

MA1522 Notes (AY24/25 Sem1)

Michael Yang

September 3, 2024

Chapter 4: Subspaces Associated to a Matrix

4.1 Column Space, Row Space, and Nullspace

Definition

Let \mathbf{A} be an $m \times n$ matrix,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

The **row space** of \mathbf{A} is the subspace of \mathbb{R}^n spanned by the rows of \mathbf{A} ,

$$\text{Row}(\mathbf{A}) = \text{span}\{(a_{11} \ a_{12} \ \cdots \ a_{1n}), (a_{21} \ a_{22} \ \cdots \ a_{2n}), \dots, (a_{m1} \ a_{m2} \ \cdots \ a_{mn})\}$$

The **column space** of \mathbf{A} is the subspace of \mathbb{R}^m spanned by the columns of \mathbf{A} ,

$$\text{Col}(\mathbf{A}) = \text{span} \left\{ \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}$$

Remark: May write the vectors in row space as column vectors.

Theorem

(Row operations preserve row space)

Suppose \mathbf{A} and \mathbf{B} are **row equivalent matrices**. Then $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{B})$.

Theorem

(Basis for row space)

For any matrix \mathbf{A} , the **nonzero rows** of the **reduced row-echelon form** of \mathbf{A} form a **basis** for the row space of \mathbf{A} .

Theorem

(Row operations preserve linear relations between columns)

Let $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n)$ and $\mathbf{B} = (\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n)$ be row equivalent $m \times n$ matrices, where \mathbf{a}_i and \mathbf{b}_i is the i -th column of \mathbf{A} and \mathbf{B} , respectively, for $i = 1, \dots, n$. Then for any coefficients c_1, c_2, \dots, c_n ,

$$c_1 \mathbf{a}_1 + c_2 \mathbf{a}_2 + \cdots + c_n \mathbf{a}_n = \mathbf{0}$$

if and only if

$$c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n = \mathbf{0}$$

Theorem

(Basis for column space)

Suppose \mathbf{R} is the reduced row-echelon form of a matrix \mathbf{A} . Then the columns of \mathbf{A} corresponding to the pivot columns in \mathbf{R} form a basis for the column space of \mathbf{A} .

The column space is the set of vectors \mathbf{v} such that $\mathbf{Ax} = \mathbf{v}$ is consistent, or the set of vectors \mathbf{v} such that $\mathbf{v} = \mathbf{Au}$ for some \mathbf{u} ,

$$\text{Col}(\mathbf{A}) = \{\mathbf{v} = \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^k\} = \{\mathbf{v} \mid \mathbf{Ax} = \mathbf{v} \text{ is consistent}\}.$$

Definition

The **nullspace** of a $m \times n$ matrix \mathbf{A} is the solution space to the homogeneous system $\mathbf{Ax} = \mathbf{0}$ with coefficient matrix \mathbf{A} . It is denoted as

$$\text{Null}(\mathbf{A}) = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{Av} = \mathbf{0}\}.$$

Whenever we come across a subspace, we are interested in its dimensions.

The **nullity** of \mathbf{A} is the dimension of the nullspace of \mathbf{A} , denoted as

$$\text{nullity}(\mathbf{A}) = \dim(\text{Null}(\mathbf{A}))$$

4.2 Rank

Theorem

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{R} its reduced row-echelon form.

$$\begin{aligned} \dim(\text{Col}(\mathbf{A})) &= \# \text{ of pivot columns in RREF of } \mathbf{A}, \\ &= \# \text{ of leading entries in RREF of } \mathbf{A}, \\ &= \# \text{ of nonzero rows in RREF of } \mathbf{A} = \dim(\text{Row}(\mathbf{A})) \end{aligned}$$

Definition

Define the **rank** of \mathbf{A} to be the dimension of its column space or row space

$$\text{rank}(\mathbf{A}) = \dim(\text{Col}(\mathbf{A})) = \dim(\text{Row}(\mathbf{A}))$$

Theorem

Rank is invariant under transpose,

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$$

Theorem

The linear system $\mathbf{Ax} = \mathbf{b}$ is **consistent** if and only if the rank of \mathbf{A} is equal to the rank of the augmented matrix $(\mathbf{A} \mid \mathbf{b})$,

$$\text{rank}(\mathbf{A}) = \text{rank}((\mathbf{A} \mid \mathbf{b})).$$

Lemma

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{B} a $n \times p$ matrix. The column space of the product \mathbf{AB} is a subspace of the column space of \mathbf{A} ,

$$\text{Col}(\mathbf{AB}) \subseteq \text{Col}(\mathbf{A})$$

Theorem

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{B} a $n \times p$ matrix. Then

$$\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$$

Theorem

If \mathbf{A} and \mathbf{B} are **row equivalent** matrices, then $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B})$.

Theorem

(Rank-Nullity Theorem)

Let \mathbf{A} be a $m \times n$ matrix. The sum of its rank and nullity is equal to the number of columns,

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n$$

Theorem

(Equivalent Statements of Invertibility)

12. \mathbf{A} is of full rank, $\text{rank}(\mathbf{A}) = n$.
13. $\text{nullity}(\mathbf{A}) = 0$.

Theorem

(Full Rank Equals Number of Columns)

Suppose \mathbf{A} is a $m \times n$ matrix. The following statements are equivalent.

1. \mathbf{A} is full rank, where the rank is equal to the number of columns, $\text{rank}(\mathbf{A}) = n$.
2. The rows of \mathbf{A} spans \mathbb{R}^n , $\text{Row}(\mathbf{A}) = \mathbb{R}^n$.
3. The columns of \mathbf{A} are linearly independent.
4. The homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, that is, $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.
5. $\mathbf{A}^T \mathbf{A}$ is an invertible matrix of order n .
6. \mathbf{A} has a left inverse.

The reduced row-echelon form of \mathbf{A} is

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{0}_{(m-n) \times n} \end{pmatrix}$$

Theorem

(Full Rank Equals Number of Rows)

Suppose \mathbf{A} is a $m \times n$ matrix. The following statements are equivalent.

1. \mathbf{A} is full rank, where the rank is equal to the number of columns, $\text{rank}(\mathbf{A}) = m$.
2. The columns of \mathbf{A} spans \mathbb{R}^m , $\text{Col}(\mathbf{A}) = \mathbb{R}^m$.
3. The rows of \mathbf{A} are linearly independent.
4. The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$.
5. $\mathbf{A}\mathbf{A}^T$ is an invertible matrix of order m .
6. \mathbf{A} has a right inverse.

The reduced row-echelon form of \mathbf{A} is

$$\mathbf{R} = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots \\ \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots \end{pmatrix}$$

Chapter 5: Orthogonality and Least Square Solution

5.1 Orthogonality

Definition

Two vectors \mathbf{u}, \mathbf{v} in \mathbb{R}^n are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

In this case, either one of the vectors is the zero vector, or that they are **perpendicular**.

Definition

A set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ of vectors is **orthogonal** if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for every $i \neq j$, that is, vectors in S are **pairwise orthogonal**.

A set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ of vectors is **orthonormal** if for all $i, j = 1, \dots, k$,

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

That is, S is **orthogonal**, and all the vectors are **unit vectors**.

Note

Orthogonal set can contain zero vector $\mathbf{0}$.

Orthonormal set cannot contain $\mathbf{0}$.

Definition

Let V be a subspace of \mathbb{R}^n . A vector $\mathbf{n} \in \mathbb{R}^n$ is **orthogonal** to V if for every \mathbf{v} in V , $\mathbf{n} \cdot \mathbf{v} = 0$, that is, \mathbf{n} is **orthogonal** to every vector in V . We will denote it as $\mathbf{n} \perp V$.

Theorem

Let V be a subspace of \mathbb{R}^n and $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a spanning set for V , $\text{span}(S)=V$. Then a vector \mathbf{w} is **orthogonal** to V if and only if $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all $i = 1, \dots, k$.

Theorem

Let V be a subspace of \mathbb{R}^n and $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a spanning set for V . Then \mathbf{w} is **orthogonal** to V if and only if \mathbf{w} is in the nullspace of \mathbf{A}^T , where $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k)$;

$$\mathbf{w} \perp V \quad \Leftrightarrow \quad \mathbf{w} \in \text{Null}(\mathbf{A}^T)$$

Definition

Let V be a subspace of \mathbb{R}^n . The **orthogonal complement** of V is the set of all vectors that are **orthogonal** to V , and is denoted as

$$V^\perp = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \text{ in } V\}$$

Theorem

Let V be a subspace of \mathbb{R}^n and $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a spanning set for V . Let $\mathbf{A} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k)$. Then the **orthogonal complement** of V is the nullspace of \mathbf{A}^T ,

$$V^\perp = \text{Null}(\mathbf{A}^T)$$

Note

Let \mathbf{A} be a $m \times n$ matrix. The nullspace of \mathbf{A} is the orthogonal complement of the row space of \mathbf{A} ,

$$\text{Row}(\mathbf{A})^\perp = \text{Null}(\mathbf{A})$$

5.2 Orthogonal and Orthonormal Bases

Definition

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is an **orthogonal set** of **nonzero** vectors. Then S is linearly independent.

Theorem

Every **orthonormal set** is **linearly independent**.

Definition

Let V be a subspace of \mathbb{R}^n . A set $S \subseteq V$ is an **orthogonal basis** (resp, **orthonormal basis**) of V if S is a basis of V and S is an **orthogonal** (resp, **orthonormal**) set.

Theorem

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be an orthogonal basis for a subspace V of \mathbb{R}^n . Then for any $\mathbf{v} \in V$,

$$\mathbf{v} = \left(\frac{\mathbf{v} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \right) \mathbf{u}_1 + \left(\frac{\mathbf{v} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \right) \mathbf{u}_2 + \dots + \left(\frac{\mathbf{v} \cdot \mathbf{u}_k}{\|\mathbf{u}_k\|^2} \right) \mathbf{u}_k$$

If further S is an **orthonormal basis**, then

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{v} \cdot \mathbf{u}_2) \mathbf{u}_2 + \dots + (\mathbf{v} \cdot \mathbf{u}_k) \mathbf{u}_k$$

that is, S orthogonal, $[\mathbf{v}]_S = \begin{pmatrix} \frac{\mathbf{v} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \\ \frac{\mathbf{v} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \\ \vdots \\ \frac{\mathbf{v} \cdot \mathbf{u}_k}{\|\mathbf{u}_k\|^2} \end{pmatrix}$, S orthonormal, $[\mathbf{v}]_S = \begin{pmatrix} \mathbf{v} \cdot \mathbf{u}_1 \\ \mathbf{v} \cdot \mathbf{u}_2 \\ \vdots \\ \mathbf{v} \cdot \mathbf{u}_k \end{pmatrix}$.

Note that this only works if $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is an orthogonal or orthonormal basis.

Note

Let V be a subspace of \mathbb{R}^n and S an **orthonormal basis** of V . For any $\mathbf{u}, \mathbf{v} \in V$,

1. $\mathbf{u} \cdot \mathbf{v} = [\mathbf{u}]_S \cdot [\mathbf{v}]_S$
2. $\|\mathbf{u} - \mathbf{v}\| = \|[u]_S - [v]_S\|$

Definition

A $n \times n$ square matrix \mathbf{A} is **orthogonal** if $\mathbf{A}^T = \mathbf{A}^{-1}$, equivalently, $\mathbf{A}^T \mathbf{A} = \mathbf{I} = \mathbf{A} \mathbf{A}^T$.

Theorem

Let \mathbf{A} be a square matrix of order n . The following statements are equivalent.

1. \mathbf{A} is an **orthogonal matrix**.
2. The **columns** of \mathbf{A} form an **orthonormal basis** for \mathbb{R}^n .
3. The **rows** of \mathbf{A} form an **orthonormal basis** for \mathbb{R}^n .

Note

The term 'orthonormal matrix' is not used.

Question

Let W be a subspace of dimension 3. We can never find an orthonormal subset of W containing 4 vectors. (**T**)

Orthonormal set is linearly independent and if W contains a set of 4 linearly independent vectors, then $3 = \dim(W) \geq 4$, a contradiction. An orthonormal set is linearly independent. Also, if U and V are subspaces such that $U \subseteq V$, then $\dim(U) \leq \dim(V)$.

Question

Which is true regarding an orthogonal set S containing 3 non-zero vectors in \mathbb{R}^3 ?

1. The set S must be linearly independent (**T**)
2. S is a basis for \mathbb{R}^3 (**T**)
3. Each pair of vectors in S are perpendicular to each other (**T**)
4. The set S must span \mathbb{R}^3 (**T**)

Nonzero orthogonal vectors are perpendicular to each other, and is thus linearly independent.

Question

An orthogonal set must be linearly independent. (**T**)

Orthogonal set can contain the zero vector, which makes the set linearly dependent.

Question

Let $S = \left\{ \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \right\}$ be a basis for a subspace V in \mathbb{R}^3 . Let $\mathbf{v} = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$. What is the norm of $[\mathbf{v}]_S$, $\|[\mathbf{v}]_S\|$?

$\sqrt{3^2 + 4^2} = 5$. If S is an orthonormal basis for V , then for any vector $v \in V$, $\|\mathbf{v}\| = \|[\mathbf{v}]_S\|$.

Question

A square matrix \mathbf{A} of order n is orthogonal if the columns or rows of \mathbf{A} form an orthogonal basis for \mathbb{R}^n . (F)

The columns and/or columns need to form an orthonormal basis, not an orthogonal basis, in order for \mathbf{A} to be orthogonal.

5.3 Orthogonal Projection

Theorem

Orthogonal projection theorem

Let V be a subspace of \mathbb{R}^n . Every vector \mathbf{w} in \mathbb{R}^n can be decomposed **uniquely** as a sum

$$\mathbf{w} = \mathbf{w}_p + \mathbf{w}_n$$

where \mathbf{w}_n is orthogonal to V and \mathbf{w}_p is in V . Moreover, if $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is an **orthogonal basis** for V , then

$$\mathbf{w}_p = \frac{\mathbf{w} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{w} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \dots + \frac{\mathbf{w} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \mathbf{u}_k.$$

Definition

Define the vector \mathbf{w}_p in the theorem above as the **orthogonal projection** (or just **projection**) of \mathbf{w} onto the subspace V .

Theorem

Best Approximation Theorem

Let V be a subspace of \mathbb{R}^n and \mathbf{w} a vector in \mathbb{R}^n . Let \mathbf{w}_p be the projection of \mathbf{w} onto V . Then \mathbf{w}_p is a vector in V closest to \mathbf{w} ; that is,

$$\|\mathbf{w} - \mathbf{w}_p\| \leq \|\mathbf{w} - \mathbf{v}\|$$

for all \mathbf{v} in V .

Definition

Gram-Schmidt Orthogonalization

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a linearly independent set. Let

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 \\ \mathbf{v}_2 &= \mathbf{u}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|^2} \right) \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{u}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_3}{\|\mathbf{v}_1\|^2} \right) \mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_3}{\|\mathbf{v}_2\|^2} \right) \mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_k &= \mathbf{u}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_k}{\|\mathbf{v}_1\|^2} \right) \mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_k}{\|\mathbf{v}_2\|^2} \right) \mathbf{v}_2 - \dots - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{u}_k}{\|\mathbf{v}_{k-1}\|^2} \right) \mathbf{v}_{k-1} \end{aligned}$$

Then $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is an **orthogonal** set (of nonzero vectors), and hence,

$$\left\{ \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \dots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|} \right\}$$

is an **orthonormal** set such that $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$.

5.4 QR Factorization

Definition

Suppose now \mathbf{A} is a $m \times n$ matrix with linearly independent columns, i.e. $\text{rank}(\mathbf{A}) = n$. Write

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n).$$

Since the set $S = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ is **linearly independent** we may apply the **Gram-Schmidt process** on S to obtain an **orthonormal** set $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$. Set

$$\mathbf{Q} = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_n).$$

Recall that for any $j = 1, 2, \dots, n$, $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_j\} = \text{span}\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j\}$. In particular, \mathbf{a}_j is in $\text{span}\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j\}$. Thus we may write

$$\mathbf{a}_j = r_{1j}\mathbf{q}_1 + r_{2j}\mathbf{q}_2 + \dots + r_{jj}\mathbf{q}_j + 0\mathbf{q}_{j+1} + \dots + 0\mathbf{q}_n = (\mathbf{q}_1 \quad \dots \quad \mathbf{q}_j \quad \dots \quad \mathbf{q}_n) \begin{pmatrix} r_{1j} \\ \vdots \\ r_{jj} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Explicitly,

$$\begin{aligned}\mathbf{a}_1 &= r_{11}\mathbf{q}_1 = (\mathbf{q}_1 \quad \dots \quad \mathbf{q}_n) \begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ \mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 = (\mathbf{q}_1 \quad \dots \quad \mathbf{q}_n) \begin{pmatrix} r_{12} \\ r_{22} \\ \vdots \\ 0 \end{pmatrix} \\ &\vdots \\ \mathbf{a}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \dots + r_{nn}\mathbf{q}_n = (\mathbf{q}_1 \quad \dots \quad \mathbf{q}_n) \begin{pmatrix} r_{1n} \\ r_{2n} \\ \vdots \\ r_{nn} \end{pmatrix}\end{aligned}$$

Thus, we may write

$$\begin{aligned}A &= (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n) \\ &= (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_n) \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{pmatrix} \\ &= \mathbf{Q}\mathbf{R}\end{aligned}$$

for some $m \times n$ matrix \mathbf{Q} with **orthonormal columns**, and a **upper triangular** $n \times n$ matrix \mathbf{R} .

Note

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n.$$

The diagonal entries of \mathbf{R} are positive, $r_{ii} > 0$ for all $i = 1, 2, \dots, n$.

The upper triangular matrix \mathbf{R} is invertible.

Theorem

(QR Factorization)

Suppose \mathbf{A} is a $m \times n$ matrix with **linearly independent** columns. Then \mathbf{A} can be written as $\mathbf{A} = \mathbf{Q}\mathbf{R}$ for some $m \times n$ matrix \mathbf{Q} such that $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n$ and **invertible** upper triangular matrix \mathbf{R} with **positive** diagonal entries.

Definition

The decomposition given in the theorem above is called a **QR factorization** of \mathbf{A} .

Algorithm to QR Factorization

Let \mathbf{A} be a $m \times n$ matrix with **linearly independent** columns.

1. Perform Gram-Schmidt on the columns of $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n)$ to obtain an **orthonormal set** $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$.
2. Set $\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$.
3. Compute $\mathbf{R} = \mathbf{Q}^T \mathbf{A}$.

Corollary

Suppose \mathbf{A} is a $m \times n$ matrix with **linearly independent** columns, i.e. $\text{rank}(\mathbf{A}) = n$. Then $\mathbf{A}^T \mathbf{A}$ is invertible, and \mathbf{A} has a **left inverse**; that is, there is a \mathbf{B} such that $\mathbf{B}\mathbf{A} = \mathbf{A}_n$

5.5 Least Square Approximation

Definition

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{b} a vector in \mathbb{R}^m . A vector \mathbf{u} in \mathbb{R}^n is a **least square solution** of $\mathbf{A}\mathbf{x} = \mathbf{b}$ if for every vector $\mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}\| \leq \|\mathbf{A}\mathbf{v} - \mathbf{b}\|$$

Theorem

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{b} a vector in \mathbb{R}^m . A vector \mathbf{u} in \mathbb{R}^n is a **least square solution** to $\mathbf{A}\mathbf{x} = \mathbf{b}$ if and only if \mathbf{u} is the **projection** of \mathbf{b} onto the column space of \mathbf{A} , $\text{Col}(\mathbf{A})$.

Theorem

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{b} a vector in \mathbb{R}^m . A vector \mathbf{u} in \mathbb{R}^n is a **least square solution** to $\mathbf{A}\mathbf{x} = \mathbf{b}$ if and only if \mathbf{u} is a solution to $\mathbf{A}^T \mathbf{A}\mathbf{x} = \mathbf{A}^T \mathbf{b}$.

Note

Least square solutions are not unique, but projection is unique.

Note

Let \mathbf{A} be a $m \times n$ matrix and \mathbf{b} a vector in \mathbb{R}^m . For any choice of **least square solution** \mathbf{u} , that is, for any solution \mathbf{u} of $\mathbf{A}^T \mathbf{A}\mathbf{x} = \mathbf{A}^T \mathbf{b}$, the projection $\mathbf{A}\mathbf{u}$ is unique.

Theorem

Let V be a subspace of \mathbb{R}^n and $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a **spanning set** for V . Set $\mathbf{A} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k)$. Let \mathbf{w} be a vector in \mathbb{R}^n , and \mathbf{u} be a least square solution to $\mathbf{A}\mathbf{x} = \mathbf{w}$. Then $\mathbf{w}_p = \mathbf{A}\mathbf{u}$ is the **orthogonal projection** of a vector \mathbf{w} onto V .

In particular, if $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is a **basis** for V , then the **orthogonal projection** of a vector \mathbf{w} onto V is

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

Note

For any $m \times n$ matrix \mathbf{A} and $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ is always consistent. (The rank of $\mathbf{A}^T \mathbf{A}$ is always equal to $(\mathbf{A}^T \mathbf{A} \quad \mathbf{A}^T \mathbf{b})$)

Question

Let \mathbf{u} be a solution to $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$. Which of the following statement is false?

- \mathbf{u} is a least square solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$ (T)
- $\mathbf{A} \mathbf{u} - \mathbf{b}$ is orthogonal to the column space of \mathbf{A} (T)
- \mathbf{u} is the unique solution to $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ (F)
- For any vector $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{A} \mathbf{u} - \mathbf{b}\| \leq \|\mathbf{A} \mathbf{v} - \mathbf{b}\|$ (T)
- $\mathbf{A} \mathbf{u}$ is the projection of \mathbf{b} onto the column space of \mathbf{A} (T)

Question

If \mathbf{u} is a solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$, then \mathbf{u} is a least square solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$. (T)

Question

Suppose $\text{rank}(\mathbf{A}) = \text{number of columns of } \mathbf{A}$. Which of the statements is always true?

- $\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ is the unique least square solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$ (F)
- $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ is the unique least square solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$ (T)
- $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ is the unique solution to $\mathbf{A} \mathbf{x} = \mathbf{b}$ (F)

Full rank implies $\mathbf{A}^T \mathbf{A}$ is invertible.

6 Eigenanalysis

6.1 Eigenvalues and Eigenvectors

Definition

Let \mathbf{A} be a **square** matrix of order n . A real number λ is an **eigenvalue** of \mathbf{A} if there is a **nonzero** vector \mathbf{v} in \mathbb{R}^n , $\mathbf{v} \neq 0$, such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

In this case, the nonzero vector \mathbf{v} is called an **eigenvector** associated to λ . Let \mathbf{A} be a **square** matrix of order n , the **characteristic polynomial** of \mathbf{A} , denoted as $\text{char}(\mathbf{A})$, is the degree n polynomial

$$\det(x\mathbf{I} - \mathbf{A}).$$

Theorem

Let \mathbf{A} be a **square** matrix of order n . $\lambda \in \mathbb{R}^n$ is an **eigenvalue** of \mathbf{A} if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A}\mathbf{x} = \mathbf{0})$ has **nontrivial** solutions.

Theorem

Let \mathbf{A} be a **square** matrix of order n . λ is an **eigenvalue** of \mathbf{A} if and only if λ is a **root** of the **characteristic polynomial** $\det(x\mathbf{I} - \mathbf{A})$.

Theorem

(Equivalent statements for invertibility)

14. A **square** matrix \mathbf{A} is **invertible** if and only if $\lambda = 0$ is **not** an **eigenvalue** of \mathbf{A} .

Definition

Let λ be an eigenvalue of \mathbf{A} . The **algebraic multiplicity** of λ is the **largest integer** r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x)$$

for some polynomial $p(x)$. Alternatively, r_λ is the **positive integer** such that in the above equation, λ is **not a root** of $p(x)$. Suppose \mathbf{A} is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be **factorized into linear factors completely**.

Then we can write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1} (x - \lambda_2)^{r_2} \dots (x - \lambda_k)^{r_k}$$

where $r_1 + r_2 + \dots + r_k = n$, and $\lambda, \lambda_2, \dots, \lambda_k$ are the **distinct eigenvalues** of \mathbf{A} .

Then the **algebraic multiplicity** of λ_i is r_i for $i = 1, \dots, k$.

Theorem

The **eigenvalues** of a **triangular matrix** are the **diagonal entries**. The **algebraic multiplicity** of the eigenvalue is the number of times it appears as a diagonal entry of \mathbf{A} .

Definition

The **eigenspace** associated to an eigenvalue λ of \mathbf{A} is

$$E_\lambda = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\} = \text{Null}(\lambda\mathbf{I} - \mathbf{A}).$$

The **geometric multiplicity** of an eigenvalue λ is the **dimension** of its associated eigenspace,

$$\dim(E_\lambda) = \text{nullity}(\lambda\mathbf{I} - \mathbf{A})$$

Note

If \mathbf{A} and \mathbf{B} are row equivalent order n square matrices, if λ is an eigenvalue of \mathbf{A} , it is not guaranteed to be an eigenvalue of \mathbf{B} . If \mathbf{v} is an eigenvector of \mathbf{A} , it is not guaranteed to be an eigenvector of \mathbf{B} .

This is because row operations affect the determinant of the matrix, so eigenvalues and eigenvectors are not preserved.

Note

Let \mathbf{A} be a $n \times n$ matrix. The characteristic polynomial of \mathbf{A} is equal to the characteristic polynomial of \mathbf{A}^T . Hence \mathbf{A} and \mathbf{A}^T has the same eigenvalues.

Let λ be an eigenvalue of \mathbf{A} . The geometric multiplicity of λ as an eigenvalue of \mathbf{A} is equal to its geometric multiplicity as an eigenvalue of \mathbf{A}^T .

6.2 Diagonalization

Definition

A square matrix \mathbf{A} of order n is **diagonalizable** if there exists an **invertible** matrix \mathbf{P} such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

is a **diagonal** matrix, OR

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

Definition

A $n \times n$ square matrix \mathbf{A} is **diagonalizable** if and only if \mathbf{A} has n **linearly independent eigenvectors**. That is, \mathbf{A} is **diagonalizable** if and only if

$$\mathbf{P} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n), \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix},$$

where μ_i is the **eigenvalue** associated to **eigenvector** \mathbf{u}_i , $i = 1, 2, \dots, n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$, and $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a **basis** for \mathbb{R}^n .

Theorem

(Eigenspaces are linearly independent)

Let \mathbf{A} be a $n \times n$ **square** matrix. Let λ_1 and λ_2 be **distinct eigenvalues** of \mathbf{A} , $\lambda_1 \neq \lambda_2$. Suppose $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a **linearly independent** subset of eigenspace associated to eigenvalue λ_1 , and $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is a **linearly independent** subset of eigenspace associated to eigenvalue λ_2 . Then the union $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_m\}$ is **linearly independent**.

Theorem

(Geometric Multiplicity is no greater than Algebraic multiplicity)

The **geometric multiplicity** of an **eigenvalue** λ of a square matrix \mathbf{A} is **no greater** than the **algebraic multiplicity**, that is,

$$1 \leq \dim(E_\lambda) \leq r_\lambda$$

Theorem

(Equivalent Statements for Diagonalizability)

Let \mathbf{A} be a square matrix of order n . The following statements are equivalent.

1. \mathbf{A} is diagonalizable.
2. There exists a **basis** $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ of \mathbb{R}^n of **eigenvectors** of \mathbf{A} .
3. The **characteristic polynomial** of \mathbf{A} **splits into linear factors**,

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}} (x - \lambda_2)^{r_{\lambda_2}} \dots (x - \lambda_k)^{r_{\lambda_k}}$$

where r_{λ_i} is the **algebraic multiplicity** of λ_i , for $i = 1, \dots, k$, and the **eigenvalues** are **distinct**, $\lambda_i \neq \lambda_j$ for all $i \neq j$, and the **geometric multiplicity** is equal to the **algebraic multiplicity** for each eigenvalue λ_i ,

$$\dim(E_{\lambda_i}) = r_{\lambda_i}$$

Definition

A **square matrix** \mathbf{A} of order n is **not diagonalizable** if either

1. the characteristic polynomial $\det(x\mathbf{I} - \mathbf{A})$ does not split into linear factors, or
2. there exists an eigenvalue λ such that $\dim(E_\lambda) < r_\lambda$.

Algorithm to Diagonalization

Let \mathbf{A} be an order n square matrix.

1. Compute the characteristic polynomial. **If the characteristic polynomial does not split into linear factors, \mathbf{A} is not diagonalizable.** Otherwise, write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}} (x - \lambda_2)^{r_{\lambda_2}} \dots (x - \lambda_k)^{r_{\lambda_k}}$$

2. For each eigenvalue λ_i of \mathbf{A} , $i = 1, \dots, k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for

the solution space of the following linear system,

$$(\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$$

Compute first the eigenspace associated to eigenvalues with algebraic multiplicity greater than 1. If $\dim(E_{\lambda_i}) < r_{\lambda}$, \mathbf{A} is not diagonalizable.

3. Let $S = \bigcup_{i=1}^k S_{\lambda_i}$. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis for \mathbb{R}^n consisting of eigenvectors of \mathbf{A} .

4. Let $\mathbf{P} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n)$, and $\mathbf{D} = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$, where μ_i is the eigenvalue associated to \mathbf{u}_i , $i = 1, \dots, n$,

$$\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$$

Then

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n) \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix} (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n)^{-1}$$

Question

Suppose \mathbf{A} is diagonalizable. Which of the following statement(s) is/are true?

If the diagonal matrix \mathbf{D} is fixed, then the invertible matrix \mathbf{P} is fixed. (F)

If the invertible matrix \mathbf{P} is fixed, then the diagonal matrix \mathbf{D} is fixed. (T)

6.3 Orthogonally Diagonalizable

Definition

An order n square matrix \mathbf{A} is **orthogonally diagonalizable** if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some **orthogonal matrix** \mathbf{P} and **diagonal matrix** \mathbf{D} .

Theorem

(The Spectral Theorem)

Let \mathbf{A} be a $n \times n$ square matrix. \mathbf{A} is **orthogonally diagonalizable** if and only if \mathbf{A} is **symmetric**.

Theorem

(Equivalent statements for orthogonally diagonalizable)

Let \mathbf{A} be a square matrix of order n . The following statements are equivalent.

1. \mathbf{A} is orthogonally diagonalizable.
2. There exists an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ of \mathbb{R}^n of **eigenvectors** of \mathbf{A} .
3. \mathbf{A} is a **symmetric** matrix.

A is **orthogonally diagonalizable** if and only if

$$\mathbf{P} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n), \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix}$$

where μ_i is the **eigenvalue** associated to **eigenvector** \mathbf{u}_i , $i = 1, \dots, n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$, and $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an **orthonormal basis** for \mathbb{R}^n .

Theorem

(Eigenspaces of a symmetric matrix is orthogonal)

If **A** is a **symmetric** matrix, then the **eigenspaces** are **orthogonal** to each other. That is, suppose λ_1 and λ_2 are **distinct eigenvalues** of a **symmetric matrix A**, $\lambda_1 \neq \lambda_2$, and \mathbf{v}_i is an eigenvector associated to eigenvalue λ_i , for $i = 1, 2$. Then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Algorithm to Orthogonal Diagonalization

Let **A** be an order n symmetric matrix. Since **A** is symmetric, it is orthogonally diagonalizable.

1. Compute the characteristic polynomial

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}} (x - \lambda_2)^{r_{\lambda_2}} \dots (x - \lambda_k)^{r_{\lambda_k}}$$

2. For each eigenvalue λ_i of **A**, $i = 1, \dots, k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,

$$(\lambda_i\mathbf{I} - \mathbf{A})\mathbf{x} = 0.$$

3. Apply Gram-Schmidt process to each basis S_{λ_i} of the eigenspace E_{λ_i} to obtain an orthonormal basis T_{λ_i} . Let $T = \bigcup_{i=1}^k T_{\lambda_i}$. Then $T = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an orthonormal basis for \mathbb{R}^n .
4. Let $\mathbf{P} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n)$, and $\mathbf{D} = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$, where μ_i is the eigenvalue associated to \mathbf{u}_i , $i = 1, \dots, n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. Then **P** is an orthogonal matrix, and

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n) \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix}.$$

6.4 Application of Diagonalization: Markov Chain

6.5 Application of Orthogonal Diagonalization: Singular Value Decomposition

7 Linear Transformation

7.1 Introduction to Linear Transformation

7.2 Range and Kernel of Linear Transformation