

Linear Algebra – Part II

A summary for MIT 18.06SC

Jiaxin Hu

July 16, 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 $\underline{S}, \underline{T}$. The subspaces S and T are orthogonal, denoted $S \perp T$, if and only if for any $s \in S$ and $t \in T$, $s^T t = t^T s = 0$.

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: , \rightarrow . (Grammar mistake: two full sentences without a conjunction)

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

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

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

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

Can you think of a direct proof?

(I have drafted a direct proof, but I'd encourage you to think of one by yourself first.)

Principle: Direct proofs are always preferred over proofs by contradiction)

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

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

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

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 that $C(\mathbf{A}) \perp N(\mathbf{A}^T)$. 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 $\text{rank}(\mathbf{A}^T \mathbf{A}) = n - \dim(N(\mathbf{A}^T \mathbf{A})) = n - \dim(N(\mathbf{A})) = \text{rank}(\mathbf{A})$. \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 with independent columns. The vector $p \in C(\mathbf{A})$ that satisfies

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

is called the projection of vector x onto the column space $C(\mathbf{A})$. For all $x \in \mathbb{R}^n$ and the corresponding projection p , the matrix $\mathbf{P} \in \mathbb{R}^{m \times m}$ that satisfies

$$p = \mathbf{P}x,$$

is called the projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space $C(\mathbf{A})$.

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

You should precisely define p and hat x in this context.

$\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Shouldn't p and hat x depend on x?

Your current statement implies the choice of (p,\hat{x}) holds for "all" x, which is wrong.

Proof. By Definition 3, the projection $p \in C(\mathbf{A})$. Then, there exists a vector $\hat{x} \in \mathbb{R}^m$ such that $p = \mathbf{A}\hat{x}$. By equation (2), for all $x \in \mathbb{R}^n$, we have

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

Therefore, the matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ is the projection matrix of \mathbf{A} from \mathbb{R}^n onto the column space $C(\mathbf{A})$. \square

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

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

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

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

Similarly, we have

$$\mathbf{P}^2 = \mathbf{P}^T \mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{P}. \quad (\text{avoid stacking math expressions with no transition})$$

□

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 of \mathbf{A} from \mathbb{R}^n onto the left nullspace $N(\mathbf{A}^T)$.

Proof. By Definition 3, for any $x \in \mathbb{R}^n$, we have

$$x - \mathbf{P}x \perp C(\mathbf{A}) \Rightarrow (\mathbf{I} - \mathbf{P})x \perp C(\mathbf{A}).$$

By Theorem 1.1, the column space of \mathbf{A} is orthogonal to the left null space of \mathbf{A} . Then,

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

Therefore, $\mathbf{I} - \mathbf{P}$ is a projection matrix of \mathbf{A} from \mathbb{R}^n onto the left nullspace $N(\mathbf{A}^T)$. □

Suppose $y=x_1+x_2$ and x_1 in S. We cannot conclude $(y-x_1) \perp S$.

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

least-squares

the squared loss

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

“which” represents what? the projection Xhat\beta_{LS}, the column space of X, or the vector \hat{\beta}_{LS}?

Capital

“which” should come immediately after the word modified. (same rule as “only”)

where $\|\cdot\|$ is the euclidean norm. We consider the vector $\mathbf{X}\hat{\beta}_{LS}$ as 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 to solve the minimization problem.

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

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

first-order condition for the least-squares optimization.

The estimate $\hat{\beta}_{LS}$ is consistent 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*). 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} , becomes $\mathbf{P} = \mathbf{I}_m$.

the i-th column

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 column of matrix \mathbf{A} , for all $i \in [n]$. Gram-Schmidt process finds the orthonormal basis for $C(\mathbf{A})$ as following,

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

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

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

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

a

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

“a number ... with any matrix” ==> different matrices share a same number

2.1 Properties of determinants

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

an

from

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. , while
3. (a) If one row of the matrix is multiplied by a constant t , the determinant of the matrix is multiplied by t .

$$\left| \begin{array}{cc} ta & tb \\ c & d \end{array} \right| = t \left| \begin{array}{cc} a & b \\ c & d \end{array} \right| \quad \text{period}$$

by row

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

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

A matrix with two identical rows has determinant 0.

4. If the matrix contains two identical rows, 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 , 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, $\det(\mathbf{A}) = 0$. \square

a multiplier

5. Subtracting ~~t times~~ of the i -th row from the j -th row does not change the determinant of the matrix, where $i \neq j$.

Proof. Take a two-by-two matrix as an example. 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 matrices is similar. \square

matrices

6. If the matrix has a row that is all zeros, the determinant of the matrix is 0.

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

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 property 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. If the square matrix is singular, the determinant of the matrix is 0.

vice versa. A matrix is singular if and only if its determinant is zero

Proof. 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 and 6, $\det(RREF(\mathbf{A})) = 0$. Therefore, by property 5, we have $\det(\mathbf{A}) = \det(RREF(\mathbf{A})) = 0$. \square

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

two full sentences without conjectures

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

Second, let \mathbf{E} be an elimination matrix. The product \mathbf{EB} means implementing a linear operation on the rows of \mathbf{B} . By property 5, linear operation does not change the determinant of \mathbf{B} . Therefore, we have

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

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

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

Combine A with B..

Combining 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

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

□

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 transports $\mathbf{L}^T, \mathbf{U}^T$ 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, combining property 9, we have $\det(\mathbf{A}) = \det(\mathbf{L}) \det(\mathbf{U}) = \det(\mathbf{U}^T) \det(\mathbf{L}^T) = \det(\mathbf{U}^T \mathbf{L}^T) = \det(\mathbf{A}^T)$. combine ... with ... □

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 ~~necessary exchanges~~ 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} by the summation of n^n matrices, denoted \mathbf{A}'_i , where $i \in [n^n]$. For any \mathbf{A}'_i , there is only one entry comes from \mathbf{A} in each row and other entries 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.

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

There are $n!$ nonsingular matrices among matrices \mathbf{A}'_i . The number of nonsingular matrices coincides with the size of \mathcal{P} . Intuitively, there are n ways to choose an element from the first row,

Intuitively → Specifically (proof should be rigorous, but by intuition.)

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 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 necessary exchanges from $(\alpha_1, \alpha_2, \dots, \alpha_n)$ to $(1, 2, \dots, n)$. \square

Definition 7 (*Cofactors, cofactor matrix, and cofactor formula*). Let $\mathbf{A} = [[a_{ij}]] \in \mathbb{R}^{n \times n}$ 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 $\mathbf{C} = [[C_{ij}]] \in \mathbb{R}^{n \times n}$ is called ^acofactor matrix. The cofactor formula of $\det(\mathbf{A})$ is
 always add articles before singular countable nouns.

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

No articles before plural countable nouns.

Example 1 (*Tridiagonal matrix*). One example of using cofactor formula is computing the determinant of a *tridiagonal matrix*. The 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, 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, $\mathbf{T}_{n,-1,-1} = \mathbf{T}_{n-1}$, 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 $\mathbf{AC}^T = \det(\mathbf{A})\mathbf{I}_n$.

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

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

Therefore, the diagonal elements of \mathbf{AC}^T are $(\mathbf{AC}^T)_{ii} = \sum_k^n a_{ik}C_{ik} = \det(\mathbf{A})$, for all $i \in [n]$. Consider the off-diagonal entries $(\mathbf{AC}^T)_{ij} = \sum_k^n a_{ik}C_{jk}$, where $k \neq i \neq j$. 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. Combining 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*). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an invertible matrix, $x, b \in \mathbb{R}^n$ be vectors. Applying Theorem 2.1 to the linear system $\mathbf{Ax} = b$, we obtain the Cramer's rule of $x = \mathbf{A}^{-1}b$, two

$$x = \frac{1}{\det(\mathbf{A})}\mathbf{C}^T b \quad \text{and} \quad x_j = \frac{\det(\mathbf{B}_j)}{\mathbf{A}_j}, \quad \text{what is } x_j?$$

where \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,

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

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

Proposition 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 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*). Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a square matrix. Suppose there is a nonzero vector $x \in \mathbb{R}^m$ such that

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

The vector x is called the eigenvector of \mathbf{A} . 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*). Let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ be a square matrix. The trace of \mathbf{A} is defined as

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

Definition 12 (*Characteristic polynomial*). Let λ denote an eigenvalue of the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. 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,

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

Since the characteristic polynomial $P(\lambda)$ is of degree n , there are n solutions to the equation $P(\lambda) = 0$. The solutions to $P(\lambda) = 0$, denoted $\lambda_1, \dots, \lambda_n$, are ~~n~~ eigenvalues of \mathbf{A} , which may be complex. If the complex eigenvalues exist, the complex eigenvalues come in conjugate pairs, because $P(\bar{\lambda}) = P(\lambda)$ is also equal to 0, where $\bar{\lambda}$ is the conjugate of λ . Note that n eigenvalues are not necessarily distinct with each other. break into two sentences the

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 ~~n~~-eigenvalues of \mathbf{A} . Then,

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

and

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

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

the Issue 1. “compare” is a vt. should use “compare A with B”; issue 2. “compared” —> comparing where λ_i are ~~n~~ eigenvalues of \mathbf{A} , for all $i \in [n]$. By equation (7), the coefficient for λ^{n-1} is equal to $(-1)^n \sum_{i=1}^n \lambda_i$. Compared with Definition 12, we have $tr(\mathbf{A}) = \sum_{i=1}^n \lambda_i$. Similarly, the constant term in equation (7) is $(-1)^n \prod_{i=1}^n \lambda_1 \cdots \lambda_n$ while the constant term in Definition 12 is $(-1)^n \det(\mathbf{A})$. Therefore, we have $\prod_{i=1}^n \lambda_i = \det(\mathbf{A})$. □

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 a invertible matrix. The matrix product $\mathbf{B}\mathbf{A}\mathbf{B}^{-1}$ is called the similar matrix of \mathbf{A} .

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

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

exists

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

=> (we use one direction only?)

Since x_i is a nonzero vector, and \mathbf{B} is invertible, the vector $\mathbf{B}x_i$ is also a nonzero vector. Therefore, λ_i is an eigenvalue of $\mathbf{B}\mathbf{A}\mathbf{B}^{-1}$ with associated eigenvector $\mathbf{B}x_i$, for all $i \in [n]$. \square

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

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

any

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

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

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

, and (λ_i, x_i) in $\mathbb{R} \times \mathbb{R}^n$ be the i -th eigenvalue-eigenvector pair of A

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

Then, (\dots) is the i -th ... pair of A^{-1} . always think about shortening the statement
Proof. Since the nullspace of \mathbf{A} contains only a zero vector, the eigenvalue $\lambda_i > 0$, for all $i \in [n]$.

For all λ_i and x_i , multiplying x_i on both sides of the equation $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ yields

Since \mathbf{A} is invertible..... (Using the same statement as the assumption makes your proof easier to understand)

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

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

distinct

Theorem 3.5 (Independence of eigenvectors). *Let \mathbf{A} be a square matrix with 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 . Consider a linear combination $c_1v_1 + \dots + c_mv_m = 0$, where $c_i \in \mathbb{R}$, for all $i \in [m]$. Then we have

$$\mathbf{A}(c_1v_1 + \dots + c_mv_m) = c_1\lambda_1v_1 + \dots + c_m\lambda_mv_m = 0, \quad (8)$$

and

$$\lambda_1(c_1v_1 + \dots + c_mv_m) = c_1\lambda_1v_1 + \dots + c_m\lambda_1v_m = 0. \quad (9)$$

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 for a triangular matrix are the entries on the diagonal.*

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

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

obtain

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

3.2 Eigenvalues and eigenvectors for symmetric matrices

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

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

symmetric matrix must be square.

real entries only

Theorem 3.7 (Eigenvalues for symmetric and antisymmetric matrices). *All the eigenvalues of a symmetric square matrix with only real entries are real. All the eigenvalues of an antisymmetric square matrix are imaginary; i.e., $\lambda = bi$, where $b \in \mathbb{R}, i = \sqrt{-1}$.*

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

real entries only

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

$$\bar{x}^T \mathbf{A} x = \bar{x}^T \lambda x, \tag{11}$$

and

$$x^T \mathbf{A} \bar{x} = x^T \bar{\lambda} \bar{x}. \tag{12}$$

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

$$\begin{aligned} \bar{x}^T \mathbf{A} x &= x^T \mathbf{A}^T \bar{x} = x^T \mathbf{A} \bar{x}. \\ (12) &\qquad (11) \end{aligned}$$

Subtracting the equation (11) from equation (12) yields A-B: subtract B from A

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

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

Next, we prove that the eigenvalues for antisymmetric matrices are imaginary. Let \mathbf{A} be an anti-symmetric 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 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 $\mathbf{A} \in \mathbb{R}^{n \times n}$ associated with different eigenvalues are orthogonal.*

Proof. Let λ_1 and λ_2 be two different 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,

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

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). *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 .*

We prove the existance by construction.

break into two sentences

Proof. First, let x_0 be a nonzero eigenvector associated with 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 $y_j \perp x_0$, for all $j \in [(n-1)]$, 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, we have $x_0^T \mathbf{A} = \lambda_0 x_0^T$. Consider the following matrix product,

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

satisfying

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} . Hence, λ_0 is also a repeated eigenvalue of $\mathbf{X}^T \mathbf{A} \mathbf{X}$. The characteristic polynomial of $\mathbf{X}^T \mathbf{A} \mathbf{X} - \lambda \mathbf{I}_n$ is

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

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 associated with lambda_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. Finally, we will have m orthogonal eigenvectors of \mathbf{A} associated with λ_0 . \square

(until when?) you need to specify the stopping criteria.

Otherwise, why not (m+1), (m+2), or even more eigenvectors?

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

$$\mathbf{A} = \mathbf{S}^{-1} \Lambda \mathbf{S}, \quad \text{consisting of comma} \quad (13)$$

where $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose entries are 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} associated with eigenvalues in Λ .

ordered in the same as the eigenvalues in \Lambda

Theorem 3.10 (Diagonalization of symmetric matrices). *Let \mathbf{A} be a symmetric matrix. The matrix \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. Therefore, the matrix \mathbf{A} can be diagonalized as equation (13). \square

(think about the writing principles.. 1. reduces the number of verbs in sentences. 2 avoid "there are...")

why do you need “eigenvectors” in the theorem statement?

(principle: cut non-referred notations/definitions. check other parts..)

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 matrix $\mathbf{A}^k \rightarrow 0$ as $k \rightarrow +\infty$ if and only if $|\lambda_i| < 1$, for all $i \in [n]$.*

Proof. We diagonalize the matrix \mathbf{A} as $\mathbf{A} = \mathbf{S}^{-1}\Lambda\mathbf{S}$. For all integer $k \geq 0$, we have $\mathbf{A}^k = \mathbf{S}^{-1}\Lambda^k\mathbf{S}$.
Then,

(principle: each prooftheorem should be self-contained. check other parts)

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

□