

Linear Algebra – Part II

A summary for MIT 18.06SC

Jiaxin Hu

July 23, 2020

1 Orthogonality

1.1 Orthogonal vectors and subspaces

Definition 1 (*Orthogonal vectors*). Let $x, y \in \mathbb{R}^n$ be two vectors. 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*). Let S, T be two subspaces. The subspaces S and T are orthogonal, denoted $S \perp T$, if and only if for any $s \in S$ and any $t \in T$, $s^T t = t^T s = 0$.

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix. 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 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). Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix. 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. That is

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

Proof. 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 $a_i \in \mathbb{R}^n$ is the i -th row of \mathbf{A} , for all $i \in [m]$. 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 $c_i \in \mathbb{R}$, for all $i \in [m]$. Taking inner product between 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 that $N(\mathbf{A}^T) \perp C(\mathbf{A})$. □

For notational convenience, we use the shorthands \mathbf{A} for the column space $C(\mathbf{A})$ and \mathbf{A}^\perp for the nullspace $N(\mathbf{A})$ in the rest of this section. Though \mathbf{A} usually denotes a matrix, the meaning of \mathbf{A} should be clear given the context. For example, $\text{rank}(\mathbf{A})$ refers the rank of matrix \mathbf{A} , while $\dim(\mathbf{A})$ refers to the dimension of the column space $C(\mathbf{A})$. Specifically, for a vector v , $v \in \mathbf{A}$ means that the vector v belongs to the column space $C(\mathbf{A})$.

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, we also rewrite the Theorem 1.1 as

$$\mathbf{A}^T \perp \mathbf{A}^\perp \quad \text{and} \quad \mathbf{A} \perp \mathbf{A}^{T,\perp}. \quad (\mathbf{A}^T)^\perp$$

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

$$(\mathbf{A}^T \mathbf{A})^\perp = \mathbf{A}^\perp \quad \text{and} \quad \text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}).$$

Proof. First, we prove that $(\mathbf{A}^T \mathbf{A})^\perp = \mathbf{A}^\perp$.

On one hand, if the vector $x \in \mathbf{A}^\perp$, then $\mathbf{A}x = 0$. Multiplying the matrix \mathbf{A}^T on both sides of the equation, we have $\mathbf{A}^T \mathbf{A}x = 0$. Therefore, for any $x \in \mathbf{A}^\perp$, the vector $x \in (\mathbf{A}^T \mathbf{A})^\perp$.

On the other hand, if the vector $x \in (\mathbf{A}^T \mathbf{A})^\perp$, we have

$$\mathbf{A}^T \mathbf{A}x = 0.$$

Then, the vector $\mathbf{A}x \in \mathbf{A}^{T,\perp}$. By the definition of matrix-to-vector multiplication, we also have $\mathbf{A}x \in \mathbf{A}$. By the Theorem 1.1, the subspaces $\mathbf{A}^{T,\perp} \perp \mathbf{A}$. Therefore, $\mathbf{A}x = 0$, and we obtain that $x \in \mathbf{A}^\perp$.

Next, we prove that $\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A})$.

Given $(\mathbf{A}^T \mathbf{A})^\perp = \mathbf{A}^\perp$, we have $\text{rank}(\mathbf{A}^T \mathbf{A}) = n - \dim((\mathbf{A}^T \mathbf{A})^\perp) = n - \dim(\mathbf{A}^\perp) = \text{rank}(\mathbf{A})$. \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. Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a matrix with independent columns; i.e., $\text{rank}(\mathbf{A}) = n$. 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. \square

1.2 Projections onto subspaces

Definition 3 (*Projection and projection matrix*). Let $x \in \mathbb{R}^n$ be a vector and $\mathbf{A}^{n \times m}$ be a matrix. The vector $p_{\mathbf{A}}(x) \in \mathbf{A}$ that satisfies

$$(x - p_{\mathbf{A}}(x)) \perp \mathbf{A}, \tag{1}$$

is called the projection of vector x onto the column space \mathbf{A} . For all $x \in \mathbb{R}^n$, the matrix $\mathbf{P}_{\mathbf{A}} \in \mathbb{R}^{n \times n}$ that satisfies

$$p_{\mathbf{A}}(x) = \mathbf{P}_{\mathbf{A}}x,$$

is called the projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space \mathbf{A} . Note that the projection $p_{\mathbf{A}}(x)$ depends on the vector x and the space \mathbf{A} , while the projection matrix $\mathbf{P}_{\mathbf{A}}$ depends on the space \mathbf{A} only.

Proposition 1 (Projection matrix of \mathbf{A}). *Let $x \in \mathbb{R}^n$ be a vector and $\mathbf{A}^{n \times m}$ be a matrix with independent columns. The projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space \mathbf{A} is*

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

Proof. Let x_0 be an arbitrary vector of \mathbb{R}^n . By Definition 3, the projection $p_{\mathbf{A}}(x_0) \in \mathbf{A}$. Then, there exists a vector $v(x_0) \in \mathbb{R}^m$ such that $p_{\mathbf{A}}(x_0) = \mathbf{A}v(x_0)$. By equation (1), we have

$$\mathbf{A}^T(x_0 - \mathbf{A}v(x_0)) = 0 \Rightarrow v(x_0) = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x_0 \Rightarrow p_{\mathbf{A}}(x_0) = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T x_0. \quad (2)$$

Since x_0 is arbitrary, the equation (2) still holds after replacing all the x_0 by x , for all $x \in \mathbb{R}^n$. Therefore, the matrix $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ is the projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space \mathbf{A} . \square

Theorem 1.3 (Properties of projection matrix). *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix, and $\mathbf{P}_{\mathbf{A}}$ be the projection matrix of \mathbf{A} onto the column space \mathbf{A} . Then,*

$$\mathbf{P}_{\mathbf{A}}^T = \mathbf{P}_{\mathbf{A}} \quad \text{and} \quad \mathbf{P}_{\mathbf{A}}^2 = \mathbf{P}_{\mathbf{A}}.$$

Proof. By Proposition 1, the projection matrix is $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Then,

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

Similarly, we have

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

\square

Corollary 2 (Projection onto $\mathbf{A}^{T,\perp}$). Suppose $\mathbf{P}_{\mathbf{A}}$ is the projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space \mathbf{A} . Then $\mathbf{I} - \mathbf{P}_{\mathbf{A}}$ is the projection matrix of \mathbf{A} from \mathbb{R}^n onto the left nullspace $\mathbf{A}^{T,\perp}$.

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

Proof. By Definition 3, ~~for all $x \in \mathbb{R}^n$~~ , $\mathbf{P}_{\mathbf{A}}x$ is the projection of x from \mathbb{R}^n onto \mathbf{A} . Then,

$$x - \mathbf{P}_{\mathbf{A}}x \perp \mathbf{A} \Rightarrow (\mathbf{I} - \mathbf{P}_{\mathbf{A}})x \perp \mathbf{A}. \quad (\text{pay attention to the use of phrase "for all } x \in \dots")$$

By Theorem 1.1, we have $\mathbf{A} \perp \mathbf{A}^{T,\perp}$. Since $\dim(\mathbf{A}) + \dim(\mathbf{A}^{T,\perp}) = n$, we have $\mathbf{A} \cup \mathbf{A}^{T,\perp} = \mathbb{R}^n$. Then, we have

$$(\mathbf{I} - \mathbf{P}_{\mathbf{A}})x \in \mathbf{A}^{T,\perp}.$$

Since $\mathbf{P}_{\mathbf{A}}x$ is the projection of x from \mathbb{R}^n to \mathbf{A} , the vector $\mathbf{P}_{\mathbf{A}}x \in \mathbf{A}$, ~~for all $x \in \mathbb{R}^n$~~ . We have,

$$x - (\mathbf{I} - \mathbf{P}_{\mathbf{A}})x = \mathbf{P}_{\mathbf{A}}x \in \mathbf{A} \Rightarrow x - (\mathbf{I} - \mathbf{P}_{\mathbf{A}})x \perp \mathbf{A}^{T,\perp}.$$

Therefore, ~~for all $x \in \mathbb{R}^n$~~ , $(\mathbf{I} - \mathbf{P}_{\mathbf{A}})x$ is the projection of x from \mathbb{R}^n onto the left nullspace $\mathbf{A}^{T,\perp}$, and $\mathbf{I} - \mathbf{P}_{\mathbf{A}}$ is the projection matrix of \mathbf{A} from \mathbb{R}^n onto $\mathbf{A}^{T,\perp}$. \square

Since the above conclusion holds for all $x \in \mathbb{R}^n$,

1.3 Projection matrices and least squares

Let $y \in \mathbb{R}^n$ be a vector and $\mathbf{X} \in \mathbb{R}^{n \times (k+1)}$ be a design matrix. We propose the linear regression model as

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

where $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ are regression coefficients and ϵ is the noise. The least-squares estimate of β is the minimizer of the squared loss; i.e.,

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

is

where $\|\cdot\|$ is the Euclidean norm. The projection of y onto \mathbf{X} minimizes the Euclidean distance from y to the column space \mathbf{X} . Hence, ~~we consider~~ the vector $\mathbf{X}\hat{\beta}_{LS}$ ~~as~~ the projection of y onto the column space \mathbf{X} . Therefore, we may use projection to solve the ~~minimization~~ problem.

By Definition 3, the projection $\mathbf{X}\hat{\beta}_{LS}$ satisfies

least squares

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

The estimate $\hat{\beta}_{LS}$ is ~~consistent with the estimates~~ solved using the first-order condition for the least-squares optimization. ~~same as the estimate~~

(we reserve the term "consistent" for estimates that converge to ground truth in high probability)

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*). Let $\mathbf{Q} \in \mathbb{R}^{m \times n}$ be a matrix. The matrix \mathbf{Q} is an orthonormal matrix, if the columns of \mathbf{Q} are orthonormal. When $m = n$, the square matrix \mathbf{Q} is an orthogonal matrix.

Suppose $\mathbf{Q} \in \mathbb{R}^{m \times n}$ is an orthonormal matrix. Then, we have $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n$. When $m = n$, \mathbf{Q} is an orthogonal matrix and $\mathbf{Q}^T = \mathbf{Q}^{-1}$. The projection matrix of \mathbf{Q} onto the column space $C(\mathbf{Q})$, denoted \mathbf{P}_Q , becomes $\mathbf{P}_Q = \mathbf{I}_m$.

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

$$\begin{aligned} u_1 &= a_1, & e_1 &= \frac{u_1}{\|u_1\|}; \\ u_2 &= a_2 - \frac{u_1^T a_2}{u_1^T u_1} u_1, & e_2 &= \frac{u_2}{\|u_2\|}; \\ u_3 &= a_3 - \frac{u_1^T a_3}{u_1^T u_1} u_1 - \frac{u_2^T a_3}{u_2^T u_2} u_2, & e_2 &= \frac{u_3}{\|u_3\|}; \\ &\vdots & & \\ u_n &= a_n - \sum_{k=1}^{n-1} \frac{u_k^T a_n}{u_k^T u_k} u_k, & e_n &= \frac{u_n}{\|u_n\|}. \end{aligned}$$

The vectors e_1, \dots, e_n are orthonormal basis of \mathbf{A} . Based on the basis obtained by Gram-Schmidt, we decompose the matrix \mathbf{A} as follows,

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

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. The matrix decomposition in equation (3) is called QR decomposition.

2 Determinants

The *determinant* is a number associated with a 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 from the first three basic properties.

1. The determinant of an identity matrix is equal to 1; i.e., $\det(\mathbf{I}) = 1$.
2. Exchanging two rows of a matrix reverses the sign of the 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 by row.

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

4. A matrix with two identical rows has the determinant 0.

Proof. Suppose \mathbf{A} has two identical rows a_i, a_j . The matrix after exchanging a_i, a_j , denoted \mathbf{A}' , is the same as the original matrix \mathbf{A} . By property 2, we have $\det(\mathbf{A}) = -\det(\mathbf{A}') = -\det(\mathbf{A})$. Therefore, $\det(\mathbf{A}) = 0$. □

preserves

5. Subtracting a multiplier of the i -th row from the j -th row ~~maintains~~ the determinant of the matrix, where $i \neq j$.

Proof. Take a 2-by-2 matrix as an example. By properties 3 and 4, for any constant $t \in \mathbb{R}$, we have

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

The proof for higher-dimension matrices is similar. □

6. A matrix with a row that is all zeros has the determinant 0.

Proof. By property 3, letting $t = 0$ leads to property 6. □

7. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a triangular matrix with diagonal elements d_1, \dots, d_n . The determinant $\det(\mathbf{A}) = \prod_{i=1}^n d_i$.

Proof. We eliminate the matrix \mathbf{A} to a diagonal matrix, denoted \mathbf{A}' . By property 5, $\det(\mathbf{A}) = \det(\mathbf{A}')$. Since \mathbf{A} is triangular, the diagonal elements of \mathbf{A}' are still d_1, \dots, d_n . By properties 3 and 1, $\det(\mathbf{A}') = \prod_{i=1}^n d_i \det(\mathbf{I}_n) = \prod_{i=1}^n d_i$. Therefore, $\det(\mathbf{A}) = \prod_{i=1}^n d_i$. \square

8. A matrix is singular if and only if its determinant is 0.

Proof. (\Rightarrow) If \mathbf{A} is a singular matrix, the reduced row echelon form of \mathbf{A} , denoted $RREF(\mathbf{A})$, has a row with all 0 entries. By property 6, the determinant $\det(RREF(\mathbf{A})) = 0$. Therefore, by property 5, we have $\det(\mathbf{A}) = \det(RREF(\mathbf{A})) = 0$.

(\Leftarrow) If the determinant $\det(\mathbf{A}) = 0$, the determinant of $RREF(\mathbf{A})$ is 0. Since the reduced row echelon form of a matrix contains only distinct nonzero rows, the zero determinant implies that there exists at least one zero row in $RREF(\mathbf{A})$. Therefore, the matrix \mathbf{A} is singular. \square

9. Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ be two matrices. The determinant of the matrix product is $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$.

Proof. First, suppose that at least one of \mathbf{A}, \mathbf{B} is singular. The product \mathbf{AB} is also singular. Then, $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}) = 0$.

Second, let \mathbf{E} be an elimination matrix. The product \mathbf{EB} preserves means implementing a linear operation on the rows of \mathbf{B} . By property 5, linear operation maintains the determinant of \mathbf{B} . Note that the determinant of an elimination matrix is 1. Therefore, we have

$$|\mathbf{EB}| = |\mathbf{E}||\mathbf{B}|. \quad (4)$$

Last, suppose that both \mathbf{A}, \mathbf{B} are invertible. By elimination, there exist a sequence of elimination matrices $\{\mathbf{E}_i\}_{i=1}^n$ such that

$$\mathbf{E}_n \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{I}.$$

By Lemma 2 in Linear Algebra-Part I, the inverse of an elimination matrix is still an elimination matrix. Let \mathbf{E}'_k denote the inverse \mathbf{E}_k^{-1} , for all $k \in [n]$. We rewrite \mathbf{A} as

$$\mathbf{A} = \mathbf{E}'_n \cdots \mathbf{E}'_1.$$

By equation (4), we have

$$|\mathbf{AB}| = |\mathbf{E}'_n \cdots \mathbf{E}'_1 \mathbf{B}| = |\mathbf{E}'_n| |\mathbf{E}'_{n-1} \cdots \mathbf{E}'_1 \mathbf{B}| = \cdots = |\mathbf{E}'_n| \cdots |\mathbf{E}'_1| |\mathbf{B}|. \quad (5)$$

Applying the equation (4) again to the term $|\mathbf{E}'_n| \cdots |\mathbf{E}'_1|$, we have

$$|\mathbf{E}'_n| \cdots |\mathbf{E}'_1| = |\mathbf{E}'_n| \cdots |\mathbf{E}'_3| |\mathbf{E}'_2 \mathbf{E}'_1| = \cdots = |\mathbf{E}'_n \cdots \mathbf{E}'_1|. \quad (6)$$

Hence, combining equation (5) with equation (6), we obtain the result

$$|\mathbf{AB}| = |\mathbf{E}'_n \cdots \mathbf{E}'_1| |\mathbf{B}| = |\mathbf{A}| |\mathbf{B}|.$$

\square

10. Let \mathbf{A} be a square matrix. The determinant $\det(\mathbf{A}^T) = \det(\mathbf{A})$.

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

2.2 Determinant computation

Proposition 2 (Big formula for computing determinant). *Let $\mathbf{A} = \llbracket a_{ij} \rrbracket \in \mathbb{R}^{n \times n}$ be a square matrix. 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 swaps from $(\alpha_1, \alpha_2, \dots, \alpha_n)$ to $(1, 2, \dots, n)$.

Proof. Let $\mathbf{A} = \llbracket a_{ij} \rrbracket \in \mathbb{R}^{n \times n}$ be a square matrix. We decompose the matrix \mathbf{A} into the sum of n^n matrices, denoted \mathbf{A}'_i , where $i \in [n^n]$. For each \mathbf{A}'_i , only one entry in each row comes from \mathbf{A} , and the other entries of \mathbf{A}'_i are 0. By property 3, the determinant $\det(\mathbf{A}) = \sum_{i=1}^{n^n} \det(\mathbf{A}'_i)$. Take a 3-by-3 square matrix as an example.

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

There are $n!$ nonsingular matrices among matrices \mathbf{A}'_i . The number of nonsingular matrices coincides with the size of \mathcal{P} . Specifically, 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 $n \times (n - 1) \times (n - 2) \times \dots \times 2 = n!$ nonsingular matrices.

By properties 7 and 2, the determinant of a nonsingular \mathbf{A}'_i 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 swaps from $(\alpha_1, \alpha_2, \dots, \alpha_n)$ to $(1, 2, \dots, n)$. The determinant of \mathbf{A} follows by the summation of the nonsingular matrices' determinants. \square

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

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

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

$$\det(\mathbf{A}) = \sum_{j=1}^n a_{ij} C_{ij} = \sum_{j=1}^n a_{ji} C_{ji}, \quad \text{for all } i \in [n].$$

Example 1 (Tridiagonal matrix). One example of using the cofactor formula is computing the determinant of a *tridiagonal matrix*. The tridiagonal matrix is a matrix in which nonzero elements lie only 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.,

located

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

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 the cofactor formula, we have

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

By the definition of tridiagonal matrix, we have $\mathbf{T}_{n,-1,-1} = \mathbf{T}_{n-1}$, and $\det(\mathbf{A}_{n,-1,-2}) = 1 \times \det(\mathbf{T}_{n-2})$. 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 the 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 $\mathbf{A}\mathbf{C}^T = \det(\mathbf{A})\mathbf{I}_n$.

Let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ be an invertible matrix, and C_{ij} be the cofactor of a_{ij} . By the definition of cofactor matrix, we have

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

Therefore, the diagonal elements of $\mathbf{A}\mathbf{C}^T$ are $(\mathbf{A}\mathbf{C}^T)_{ii} = \sum_{k=1}^n a_{ik}C_{ik} = \det(\mathbf{A})$, for all $i \in [n]$. Consider the off-diagonal entries $(\mathbf{A}\mathbf{C}^T)_{ij} = \sum_{k=1}^n a_{ik}C_{jk}$, for all $i \neq j$ and $i, j \in [n]$. By the cofactor formula, the sum $\sum_{k=1}^n a_{ik}C_{jk}$ 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 obtain $\mathbf{A}\mathbf{C}^T = \det(\mathbf{A})\mathbf{I}_n$. \square

Definition 8 (*Cramer's rule*). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an invertible matrix, and $x, b \in \mathbb{R}^n$ be two vectors. Applying Theorem 2.1 to the linear system $\mathbf{A}x = b$, we obtain the Cramer's rule of $x = \mathbf{A}^{-1}b$,

$$x = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T b \quad \text{and} \quad x_j = \frac{\det(\mathbf{B}_j)}{\det(\mathbf{A})}, \quad \text{for all } j \in [n],$$

where x_j is j -th entry of x , and \mathbf{B}_j is the matrix \mathbf{A} after replacing the j -th column by b .

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

$$P = \{a_1 v_1 + \cdots + a_n v_n : 0 \leq a_1, \dots, a_n \leq 1\}.$$

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

Proposition 3 (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 half of the area of the parallelepiped determined by v_1, v_2 . Therefore, the area of the triangle is

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

□

3 Eigenvalues and eigenvectors

Definition 10 (*Eigenvalues and eigenvectors*). Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a square matrix. Suppose there exists a nonzero vector $x \in \mathbb{R}^m$ 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 the vector x is called the eigenvector of \mathbf{A} associated with eigenvalue λ .

Usually, eigenvectors are normalized; i.e., $\|x\| = 1$, where $\|\cdot\|$ is the Euclidean norm. 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*). Let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ be a square matrix. The trace of \mathbf{A} is defined as

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}.$$

Definition 12 (*Characteristic polynomial*). Let λ denote an eigenvalue of the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. The determinant of $\mathbf{A} - \lambda \mathbf{I}_n$ is a polynomial of λ , denoted $P(\lambda)$. We call the polynomial $P(\lambda)$ as the characteristic polynomial of \mathbf{A} . Specifically,

$$P(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \text{tr}(\mathbf{A}) \lambda^{n-1} + \cdots + (-1)^n \det(\mathbf{A}).$$

Since the characteristic polynomial $P(\lambda)$ is of degree n , there exist n solutions to the equation $P(\lambda) = 0$. The solutions to $P(\lambda) = 0$, denoted $\lambda_1, \dots, \lambda_n$, are eigenvalues of \mathbf{A} .

The eigenvalues may be complex, and the complex eigenvalues come in conjugate pairs. Let λ be a complex eigenvalue, $\bar{\lambda}$ be the conjugate of λ , and $\overline{P(\cdot)}$ be the conjugate of the characteristic polynomial. Because the characteristic polynomial satisfies $P(\bar{\lambda}) = \overline{P(\lambda)}$, the conjugate eigenvalue $\bar{\lambda}$ is also a solution to $P(\lambda) = 0$.

The eigenvalues of a matrix are not necessarily distinct with each other.

3.1 Properties for eigenvectors and eigenvalues

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

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

Proof. First, we re-write the characteristic polynomial $P(\lambda)$ of \mathbf{A} ~~as a factor formula,~~ using the factorization form

$$P(\lambda) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n), \quad (7)$$

where λ_i are the eigenvalues of \mathbf{A} , for all $i \in [n]$. In equation (7), the coefficient for λ^{n-1} is equal to $(-1)^n \sum_{i=1}^n \lambda_i$. Comparing the equation (7) with Definition 12, we obtain $\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$. Similarly, the constant term in equation (7) is $(-1)^n \prod_{i=1}^n \lambda_i$, while the constant term in Definition 12 is $(-1)^n \det(\mathbf{A})$. Therefore, we obtain $\prod_{i=1}^n \lambda_i = \det(\mathbf{A})$. \square

Definition 13 (Similar matrices). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square matrix, and $\mathbf{B} \in \mathbb{R}^{n \times n}$ be an invertible matrix. The matrix product \mathbf{BAB}^{-1} is called the similar matrix of \mathbf{A} .

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

Proof. Let \mathbf{BAB}^{-1} be a similar matrix of \mathbf{A} , and (λ_i, x_i) be the i -th eigenvalue-eigenvector pair of \mathbf{A} , for all $i \in [n]$. Then, we have

$$\lambda_i x_i = \mathbf{A}x_i \quad \Rightarrow \quad \lambda_i \mathbf{B}x_i = \mathbf{BAB}^{-1} \mathbf{B}x_i, \quad \text{for all } i \in [n].$$

Since that x_i is a nonzero vector, and \mathbf{B} is invertible, the vector $\mathbf{B}x_i$ is also a nonzero vector. Therefore, $(\lambda_i, \mathbf{B}x_i)$ is the i -th eigenvalue-eigenvector pair of \mathbf{BAB}^{-1} , for all $i \in [n]$. \square

Theorem 3.3 (Eigenvalues for powers of the matrix). *Let \mathbf{A} be a matrix, and (λ, x) be an eigenvalue-eigenvector pair of \mathbf{A} . For any polynomial P , we have*

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

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

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

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

Theorem 3.4 (Eigenvalues for the inverse). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an invertible matrix, and (λ_i, x_i) be the i -th eigenvalue-eigenvector pair of \mathbf{A} , for all $i \in [n]$. Then, (λ_i^{-1}, x_i) is the i -th eigenvalue-eigenvector pair of the inverse \mathbf{A}^{-1} , for all $i \in [n]$.*

Proof. Since \mathbf{A} is invertible, the eigenvalue $\lambda_i \neq 0$, for all $i \in [n]$. Multiplying x_i on both sides of the equation $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}_n$ yields

$$\mathbf{A}^{-1} \mathbf{A}x_i = x_i \quad \Rightarrow \quad \mathbf{A}^{-1} x_i = \lambda_i^{-1} x_i, \quad \text{for all } i \in [n].$$

Therefore, (λ_i^{-1}, x_i) is the eigenvalue-eigenvector pair of \mathbf{A}^{-1} , for all $i \in [n]$. \square

Theorem 3.5 (Independence of eigenvectors). *The eigenvectors of a matrix associated with distinct eigenvalues are independent.*

Proof. Suppose the square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has m distinct eigenvalues $\lambda_1, \dots, \lambda_m$ with associated eigenvectors v_1, \dots, v_m . Consider a linear combination $c_1 v_1 + \dots + c_m v_m = 0$, where $c_i \in \mathbb{R}$, for all $i \in [m]$. Then, we have

$$\mathbf{A}(c_1 v_1 + \dots + c_m v_m) = c_1 \lambda_1 v_1 + \dots + c_m \lambda_m v_m = 0, \quad (8)$$

and

$$\lambda_1(c_1 v_1 + \dots + c_m v_m) = c_1 \lambda_1 v_1 + \dots + c_m \lambda_1 v_m = 0. \quad (9)$$

Subtracting the equations (8) from (9) yields

$$c_2(\lambda_1 - \lambda_2)v_2 + \dots + c_m(\lambda_1 - \lambda_m)v_m = 0. \quad (10)$$

Consider the new linear combination $c_2(\lambda_1 - \lambda_2)v_2 + \dots + c_m(\lambda_1 - \lambda_m)v_m = 0$, and repeat the steps from equation (8) to equation (10). After m iterations, we have

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

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

Theorem 3.6 (Eigenvalues for triangular matrices). *The eigenvalues of a triangular matrix are the diagonal elements.*

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

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

To let $\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0$, we obtain $\lambda = a_{ii}$, for all $i \in [n]$. Therefore, the diagonal elements are the eigenvalues of the triangular matrix. \square

3.2 Eigenvalues and eigenvectors for symmetric matrices

Definition 14 (Antisymmetric matrices). The matrix \mathbf{A} is an antisymmetric matrix if

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

Theorem 3.7 (Eigenvalues for symmetric and antisymmetric matrices). *All the eigenvalues of a symmetric matrix with real entries only are real. All the eigenvalues of an antisymmetric matrix are imaginary.*

Proof. First, we prove by contradiction that a symmetric matrix with real entries only has real eigenvalues.

Let \mathbf{A} be a symmetric matrix with real entries only. Suppose λ is a complex eigenvalue of \mathbf{A} with a nonzero imaginary part, and x is the eigenvector associated with λ . Since all the entries of \mathbf{A} are real, the eigenvector x is also a complex vector with a nonzero imaginary part. Let $\bar{\lambda}$ and \bar{x} denote the conjugate eigenvalue and eigenvector, respectively. Then, we have

$$\bar{x}^T \mathbf{A} x = \bar{x}^T \lambda x, \quad (11)$$

and

$$x^T \mathbf{A} \bar{x} = x^T \bar{\lambda} \bar{x}. \quad (12)$$

By the symmetry of \mathbf{A} , we have

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

Subtracting the equation (12) from equation (11) yields

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

Since that $x^T \bar{x}$ is a nonzero real number, the imaginary part of λ is equal to 0, which contradicts the assumption that λ has a nonzero imaginary part. Therefore, the eigenvalues of \mathbf{A} are real.

Next, we prove that the eigenvalues for antisymmetric matrices are imaginary. Let \mathbf{A} be an antisymmetric matrix, λ be a complex eigenvalue of \mathbf{A} , and x be the complex eigenvector associated with λ . By the antisymmetry of \mathbf{A} , 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 that $x^T \bar{x}$ is a nonzero real number, the real part of λ is equal to 0. Therefore, the eigenvalues for an antisymmetric matrix are imaginary. \square

Theorem 3.8 (Orthogonality of eigenvectors for symmetric matrices). *The eigenvectors of a symmetric matrix associated with distinct eigenvalues are orthogonal.*

Proof. Let λ_1 and λ_2 be two distinct eigenvalues of \mathbf{A} , and x_1, x_2 are two eigenvectors associated with λ_1 and λ_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, we have

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

Since λ_1, λ_2 are distinct, 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). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Suppose λ_0 is a repeated eigenvalue of \mathbf{A} with multiplicity m , where $2 \leq m \leq n$. There exist m orthonormal eigenvectors associated with λ_0 .*

Proof. We prove the existence by construction. Let x_0 be an eigenvector associated with λ_0 . Then, there exist $n - 1$ additional orthogonal vectors $y_1, \dots, y_{n-1} \in \mathbb{R}^n$ such that $(x_0, y_1, \dots, y_{n-1})$ forms an orthonormal 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, we have $x_0^T \mathbf{A} = \lambda_0 x_0^T$. Consider the following matrix product,

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

Note that \mathbf{X} is an orthogonal matrix satisfying $\mathbf{X}^T = \mathbf{X}^{-1}$. The product $\mathbf{X}^T \mathbf{A} \mathbf{X}$ is a similar matrix of \mathbf{A} . By Theorem 3.2, λ_0 is also a repeated eigenvalue of $\mathbf{X}^T \mathbf{A} \mathbf{X}$ with multiplicity m . The characteristic polynomial of $\mathbf{X}^T \mathbf{A} \mathbf{X}$ is

$$P(\lambda) = \det(\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda \mathbf{I}_n) = (\lambda_0 - \lambda) \det(\mathbf{Y}^T \mathbf{A} \mathbf{Y} - \lambda \mathbf{I}_{n-1}). \quad (13)$$

Since λ_0 has multiplicity $m \geq 2$, the term $\det(\mathbf{Y}^T \mathbf{A} \mathbf{Y} - \lambda_0 \mathbf{I}_{n-1}) = 0$. 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 . Therefore, we have obtained two orthonormal eigenvectors of \mathbf{A} associated with λ_0 .

Now, replace x_0 by two orthogonal vectors $x_1 = \mathbf{X}v_1, x_2 = \mathbf{X}v_2$ and repeat the above steps.

After $(m-1)$ repetitions, we have obtained m orthogonal eigenvectors of \mathbf{A} associated with λ_0 . Let $\mathbf{X}^{(m)}, \mathbf{Y}^{(m)}$ denote the matrix \mathbf{X}, \mathbf{Y} after the $(m-1)$ -th repetition, respectively. The polynomial (13) becomes

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

Since λ_0 has multiplicity m , the determinant $\det(\mathbf{Y}^{(m),T} \mathbf{A} \mathbf{Y}^{(m)} - \lambda \mathbf{I}_{n-m}) \neq 0$. The dimension of the nullspace of $\mathbf{X}^{(m),T} \mathbf{A} \mathbf{X}^{(m)} - \lambda_0 \mathbf{I}_n$ is m . Therefore, we are unable to find the $(m+1)$ -th orthonormal eigenvector associated with λ_0 , and the repetition stops. \square

3.3 Diagonalization

Definition 15 (Diagonalization). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square matrix with n independent eigenvectors. We diagonalize the matrix \mathbf{A} as

$$\mathbf{A} = \mathbf{S}^{-1} \mathbf{\Lambda} \mathbf{S}, \quad (14)$$

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

Theorem 3.10 (Diagonalization of symmetric matrices). *Symmetric matrices are diagonalizable.*

Proof. By Theorems 3.8 and 3.9, for any symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, \mathbf{A} have n orthogonal eigenvectors with possibly repeated eigenvalues. Therefore, the matrix \mathbf{A} is able to be diagonalized as equation (14). \square

Theorem 3.11 (Descent of matrix powers). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square matrix with eigenvalues $\lambda_1, \dots, \lambda_n$, and n independent eigenvectors associated with the eigenvalues. The power of the matrix $\mathbf{A}^k \rightarrow 0$ as $k \rightarrow +\infty$ if and only if $|\lambda_i| < 1$, for all $i \in [n]$.*

The theorem says nothing about the eigenvectors.

Proof. Since \mathbf{A} has n independent eigenvectors, the matrix \mathbf{A} is diagonalizable. Let $\mathbf{A} = \mathbf{S}^{-1} \mathbf{\Lambda} \mathbf{S}$, where $\mathbf{\Lambda}, \mathbf{S}$ are defined the same as Definition 15. For all integer $k \geq 0$, we have $\mathbf{A}^k = \mathbf{S}^{-1} \mathbf{\Lambda}^k \mathbf{S}$. Then,

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

\square