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

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

, —> . (Grammar mistake: two full sentences without a conjection)

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

Theorem 1.1 (Orthogonality of matrix subspaces)**.** *Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a matrix. The row space $C(\boldsymbol{A}^T)$ and the nullspace $N(\boldsymbol{A})$ are orthogonal. The column space $C(\boldsymbol{A})$ and the left nullspace $N(\boldsymbol{A}^T)$ are orthogonal~~, i.e.,~~* . That is,

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

Proof. For any vector $x \in N(\boldsymbol{A})$, we have $\boldsymbol{A}x = 0$. Specifically,

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

where $a_i \in \mathbb{R}^n$ is the i-th row of $\boldsymbol{A}$, for all $i \in [m]$. By Definition 1, x is orthogonal to the rows in matrix $\boldsymbol{A}$. For any vector $v \in C(\boldsymbol{A}^T)$, v is a linear combination of the rows $v = c_1 a_1 + ... + c_m a_m$, where $c_i \in \mathbb{R}$, for all $i \in [m]$. Taking inner product between vectors v and x, we have

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

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

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

$$N(\boldsymbol{A}^T\boldsymbol{A}) = N(\boldsymbol{A})\overset{\text{and}}{\cancel{;}} \quad rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}).$$

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. Fxirst, we prove that $N(\boldsymbol{A}^T\boldsymbol{A}) = N(\boldsymbol{A})$.
On one hand, if $x \in N(\boldsymbol{A})$, then $\boldsymbol{A}x = 0$, which implies that $\boldsymbol{A}^T\boldsymbol{A}x = \boldsymbol{A}^T 0 = 0$. Therefore, for any $x \in N(\boldsymbol{A})$, the vector $x \in N(\boldsymbol{A}^T\boldsymbol{A})$.
On the other hand, we prove by contradiction that for any $x \in N(\boldsymbol{A}^T\boldsymbol{A})$ the vector $x \in N(\boldsymbol{A})$. Suppose there is a vector $x \in N(\boldsymbol{A}^T\boldsymbol{A}^T)$ but $x \notin N(\boldsymbol{A})$. We have

$$\boldsymbol{A}x = b \neq 0, \quad \boldsymbol{A}^T\boldsymbol{A}x = 0 \quad \Rightarrow \quad \boldsymbol{A}^T b = 0. \tag{1}$$

By the first equality in (1), $b \in C(\boldsymbol{A})$, and by the third equation in (1), $b \in N(\boldsymbol{A}^T)$. This contradicts the Theorem 1.1 that $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$. Therefore, for any $x \in N(\boldsymbol{A}^T\boldsymbol{A})$, the vector $x \in N(\boldsymbol{A})$.
Next, given $N(\boldsymbol{A}^T\boldsymbol{A}) = N(\boldsymbol{A})$, we have $rank(\boldsymbol{A}^T\boldsymbol{A}) = n - dim(N(\boldsymbol{A}^T\boldsymbol{A})) = n - dim(N(\boldsymbol{A})) = rank(\boldsymbol{A})$. □

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

Proof. Suppose $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a matrix with independent columns; i.e., $rank(\boldsymbol{A}) = n$. By Theorem 1.2, $rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}) = n$. Since $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a square matrix, $\boldsymbol{A}^T\boldsymbol{A}$ is invertible. □

1.2 Projections onto subspaces

Definition 3 (*Projection and projection matrix*)**.** Let $x \in \mathbb{R}^n$ be a vector and $\boldsymbol{A}^{n\times m}$ be a matrix with independent columns. The vector $p \in C(\boldsymbol{A})$ that satisfies

$$(x - p) \perp C(\boldsymbol{A}), \tag{2}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Similarly, we have

(avoid stacking math expressions with no transition)

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

□

Corollary 2 (Projection onto $N(\boldsymbol{A}^T)$)**.** Suppose $\boldsymbol{P}$ is a projection matrix in Theorem 1.3. Then $\boldsymbol{I} - \boldsymbol{P}$ is ~~also~~ a projection matrix of $\boldsymbol{A}$ from $\mathbb{R}^n$ onto the left nullspace $N(\boldsymbol{A}^T)$.

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

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

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

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

why?

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

Suppose y=x1+x2 and x1\in S. We cannot conclude (y-x1) \perp S.

1.3 Projection matrices and least squares

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

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

^T

where $\beta = (\beta_0, \beta_1, ..., \beta_k)$ 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

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

"...which" represents what? the projection X\hat \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 $\boldsymbol{X}\hat{\beta}_{LS}$ as the projection of y onto the column space of $\boldsymbol{X}$, which minimizes the distance from y to the column space $C(\boldsymbol{X})$. Therefore, we may use projection to solve the minimization problem.
By Definition 3, the projection $\boldsymbol{X}\hat{\beta}_{LS}$ satisfies

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

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 $q_1, ..., q_n$ 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 $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ be a matrix. The matrix $\boldsymbol{Q}$ is an orthonormal matrix, if the columns of $\boldsymbol{Q}$ are orthonormal. When $m = n$, the square matrix $\boldsymbol{Q}$ is an orthogonal matrix.

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

the i-th column

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

follows

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

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

$$
\boldsymbol{A} = [a_1, ..., a_n] = [e_1, ..., e_2]\begin{bmatrix} e_1^T a_1 & e_1^T a_2 & \cdots & e_1^T a_n \\ 0 & e_2^T a_2 & \cdots & e_2^T a_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e_n^T a_n \end{bmatrix} = \boldsymbol{Q}\boldsymbol{R}, \tag{3}
$$

where $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ is an orthonormal matrix and $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ 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 $\boldsymbol{A}$, let $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$ denote the determinant of matrix $\boldsymbol{A}$.

"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., $\det(\boldsymbol{I}) = 1$.
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.

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

period

by row

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

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$ 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 $\boldsymbol{A}$ has two identical rows a_i, a_j. The matrix after exchanging a_i, a_j, denoted $\boldsymbol{A}'$, is the same as the original matrix $\boldsymbol{A}$. By property 2, we have $\det(\boldsymbol{A}) = -\det(\boldsymbol{A}') = -\det(\boldsymbol{A})$. Therefore, $\det(\boldsymbol{A}) = 0$. □

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 matri~~x~~ is similar. □

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. □

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

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

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

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

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

9. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ be two square matrices, the determinant of the matrix product is $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$. two full sentences without conjections

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

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

$$|\boldsymbol{EB}| = |\boldsymbol{E}||\boldsymbol{B}| \tag{4}$$ period

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

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

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

Combine A with B..

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

Combing equation (4), we have

$$|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{E}_n' \cdots \boldsymbol{E}_1' \boldsymbol{B}| = |\boldsymbol{E}_n'||\boldsymbol{E}_{n-1}' \cdots \boldsymbol{E}_1' \boldsymbol{B}| = \cdots = |\boldsymbol{E}_n'| \cdots |\boldsymbol{E}_1'||\boldsymbol{B}|. \tag{5}$$

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

$$|\boldsymbol{E}_n'| \cdots |\boldsymbol{E}_1'| = |\boldsymbol{E}_n'| \cdots |\boldsymbol{E}_3'||\boldsymbol{E}_2' \boldsymbol{E}_1'| = \cdots = |\boldsymbol{E}_n' \cdots \boldsymbol{E}_1'|. \tag{6}$$

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

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

□

10. Let $\boldsymbol{A}$ be a square matrix, the determinant $\det(\boldsymbol{A}^T) = \det(\boldsymbol{A})$.

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

2.2 Determinant computation

Proposition 2 (Big formula for computing determinant)**.** *Let $\boldsymbol{A} = [\![a_{ij}]\!] \in \mathbb{R}^{n\times n}$ be a square matrix. The big formula for computing the determinant of $\boldsymbol{A}$ is*

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

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 $(1, 2, \dots, n)$.

swaps into the

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

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

There are $n!$ nonsingular matrices among matrices $\boldsymbol{A}_i'$. 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 $\boldsymbol{A}_i'$ follows the formula

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

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

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

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

a

The matrix $\boldsymbol{C} = [\![C_{ij}]\!] \in \mathbb{R}^{n\times n}$ is called cofactor matrix. The cofactor formula of $\det(\boldsymbol{A})$ is

always add articles before singular countable nouns.

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

j=1

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 $\boldsymbol{T}_n \in \mathbb{R}^{n\times n}$ denote the tridiagonal matrix of 1's; i.e.,

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

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

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

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

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

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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be an invertible matrix, and $\boldsymbol{C}$ be the cofactor matrix of $\boldsymbol{A}$. The inverse matrix of $\boldsymbol{A}$ satisfies*

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

Proof. It is equivalent to show that $\boldsymbol{A}\boldsymbol{C}^T = \det(\boldsymbol{A})\boldsymbol{I}_n$.
Let $\boldsymbol{A} = [\![a_{ij}]\!] \in \mathbb{R}^{n\times n}$ be an invertible matrix, and C_{ij} be the cofactor of a_{ij}. By the definition of cofactor matrix, we have

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

Therefore, the diagonal elements of $\boldsymbol{A}\boldsymbol{C}^T$ are $(\boldsymbol{A}\boldsymbol{C}^T)_{ii} = \sum_k^n a_{ik}C_{ik} = \det(\boldsymbol{A})$, for all $i \in [n]$. Consider the off-diagonal entries $(\boldsymbol{A}\boldsymbol{C}^T)_{ij} = \sum_k^n a_{ik}C_{jk}$, where k=1 $i \neq j$. 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 $\boldsymbol{A}\boldsymbol{C}^T$ are 0. Then, we have $\boldsymbol{A}\boldsymbol{C}^T = \det(\boldsymbol{A})\boldsymbol{I}_n$. ∎

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

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

what is x_j?

where $\boldsymbol{B}_j$ is the matrix $\boldsymbol{A}$ after replacing the j-th column by b.

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

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

We use $vol(P)$ to denote the volume of the parallelepiped P.

Proposition 3 (Determinants and volumes)**.** *The absolute determinant of a square matrix $\boldsymbol{A}$ is the volume of parallelepiped determined by the rows of $\boldsymbol{A}$; i.e.,*

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

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

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

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

∎

3 Eigenvalues and eigenvectors

Definition 10 (*Eigenvalues and eigenvectors*)**.** Let $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ be a square matrix. Suppose there is a nonzero vector $x \in \mathbb{R}^m$ such that

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

The vector x is called the eigenvector of $\boldsymbol{A}$. The value λ is called the eigenvalue of $\boldsymbol{A}$, 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 $\boldsymbol{A}$.

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

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

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

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

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 $\lambda_1, ..., \lambda_n$, are $\cancel{n}$ eigenvalues of $\boldsymbol{A}$, which may be complex. If the complex eigenvalues exist, the complex eigenvalues come in conjugate pairs, because $P(\bar{\lambda}) = \overline{P(\lambda)}$ 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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a square matrix, and $\lambda_1, ...\lambda_n$ be the ~~n~~ eigenvalues of $\boldsymbol{A}$. Then,*

and

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

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

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

the Issue 1. “compare” is an vt. should use “compare A with B” ; issue 2. “compared” —> comparing

where λ_i are $\cancel{n}$ eigenvalues of $\boldsymbol{A}$, 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 $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$. Similarly, the constant term in equation (7) is $(-1)^n \prod_{i=1}^{n} \lambda_1 \cdots \lambda_n$ while the constant term in Defintion 12 is $(-1)^n \det(\boldsymbol{A})$. Therefore, we have $\prod_{i=1}^{n} \lambda_i = \det(\boldsymbol{A})$. □

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

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

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

exists

=> (we use one direction only?)

$$\lambda_i x_i = \boldsymbol{A} x_i \quad \Leftrightarrow \quad \lambda_i \boldsymbol{B} x_i = \boldsymbol{BAB}^{-1} \boldsymbol{B} x_i.$$

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

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

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

any

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

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

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

, and (\lambda_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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be an invertible matrix, $\lambda_1, ..., \lambda_n$ be the eigenvalues of $\boldsymbol{A}$, and $x_1, ...x_n$ be the eigenvectors associated with $\lambda_1, ..., \lambda_n$. Then, $\lambda_1^{-1}, ..., \lambda_n^{-1}$ are the eigenvalues of $\boldsymbol{A}^{-1}$, and $x_1, ...x_n$ are eigenvectors associated with $\lambda_1^{-1}, ..., \lambda_n^{-1}$.*

Then, (….) is the i-th … pair of A^{-1}. always think about shortening the statement

Proof. Since ~~the nullspace of $\boldsymbol{A}$ 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 $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$ yields

Since A is invertible….. (Using the same statement as the assumption makes your proof easier to understand)

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

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

distinct

Theorem 3.5 (Independence of eigenvectors)**.** *Let $\boldsymbol{A}$ be a square matrix with at least two different eigenvalues. The eigenvectors associated with different eigenvalues are independent.*

stick with one choice of word. (you use distinct in Def 13)

Proof. Suppose the square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has m different eigenvalues $\lambda_1, ..., \lambda_m$ with associated eigenvectors $v_1, ..., v_m$. Consider a linear combination $c_1v_1 + ... + c_mv_m = 0$, where $c_i \in \mathbb{R}$, for all $i \in [m]$. Then we have

$$\boldsymbol{A}(c_1v_1 + ... + c_mv_m) = c_1\lambda_1v_1 + ... + c_m\lambda_mv_m = 0, \tag{8}$$

and

rewrite in terms of vector/matrix notion, not in entry-wise fashion.

$$\lambda_1(c_1v_1 + ... + c_mv_m) = c_1\lambda_1v_1 + ... + c_m\lambda_1v_m = 0. \tag{9}$$

Subtracting the equations (8) from (9) yields

$$c_2(\lambda_1 - \lambda_2)v_2 + \cdots + c_m(\lambda_1 - \lambda_m)v_m = 0. \tag{10}$$

Consider the new linear combination $c_2(\lambda_1 - \lambda_2)v_2 + \cdots + c_m(\lambda_1 - \lambda_m)v_m = 0$ 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 $\lambda_1, ..., \lambda_m$ are distinct, and v_m is a nonzero vector, the coefficient $c_m = 0$. Plugging $c_m = 0$ back to previous steps, we have $c_1 = \cdots = c_m = 0$. Therefore, $v_1, ..., v_m$ are independent. □

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

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

and

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

obtain

To let $\det(\boldsymbol{A} - \lambda \boldsymbol{I}_n) = 0$, 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 $\boldsymbol{A}$ is an antisymmetric matrix if $\boldsymbol{A}$ satisfies

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

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

$$\overline{x}^T \boldsymbol{A} x = \overline{x}^T \lambda x, \tag{11}$$

and

$$x^T \boldsymbol{A} \overline{x} = x^T \overline{\lambda} \overline{x}. \tag{12}$$

By the symmetry of $\boldsymbol{A}$, we have

$$\overline{x}^T \boldsymbol{A} x = x^T \boldsymbol{A}^T \overline{x} = x^T \boldsymbol{A} \overline{x}.$$

(12) (11)

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

$$0 = (\overline{x}^T \boldsymbol{A} x - x^T \boldsymbol{A} \overline{x}) = (\lambda - \overline{\lambda}) x^T \overline{x}.$$

Since that $x^T \overline{x}$ 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 $\boldsymbol{A}$ be an antisymmetric matrix, λ be a complex eigenvalue of $\boldsymbol{A}$, and x be the complex eigenvector associated with λ. By the antisymmetry of $\boldsymbol{A}$, we have

$$\overline{x}^T \boldsymbol{A} x = x^T \boldsymbol{A}^T \overline{x} = -x^T \boldsymbol{A} \overline{x}.$$

Then,

$$0 = (\overline{x}^T \boldsymbol{A} x + x^T \boldsymbol{A} \overline{x}) = (\lambda + \overline{\lambda}) x^T \overline{x}.$$

Since that $x^T \overline{x}$ 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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ associated with different eigenvalues are orthogonal.*

Proof. Let λ_1 and λ_2 be two different eigenvalues of $\boldsymbol{A}$, and x_1, x_2 are two eigenvectors associated with λ_1 and λ_2, respectively. By the symmetry of $\boldsymbol{A}$, we have

$$x_2^T \boldsymbol{A} x_1 = x_1^T \boldsymbol{A}^T x_2 = x_1^T \boldsymbol{A} x_2.$$

Then,

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

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. □

Theorem 3.9 (Eigenvectors for repeated eigenvalues of symmetric matrices)**.** *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Suppose λ_0 is a repeated eigenvalue of $\boldsymbol{A}$ 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 $y_1, ..., y_{n-1}$ such that $y_j \perp x_0$, for all $j \in [(n-1)]$, and $(x_0, y_1, ..., y_{n-1})$ forms the basis of $\mathbb{R}^n$. Let $\boldsymbol{Y} = [y_1, ... y_{n-1}]$ and $\boldsymbol{X} = [x_0, Y]$. Since $\boldsymbol{A}$ is a symmetric matrix, we have $x_0^T \boldsymbol{A} = \lambda_0 x_0^T$. Consider the following matrix product,

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

Note that $\boldsymbol{X}$ is an orthogonal matrix ~~which satisfies~~ satisfying $\boldsymbol{X}^T = \boldsymbol{X}^{-1}$. By Theorem 3.2, $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}$ is a similar matrix of $\boldsymbol{A}$. Hence, λ_0 is also a repeated eigenvalue of $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}$. The characteristic polynomial of $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X} - \lambda \boldsymbol{I}_n$ is

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

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

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

(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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a square matrix with n independent eigenvectors. We diagonalize the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{S}^{-1} \Lambda \boldsymbol{S}, \tag{13}$$

, comma

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

ordered in the same as the eigenvalues in \Lambda

Theorem 3.10 (Diagonalization of symmetric matrices)**.** *Let $\boldsymbol{A}$ be a symmetric matrix. The matrix $\boldsymbol{A}$ can be diagonalized.*

Proof. By Theorems 3.8 and Theorem 3.9, for any symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A}$ have n orthogonal eigenvectors, even though there may have repeated eigenvalues. Therefore, the matrix $\boldsymbol{A}$ can be diagonalized as equation (13). □

with possibly repeated eigenvalues

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

why do you need to mention "eigenvectors" in the theorem statement?
(principle: cut non-referred notations/definitions. check other parts..)

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

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

(principle: each proof/theorem should be self-contained. check other parts)

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

$\square$