(\mathbf{AC}^T)_{ij} = \sum_k^n a_{ik}C_{jk}$ , where  $i \neq j$ . By cofactor formula, the summation  $\sum_k^n a_{ik}C_{jk}$  is equal to the determinant of a matrix whose  $i$ -th row and  $j$ -th row are identical. By property 6, the off-diagonals of  $\mathbf{AC}^T$  are 0. Then, we have  $\mathbf{AC}^T = \det(\mathbf{A})\mathbf{I}_n$ .  $\square$

**Definition 8** (Cramer's rule). Consider the linear system  $\mathbf{Ax} = b$ , where  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is an invertible matrix, and  $x, b \in \mathbb{R}^n$  are vectors. Applying theorem 2.1, we obtain the Cramer's rule of  $x = \mathbf{A}^{-1}b$ .

$$x = \frac{1}{\det(\mathbf{A})}\mathbf{C}^T b; \quad x_j = \frac{\det(\mathbf{B}_j)}{\mathbf{A}_j},$$

where  $B_j$  the matrix that replaces the  $j$ -th column of  $\mathbf{A}$  by  $b$ .

**Definition 9** (Parallelepiped). The parallelepiped determined by  $n$  vectors  $v_1, \dots, v_n \in \mathbb{R}^n$  is the subset

$$P = \{a_1v_1 + \dots + a_nv_n : 0 \leq a_1, \dots, a_n \leq 1\}.$$

We use  $\text{vol}(P)$  to denote the volume of the parallelepiped  $P$ .

**Proposition 2** (Determinants and volumes). *The absolute determinant of a square matrix  $\mathbf{A}$  is the volume of parallelepiped determined by the rows of  $\mathbf{A}$ , i.e.*

$$|\det(\mathbf{A})| = \text{vol}(P),$$

where  $P$  is the parallelepiped determined by the rows of  $\mathbf{A}$ .

**Example 2.** The area of triangle with vertices at  $(x_1, y_1), (x_2, y_2), (x_3, y_2)$  is

$$\frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

*Proof.* The two edges of the triangle are  $v_1 = (x_2 - x_1, y_2 - y_1)$  and  $v_2 = (x_3 - x_1, y_3 - y_1)$ . The area of the triangle is half of the area of parallelepiped determined by  $v_1, v_2$ . Therefore, the area of triangle is

$$\frac{1}{2} \begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} = \frac{1}{2} \left( \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} - \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} \right) = \frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

$\square$