

Linear Algebra – Part II

A summary for MIT 18.06SC

Jiaxin Hu

July 10, 2020

1 Orthogonality

1.1 Orthogonal vectors and subspaces

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

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

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, there are four subspaces related to \mathbf{A} : **columnspace** $C(\mathbf{A})$, row space $C(\mathbf{A}^T)$, nullspace $N(\mathbf{A})$, and left nullspace $N(\mathbf{A}^T)$. Suppose the 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). Consider 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.

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

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

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

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

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

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

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 a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. First, we prove that $N(\mathbf{A}^T \mathbf{A}) = N(\mathbf{A})$.

On one hand, if $x \in N(\mathbf{A})$, then $\mathbf{A}x = 0 \Rightarrow \mathbf{A}^T \mathbf{A}x = \mathbf{A}^T 0 = 0$. Therefore, for any $x \in N(\mathbf{A})$, the vector $x \in N(\mathbf{A}^T \mathbf{A})$. prove by contraction that, for

On the other hand, we prove that for any $x \in N(\mathbf{A}^T \mathbf{A})$ the vector $x \in N(\mathbf{A})$ by contradiction.

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

$$\mathbf{A}x = b \neq 0, \quad \mathbf{A}^T \mathbf{A}x = 0 \quad \Rightarrow \quad \mathbf{A}^T b = 0. \quad (1)$$

By the first equality in (1), $b \in C(\mathbf{A})$, and by the third equation in (1), $b \in N(\mathbf{A}^T)$. This contradicts the Theorem 1.1, $C(\mathbf{A}) \perp N(\mathbf{A}^T)$. Therefore, for any $x \in N(\mathbf{A}^T \mathbf{A})$, the vector $x \in N(\mathbf{A})$.

Next, given $N(\mathbf{A}^T \mathbf{A}) = N(\mathbf{A})$, we have $\text{rank}(\mathbf{A}^T \mathbf{A}) = n - \dim(N(\mathbf{A}^T \mathbf{A})) = n - \dim(N(\mathbf{A})) = \text{rank}(\mathbf{A})$. □

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 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. □

1.2 Projections onto subspaces

Let $x \in \mathbb{R}^n$ be a vector,

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

reduce the use of "given", especially when starting a paragraph.

$$(x - p) \perp C(\mathbf{A}), \quad (2)$$

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 (2), we have

rewrite to make the statement concise. No derivation in Definition.

$$\mathbf{A}^T(x - p) = \mathbf{A}^T(x - \mathbf{A}\hat{x}) = 0 \quad \Rightarrow \quad \hat{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x, \quad p = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x.$$

The matrix $\mathbf{P} \triangleq \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ is called a 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,

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix, and \mathbf{P} the projection matrix of \mathbf{A} onto the column space...

"Let ..." is a common statement to start Definition/Theorem.

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

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

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

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

□

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

Proof. For any $x \in \mathbb{R}^n$, we have

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

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

1.3 Projection matrices and least squares

Given an observation vector $y \in \mathbb{R}^n$ and a 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 regression coefficients and ϵ is the noise. The least square estimate of β is the minimizer of the loss, i.e.

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

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} , which minimizes the distance from y to the column space $C(\mathbf{X})$. Therefore, we may use projection tools to solve the least square estimate.

The projection $\mathbf{X}\hat{\beta}_{LS}$ satisfies “solve the estimate”? solve the problem, or solve for the estimator.

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

The estimate $\hat{\beta}_{LS}$ consists with the estimates solved by 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 an 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}

$$\mathbf{A} = [a_1, \dots, a_n] = [e_1, \dots, e_n] \begin{bmatrix} e_1^T a_1 & e_1^T a_2 & \cdots & e_1^T a_n \\ 0 & e_2^T a_2 & \cdots & e_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_n^T a_n \end{bmatrix} = \mathbf{Q}\mathbf{R} \quad (3)$$

where $\mathbf{Q} \in \mathbb{R}^{m \times n}$ is an orthonormal matrix and $\mathbf{R} \in \mathbb{R}^{n \times n}$ is an upper triangular matrix. We call the matrix decomposition in equation (3) 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 . The matrix after exchanging a_i, a_j , \mathbf{A}' , is the same as the original matrix \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 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 . Then, the determinant $\det(\mathbf{A}) = \prod_{i=1}^n d_i$.

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 entries. 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 $\mathbf{D}_A, \mathbf{D}_B$ by elimination. Suppose a_1, \dots, a_n and b_1, \dots, b_n are diagonal elements for \mathbf{D}_A and \mathbf{D}_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} is*

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

where $(\alpha_1, \alpha_2, \dots, \alpha_n)$ is the permutation of $(1, 2, \dots, n)$, \mathcal{P} is the collection of all possible $(\alpha_1, \dots, \alpha_n)$, and $N(\alpha_1, \dots, \alpha_n)$ is the number of necessary 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 $n!$ nonsingular matrices among the n^n matrices. The number of nonsingular 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 $n \times (n - 1) \times (n - 2) \times \cdots \times 2 = n!$ nonsingular matrices.

By property 7 and 2, the determinant for the nonsingular 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 $(1, 2, \dots, n)$, \mathcal{P} is the collection of all possible $(\alpha_1, \dots, \alpha_n)$, and $N(\alpha_1, \dots, \alpha_n)$ is the number of necessary 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 of cofactor formula is computing the determinant of *tridiagonal matrix*. A tridiagonal matrix is a matrix in which only nonzero 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 an invertible matrix. Here, we apply cofactor formula to compute 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 elements of $\mathbf{A}\mathbf{C}^T$ are $(\mathbf{A}\mathbf{C}^T)_{ii} = \sum_k^n a_{ik}C_{ik} = \det(\mathbf{A}), \forall i \in [n]$. Consider the off-diagonal entries $(\mathbf{A}\mathbf{C}^T)_{ij} = \sum_k^n a_{ik}C_{jk}$, 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 $\mathbf{A}\mathbf{C}^T$ are 0. Then, we have $\mathbf{A}\mathbf{C}^T = \det(\mathbf{A})\mathbf{I}_n$. \square

Definition 8 (*Cramer's rule*). Consider the linear system $\mathbf{A}x = b$, 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 \mathbf{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

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

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.*

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

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

Example 2. The area of a 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 a half of the area of the parallelepiped determined by v_1, v_2 . Therefore, the area of the 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

3 Eigenvalues and eigenvectors

Definition 10 (*Eigenvalues and eigenvectors*). Consider a square matrix \mathbf{A} , an eigenvector of \mathbf{A} is a nonzero vector such that:

$$\mathbf{A}x = \lambda x, \quad \text{for some } \lambda \in \mathbb{C}.$$

The value λ is called the eigenvalue of \mathbf{A} , and x is the eigenvector associated with eigenvalue λ . Usually, eigenvectors are normalized, i.e. $\|x\| = 1$, where $\|\cdot\|$ is the euclidean distance. For simplicity, all the eigenvectors mentioned below are normalized. Besides, the eigenvectors associated with eigenvalue 0 span the nullspace of \mathbf{A} .

Definition 11 (*Trace of square matrix*). Consider a square matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$. The trace of \mathbf{A} is defined as

$$\text{tr}(\mathbf{A}) = \sum_i^n a_{ii}.$$

Definition 12 (*Characteristic polynomial*). Let λ denote an eigenvalue of the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then the matrix $\mathbf{A} - \lambda \mathbf{I}$ is a singular matrix. Therefore,

$$\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0 \quad \Leftrightarrow \quad (-1)^n \lambda^n + \text{tr}(\mathbf{A}) \lambda^{n-1} + \cdots + (-1)^n \det(\mathbf{A}) \triangleq P(\lambda) = 0.$$

The polynomial $P(\lambda)$ is called the characteristic equation of \mathbf{A} .

Since the characteristic polynomial $P(\lambda)$ is of degree n , there are n solutions to the equation $P(\lambda) = 0$. The solutions $\lambda_1, \dots, \lambda_n$ are n eigenvalues of \mathbf{A} , which may be complex and may not be distinct with each other. Note that the complex eigenvalues come in conjugate pairs, because $P(\lambda) = P(\bar{\lambda})$.

3.1 Properties for eigenvectors and eigenvalues

Theorem 3.1 (Summation and production of eigenvalues). For a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, let $\lambda_1, \dots, \lambda_n$ denote the n eigenvalues of \mathbf{A} . Then

$$\sum_{i=1}^n \lambda_i = \text{tr}(\mathbf{A}); \quad \prod_{i=1}^n \lambda_i = \det(\mathbf{A}).$$

Proof. First, we re-write the characteristic polynomial $P(\lambda)$ of \mathbf{A} as

$$P(\lambda) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n), \tag{4}$$

where $\lambda_i, i \in [n]$ are n eigenvalues of \mathbf{A} . By equation (4), the coefficient for λ^{n-1} is equal to $(-1)^n \sum_{i=1}^n \lambda_i$. Compared with definition 12, we have $\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$. Similarly, the constant term in equation (4) is $(-1)^n \prod_{i=1}^n \lambda_i$. Therefore, $\prod_{i=1}^n \lambda_i = \det(\mathbf{A})$. \square

Definition 13 (*Similar matrices*). Consider a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. For any invertible matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$, the matrix product \mathbf{BAB}^{-1} is called the similar matrix of \mathbf{A} .

Theorem 3.2 (Eigenvalues for similar matrices). Consider a square matrix \mathbf{A} . The similar matrices of \mathbf{A} share the same eigenvalues of \mathbf{A} .

Proof. Let \mathbf{BAB}^{-1} be the similar matrix of \mathbf{A} , and $\lambda_1, \dots, \lambda_n$ be the eigenvalues of \mathbf{A} . Then, for any λ_i , there is an eigenvector, such that,

$$\lambda x = \mathbf{A}x \Leftrightarrow \lambda \mathbf{B}x = \mathbf{BAB}^{-1}\mathbf{B}x.$$

Since x is a nonzero vector, and \mathbf{B} is invertible, the vector $\mathbf{B}x$ is also a nonzero vector. Therefore, λ is an eigenvalue of \mathbf{BAB}^{-1} with associated eigenvector $\mathbf{B}x$. \square

Theorem 3.3 (Eigenvalues for powers of the matrix). *Consider a matrix \mathbf{A} with a eigenvalue λ and the associated eigenvector x . For any polynomial P ,*

$$P(\mathbf{A})x = P(\lambda)x.$$

Proof. For any integer $k \geq 0$ and constant $c \in \mathbb{R}$,

$$c\mathbf{A}^k x = c\mathbf{A}^{k-1}\mathbf{A}x = c\lambda\mathbf{A}^{k-1}x = \dots = c\lambda^k x.$$

Therefore, for any polynomial P , $P(\mathbf{A})x = P(\lambda)x$. \square

Theorem 3.4 (Eigenvalues for the inverse). *Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an invertible matrix. Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalue of \mathbf{A} , and x_1, \dots, x_n denote the associated eigenvectors. Then, $\lambda_1^{-1}, \dots, \lambda_n^{-1}$ and x_1, \dots, x_n are the eigenvalues and associated eigenvectors of \mathbf{A}^{-1} , respectively.*

Proof. Since the nullspace of \mathbf{A} contains only a zero vector, the eigenvalues $\lambda_i > 0, \forall i \in [n]$. For any λ_i and x_i , multiplying x_i on both sides of the equation $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$, we have

$$\mathbf{A}^{-1}\mathbf{A}x_i = x_i \Rightarrow \mathbf{A}^{-1}x_i = \lambda_i^{-1}x_i.$$

Therefore, \mathbf{A}^{-1} share the same eigenvectors set with \mathbf{A} , and the reciprocal of the eigenvalues of \mathbf{A} are the eigenvalues of \mathbf{A}^{-1} . \square

Theorem 3.5 (Independence of eigenvectors). *Suppose the square matrix \mathbf{A} has at least two different eigenvalues. The eigenvectors associated with different eigenvalues are independent.*

Proof. Suppose the square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has m different eigenvalues $\lambda_1, \dots, \lambda_m$ with associated eigenvectors v_1, \dots, v_m . Let the linear combination $c_1v_1 + \dots + c_mv_m = 0$. Then we have

$$\mathbf{A}(c_1v_1 + \dots + c_mv_m) = c_1\lambda_1v_1 + \dots + c_m\lambda_mv_m = 0; \quad (5)$$

$$\lambda_1(c_1v_1 + \dots + c_mv_m) = c_1\lambda_1v_1 + \dots + c_m\lambda_1v_m = 0. \quad (6)$$

Subtracting the equations (5) from (6),

$$c_2(\lambda_1 - \lambda_2)v_2 + \dots + c_m(\lambda_1 - \lambda_m)v_m = 0 \quad (7)$$

Repeating the steps from equation (5) to equation (7), finally we have

$$c_m(\lambda_1 - \lambda_m)(\lambda_2 - \lambda_m) \dots (\lambda_{m-1} - \lambda_m)v_m = 0.$$

Since $\lambda_1, \dots, \lambda_m$ are distinct, and v_m is a nonzero vector, we have $c_m = 0$. Plugging $c_m = 0$ back to previous steps leads to that $c_1 = \dots = c_m = 0$. Therefore, v_1, \dots, v_m are independent. \square

Theorem 3.6 (Eigenvalues for triangular matrices). *For triangular matrices, the eigenvalues are the entries on the diagonal.*

Proof. Suppose $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ is a triangular matrix. By the property of determinant 7,

$$\det(\mathbf{A}) = \prod_{i=1}^n a_{ii}; \quad \det(\mathbf{A} - \lambda \mathbf{I}_n) = \prod_{i=1}^n (a_{ii} - \lambda).$$

To let $\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0$, $\lambda = a_{ii}, \forall i \in [n]$. Therefore, the entries on the diagonal are the eigenvalues of a triangular matrix. \square

3.2 Eigenvalues and eigenvectors for symmetric matrices

Theorem 3.7 (Eigenvalues for symmetric and antisymmetric matrices). *Symmetric square matrices with all real entries have real eigenvalues. For antisymmetric matrices which satisfy $\mathbf{A}^T = -\mathbf{A}$, all eigenvalues are imaginary, i.e. $\lambda = bi$, where $b \in \mathbb{R}, i = \sqrt{-1}$.*

Proof. First, we prove that symmetric square matrices with all real entries have real eigenvalues by contradiction.

Suppose a symmetric real matrix \mathbf{A} has a complex eigenvalue λ with eigenvector x . Since \mathbf{A} is real, then the eigenvector x is also complex. Let $\bar{\lambda}$ and \bar{x} denote the conjugate eigenvalue and eigenvector, respectively. Then we have

$$\mathbf{A}x = \lambda x; \quad \mathbf{A}\bar{x} = \bar{\lambda}\bar{x}. \quad (8)$$

Premultiply \bar{x}^T on both sides of the first equation in (8) and premultiply x^T on both sides of the second equation in (8). By the symmetry of \mathbf{A} ,

$$\bar{x}^T \mathbf{A}x = x^T \mathbf{A}^T \bar{x} = x^T \mathbf{A}\bar{x}.$$

Subtracting the two equations in (8), we have

$$0 = (\bar{x}^T \mathbf{A}x - x^T \mathbf{A}\bar{x}) = (\lambda - \bar{\lambda})x^T \bar{x}.$$

By the fact that $x^T \bar{x}$ is a real number, the imaginary part of λ is equal to 0, which contradicts that λ is complex.

Next, we prove that the eigenvalues for antisymmetric matrices are imaginary. Suppose \mathbf{A} is an antisymmetric matrix with a complex eigenvalue λ and complex eigenvector x . By similar procedures for symmetric matrices, we have

$$\bar{x}^T \mathbf{A}x = x^T \mathbf{A}^T \bar{x} = -x^T \mathbf{A}\bar{x}.$$

Then,

$$0 = (\bar{x}^T \mathbf{A}x + x^T \mathbf{A}\bar{x}) = (\lambda + \bar{\lambda})x^T \bar{x}.$$

Since $x^T \bar{x}$ is a real number, the real part of λ is equal to 0. Therefore, eigenvalues for antisymmetric matrices are imaginary. \square

Theorem 3.8 (Orthogonality of eigenvectors for symmetric matrices). *The eigenvectors of a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ associated with different eigenvalues are orthogonal.*

Proof. Suppose λ_1 and λ_2 are two different eigenvalues of \mathbf{A} with associated eigenvector x_1, x_2 , respectively. By the symmetry of \mathbf{A} , we have

$$x_2^T \mathbf{A}x_1 = x_1^T \mathbf{A}^T x_2 = x_1^T \mathbf{A}x_2,$$

then

$$0 = (x_2^T \mathbf{A}x_1 - x_1^T \mathbf{A}x_2) = (\lambda_1 - \lambda_2)x_2^T x_1.$$

Since λ_1, λ_2 are different, we have $\lambda_1 - \lambda_2 \neq 0$ and $x_2^T x_1 = 0$. Therefore, x_1 and x_2 are orthogonal. \square

Theorem 3.9 (Eigenvectors for repeated eigenvalues of symmetric matrices). *Consider a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. If λ_0 is a repeated eigenvalue of \mathbf{A} with multiplicity $2 \leq m \leq n$, then there exist m orthonormal eigenvectors corresponding to λ_0 .*

Proof. First, there is at least one nonzero eigenvector x_0 for λ_0 . For any $x_0 \in \mathbb{R}^n$, there are $n - 1$ additional orthogonal vectors y_1, \dots, y_{n-1} , such that $y_j \perp x_0, \forall j \in [(n - 1)]$ and $(x_0, y_1, \dots, y_{n-1})$ forms the basis of \mathbb{R}^n . Let $\mathbf{Y} = [y_1, \dots, y_{n-1}]$ and $\mathbf{X} = [x_0, \mathbf{Y}]$. Since \mathbf{A} is a symmetric matrix, $x_0^T \mathbf{A} x_0 = \lambda_0 x_0^T x_0$. We have

$$\mathbf{X}^T \mathbf{A} \mathbf{X} = \begin{bmatrix} x_0^T \mathbf{A} x_0 & x_0^T \mathbf{A} \mathbf{Y} \\ \mathbf{Y}^T \mathbf{A} x_0 & \mathbf{Y}^T \mathbf{A} \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \mathbf{Y}^T \mathbf{A} \mathbf{Y} \end{bmatrix}.$$

Note that \mathbf{X} is an orthogonal matrix which satisfies $\mathbf{X}^T = \mathbf{X}^{-1}$. By Theorem 3.2, $\mathbf{X}^T \mathbf{A} \mathbf{X}$ is a similar matrix of \mathbf{A} , and $\mathbf{X}^T \mathbf{A} \mathbf{X}$ share the same eigenvalues of \mathbf{A} . Then, λ_0 is also a repeated eigenvalue of $\mathbf{X}^T \mathbf{A} \mathbf{X}$. The determinant of $\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda \mathbf{I}_n$ is

$$\det(\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda \mathbf{I}_n) = (\lambda_0 - \lambda) \det(\mathbf{Y}^T \mathbf{A} \mathbf{Y} - \lambda \mathbf{I}_{n-1}).$$

Since λ_0 has multiplicity $m \geq 2$, then $\det(\mathbf{Y}^T \mathbf{A} \mathbf{Y} - \lambda_0 \mathbf{I}_{n-1}) = 0$. Therefore, the dimension of the nullspace of $\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda_0 \mathbf{I}_n$ is larger than 2. Let v_1, v_2 be the two orthogonal vectors in the nullspace of $\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda_0 \mathbf{I}_n$. Then v_1, v_2 are two eigenvectors of $\mathbf{X}^T \mathbf{A} \mathbf{X}$, and $\mathbf{X} v_1, \mathbf{X} v_2$ are two orthogonal eigenvectors of \mathbf{A} associated with λ_0 .

Replace x_0 by two orthogonal vectors $x_1 = \mathbf{X} v_1, x_2 = \mathbf{X} v_2$ and repeat the steps above. Finally, we will have m orthogonal eigenvectors of \mathbf{A} associated with λ_0 . \square

3.3 Diagonalization

Definition 14 (Diagonalization). Suppose the square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has n independent eigenvectors. Then the matrix \mathbf{A} can be diagonalized as

$$\mathbf{A} = \mathbf{S}^{-1} \mathbf{\Lambda} \mathbf{S},$$

where $\mathbf{S} \in \mathbb{R}^{n \times n}$ is a matrix whose columns are n independent eigenvectors of \mathbf{A} and $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose entries are n eigenvalues of \mathbf{A} .

Theorem 3.10 (Diagonalization of symmetric matrices). Suppose \mathbf{A} is a symmetric matrix, then \mathbf{A} can be diagonalized.

Proof. By Theorems 3.8 and Theorem 3.9, for any symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, \mathbf{A} have n orthogonal eigenvectors, even though there may have repeated eigenvalues. \square

Theorem 3.11 (Descent of matrix powers). If a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has n independent eigenvectors with eigenvalues $\lambda_1, \dots, \lambda_n$, then $\mathbf{A}^k \rightarrow 0$ as $k \rightarrow +\infty$ if and only if all $|\lambda_i| < 1$.

Proof. We diagonalize the matrix \mathbf{A} as $\mathbf{A} = \mathbf{S}^{-1} \mathbf{\Lambda} \mathbf{S}$. For any positive integer k , $\mathbf{A}^k = \mathbf{S}^{-1} \mathbf{\Lambda}^k \mathbf{S}$. Then

$$\lim_{k \rightarrow +\infty} \mathbf{A}^k = \lim_{k \rightarrow +\infty} \mathbf{S}^{-1} \mathbf{\Lambda}^k \mathbf{S} = \lim_{k \rightarrow +\infty} \mathbf{S}^{-1} \begin{bmatrix} |\lambda_1|^k & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & |\lambda_n|^k \end{bmatrix} \mathbf{S} = 0 \quad \Leftrightarrow \quad |\lambda_i| < 1, \forall i \in [n].$$

\square