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,

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

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

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

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 $Row(\mathbf{A}) = Row(\mathbf{B})$.

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 $\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n)$ and $\mathbf{B} = (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_n)$ 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 $c_1, c_2, \dots, c_n$,

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

if and only if

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

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 $\mathbf{u}$,

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

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

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

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

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

4.2 Rank

Theorem

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

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

Definition

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

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

Theorem

Rank is invariant under transpose,

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

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 $(\mathbf{A} \mid \mathbf{b})$,

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

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

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

Theorem

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

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,

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

Theorem

(Equivalent Statements of Invertibility)

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

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, $rank(\mathbf{A}) = n$.
2. The rows of $\mathbf{A}$ spans $\mathbb{R}^n$, 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, $Null(\mathbf{A}) = \{\mathbf{0}\}$.
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, $rank(\mathbf{A}) = m$.
2. The columns of $\mathbf{A}$ spans $\mathbb{R}^m$, Col($\mathbf{A}$)= $\mathbb{R}^m$.
3. The rows of $\mathbf{A}$ are linearly independent.
4. The linear system $\mathbf{Ax = b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$.
5. $\mathbf{AA}^T$ 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 $n \in \mathbb{R}^n$ 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 $\mathbf{n} \perp \mathbf{V}$.

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, 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 $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k)$;

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

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 $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k)$. Then the **orthogonal complement** of V is the nullspace of $\mathbf{A}^T$,

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

Note

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

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

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 $\mathbf{v} \in V$,

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

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

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

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

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

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

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

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

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

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

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. $rank(\mathbf{A}) = n$. 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

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

Explicitly,

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

Thus, we may write

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

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 $\mathbf{A} = \mathbf{QR}$ 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 $\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n)$ 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. $rank(\mathbf{A}) = n$. Then $\mathbf{A}^T\mathbf{A}$ is invertible, and $\mathbf{A}$ has a **left inverse**; that is, there is a $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{A}_n$

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 $\mathbf{Ax} = \mathbf{b}$ if for every vector $\mathbf{v} \in \mathbb{R}^n$,

$$\| \mathbf{Au} - \mathbf{b} \| \leq \| \mathbf{Av} - \mathbf{b} \|$$

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 $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{u}$ is the **projection** of $\mathbf{b}$ onto the column space of $\mathbf{A}$, $Col(A)$.

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 $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{u}$ is a solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.

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 $\mathbb{R}^n$. For any choice of **least square solution** $\mathbf{u}$, that is, for any solution $\mathbf{u}$ of $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$., the projection $\mathbf{Au}$ is unique.

Theorem

Let $\mathbf{V}$ 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 $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k)$. Let $\mathbf{w}$ be a vector in $\mathbb{R}^n$, and $\mathbf{u}$ be a least square solution to $\mathbf{Ax} = \mathbf{w}$. Then $\mathbf{w}_p = \mathbf{Au}$ is the **orthogonal projection** of a vector $\mathbf{w}$ onto V.

In particular, if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ 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 $\mathbf{Ax} = \mathbf{b}$ (**T**)
- $\mathbf{Au}$ **-** $\mathbf{b}$ 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 $\mathbf{v} \in \mathbb{R}^n$, $\| \mathbf{Au} - \mathbf{b} \| \leq \| \mathbf{Av} - \mathbf{b} \|$ (**T**)
- $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$ (**T**)

Question

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

Question

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

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \ldots (x - \lambda_k)^{r_{1k}}$$

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

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

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

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

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

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

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 Diagonalizaton

Definition

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

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

is a **diagonal** matrix, OR

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

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** $\mathbf{u}_i$, $i = 1, 2, \dots, n$,$\mathbf{Au}_i = \mu_i\mathbf{u}_i$, 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 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 $\dim(E_{\lambda_i}) < r_\lambda$ **, A is not diagonalizable.**

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

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

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

Then

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

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 P** and **diagonal** matrix $\mathbf{D}$.

Theorem

(The Spectral Theorem)

Let $\mathbf{A}$ be a $n \times n$ square matrix. $\mathbf{A}$ is **orthgonally 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 $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of $\mathbb{R}^n$ of **eigenvectors** of $\mathbf{A}$.
3. $\mathbf{A}$ is a **symmetric** matrix.

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

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

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**, $\lambda_2 \neq \lambda_2$, 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 $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{pmatrix}$, and $\mathbf{D} = diag(\mu_1, \mu_2, \dots, \mu_n)$, 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

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

6.4 Application of Diagonalization: Markov Chain

Theorem

Suppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. Then $\mathbf{A}^m = \mathbf{P}\mathbf{D}^m\mathbf{P}^{-1}$.

Theorem

(Powers of diagonal matrices)

Let $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$ be a diagonal matrix. Then for any positive integer m, $\mathbf{D}^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{pmatrix}$.

Corollary

(Powers of diagonalizable matrices)

Suppose $\mathbf{A}$ is **diagonalizable**. Write $\mathbf{A} = \mathbf{P} \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix} \mathbf{P}^{-1}$. Then for any positive integer $k > 0$,

$$\mathbf{A}^k = \mathbf{P} \begin{pmatrix} \mu_1^k & 0 & \cdots & 0 \\ 0 & \mu_2^k & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n^k \end{pmatrix} \mathbf{P}^{-1}$$

Moreover, if $\mathbf{A}$ is **invertible**, then the identity above holds for any integer $k \in \mathbb{Z}$.

Definition

1. A vector $\mathbf{v} = (v_i)_n$ with **nonnegative** coordinates that add up to 1, $\sum_{i=1}^n v_i = 1$, is called a **probability vector**.
2. A **stochastic matrix** is a square matrix whose columns are **probability vectors**.
3. A **Markov chain** is a sequence of **probability vectors** $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k, \dots$, together with a **stochastic matrix P** such that
$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, \dots$$

A **steady-state vector**, or **equilibrium vector** for a **stochastic matrix P** is a **probability vector** that is an **eigenvector** associated to eigenvalue 1.

Theorem

Let $\mathbf{P}$ be a $n \times n$ **stochastic matrix** and

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}$$

be a **Markov chain** for some probability vector $\mathbf{x}_0$. If the Markov chain **converges**, it will converge to an **equilibrium vector**.

Definition

(Google PageRank Algorithm)

Suppose the set S contains n sites.

We define the **adjacency matrix** for S for be the order n square matrix $\mathbf{A} = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 & \text{if site } j \text{ has an outgoing link to site } i; \\ 0 & \text{if site } j \text{ does not have an outgoing link to site } i. \end{cases}$$

From the **adjacency matrix A**, we define the **probability transition matrix P** $= (p_{ij})$ by dividing each entry of $\mathbf{A}$ by the sum of the entries in the same column; that is

$$p_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}$$

Definition

A stochastic matrix is **regular** if for some positive integer $k > 0$, the matrix power $\mathbf{P}^k$ has positive entries,

$$\mathbf{P}^k = (a_{ij})_n, \quad a_{ij} > 0 \text{ for all } i, j = 1, \dots, n.$$

Theorem

Suppose

$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, \dots$$

is a Markov chain and $\mathbf{P}$ is a **regular stochastic matrix**. Then The Markov chain **will converge** to the **unique equilibrium vector**.

Algorithm to Computing Equilibrium vector

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix.

1. Find an eigenvector $\mathbf{u}$ associate to eigenvalue $\lambda = 1$, that is, find a nontrivial solution to the homogeneous system $(\mathbf{I} - \mathbf{P})\mathbf{x} = \mathbf{0}$.

2. Write $\mathbf{u} = (u_i)$. Then

$$\mathbf{v} = \frac{1}{\sum_{k=1}^{n} u_k}\mathbf{u}$$

will be an equilibrium vector. Indeed, the i-th coordinate of $\mathbf{v}$ is $\frac{u_i}{\sum_{k=1}^{n} u_k}$ and hence, the sum of the coordinates of $\mathbf{v}$ is

$$\sum_{i=1}^{n} \frac{u_i}{\sum_{k=1}^{n} u_k} = \frac{\sum_{k=1}^{n} u_i}{\sum_{k=1}^{n} u_k} = 1$$

Alternatively, the equilibrium eigenvectors are solutions to the equation

$$\begin{pmatrix} & \mathbf{P} - \mathbf{I}_n & \\ 1 & 1 \quad \dots & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Here $\begin{pmatrix} & \mathbf{P}-\mathbf{I}_n & & \\ 1 & 1 & \dots & 1 \end{pmatrix}$ is the $(n+1) \times n$ matrix whose first n rows are the matrix $\mathbf{P}-\mathbf{I}_n$, and the last row has all entries 1.

6.5 Application of Orthogonal Diagonalization: Singular Value Decomposition

Theorem

(Singular value decomposition)

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U}$ is an order m orthogonal matrix, $\mathbf{V}$ an order n orthogonal matrix, and the matrix $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}$$

for some diagonal matrix $\mathbf{D}$ of order r, where $r \leq \min\{m, n\}$.

Algorithm to Singular Value Decomposition

Let $\mathbf{A}$ be a $m \times n$ matrix with $rank(\mathbf{A}) = r$.

1. The matrix $\mathbf{A}^T\mathbf{A}$ is a symmetric matrix, and is thus orthogonally diagonalizable. Find the eigenvalues of $\mathbf{A}^T\mathbf{A}$. Arrange the nonzero eigenvalues in descending order (counting multiplicity)

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0 = \mu_{r+1} = \cdots = \mu_n$$

2. Let $\sigma_i = \sqrt{\mu_i}, i = 1, \dots, r$,

$$\sigma_1 = \sqrt{\mu_1} \geq \sigma_2 = \sqrt{\mu_2} \geq \cdots \geq \sigma_r = \sqrt{\mu_r}$$

   These are the positive **singular values** of $\mathbf{A}$. Set

$$\mathbf{\Sigma} = \begin{pmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{matrix} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}$$

3. Proceed to find an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^T\mathbf{A}$ (section 6.3) such that $\mathbf{v}_i$ is an eigenvector associated to μ_i. Set

$$\mathbf{V} = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n).$$

4. Let $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ for $i = 1, \dots, r$. Then the set $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is an orthonormal basis for the column space of $\mathbf{A}$. If $r = m$, set $\mathbf{U} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_r)$.

   Otherwise, extend $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to an orthonormal basis for $\mathbb{R}^m$ as such. Find a basis for the solution space of

$$(\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r)^T\mathbf{x} = \mathbf{0}$$

   Perform Gram-Schmidt process on the basis found to obtain an orthonormal set $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$. This set is an orthonormal basis for the orthogonal complement of the column space of $\mathbf{A}$.

Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for $\mathbb{R}^m$. Set

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}.$$

5. Then

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_r \end{matrix} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix}$$

Note

$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \begin{pmatrix} \mu_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \mu_2 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mu_r & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \ldots & \mu_n \end{pmatrix}$$

where $\mu_i, i = 1, \ldots, n$ is the eigenvalues of $\mathbf{A}^T\mathbf{A}$; that is $\mathbf{A}^T\mathbf{A} = \mathbf{P}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{P}^T$, where $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{pmatrix}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}^T\mathbf{A}$.

Note

$\mathbf{A}\mathbf{v}_i \neq \mathbf{0}$ for all $i \leq r$ and $\mathbf{A}\mathbf{v}_i = \mathbf{0}$ for all $i > r$.

Question

$rank(\mathbf{A}) = n$ if and only if all the singular values of $\mathbf{A}$ are positive. (?)

$rank(\mathbf{A}) = m$ if and only if all the singular values of $\mathbf{A}^T$ are positive. (?)

Note

If $\mathbf{A}$ is a symmetric matrix, then the singular values of $\mathbf{A}$ are the absolute value of the eigenvalues of $\mathbf{A}$.

7 Linear Transformation

7.1 Introduction to Linear Transformation

Definition

A mapping (function) $T : \mathbb{R}^n \to \mathbb{R}^m$, is a **linear transformation** if for all vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$, and scalars α, β ,

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$$

The Euclidean space $\mathbb{R}^n$ is called the **domain** of the mapping, and the Euclidean space $\mathbb{R}^m$ is called the **codomain** of the mapping.

Equivalently, a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$, is a **linear transformation** if it satisfies the following properties.

1. For any vector $\mathbf{u}$ in $\mathbb{R}^n$ and scalar α,
$$T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$$
2. For any vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$,
$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

By induction, we have that for any vectors $\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_k$ in $\mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_k$,

$$T(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) = c_1T(\mathbf{u}_1) + c_2T(\mathbf{u}_2) + \cdots + c_kT(\mathbf{u}_k)$$

Any $m \times n$ matrix $\mathbf{A}$ defines a linear transformation $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ by multiplication,

$$T_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}\mathbf{u} \text{ for any } \mathbf{u} \in \mathbb{R}^n$$

A mapping $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is **not a linear transformation** if any of the following statements hold.

1. $\mathbf{T}$ does not map the zero vector to the zero vector, $\mathbf{T}(\mathbf{0}) \neq \mathbf{0}$.
2. There is a scalar α and a vector $\mathbf{u}$ in $\mathbb{R}^n$ such that $\mathbf{T}(\alpha\mathbf{u}) \neq \alpha\mathbf{T}(\mathbf{u})$.
3. There are vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ such that $\mathbf{T}(\mathbf{u} + \mathbf{v}) \neq \mathbf{T}(\mathbf{u}) + \mathbf{T}(\mathbf{v})$.

Theorem

(Standard matrix of linear transformation)

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** if and only if there is a **unique** $m \times n$ matrix $\mathbf{A}$ such that

$$T(\mathbf{u}) = \mathbf{A}\mathbf{u} \quad \text{for all vectors } \mathbf{u} \text{ in } \mathbb{R}^n.$$

The matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{pmatrix}$$

where $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the **standard basis** for $\mathbb{R}^n$. That is, the i-th column of $\mathbf{A}$ is $T(\mathbf{e}_1)$, for $i = 1, \ldots, n$.

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The unique $m \times n$ matrix $\mathbf{A}$ such that

$$T(\mathbf{u}) = \mathbf{A}\mathbf{u} \quad \text{for all } \mathbf{u} \text{ in } \mathbb{R}^n$$

is called the **standard matrix**, or **matrix representation** of T.

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation** and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. The **representation of T with respect to basis S**, denoted as $[T]_S$, is defined to be the $m \times n$ matrix

$$[T]_S = \begin{pmatrix} T(\mathbf{u}_1) & T(\mathbf{u}_2) & \ldots & T(\mathbf{u}_n) \end{pmatrix}$$

We are only able to find the standard matrix or the formula of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ if and only if we are given the image of T on a basis of $\mathbb{R}^n$.

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. Then for any vector $\mathbf{v}$ in $\mathbb{R}^n$,

$$T(\mathbf{v}) = [T]_S[\mathbf{v}]_S$$

that is, the image $T(\mathbf{v})$ is the product of the representation of T with respect to basis S with the coordinates $\mathbf{v}$ with respect to basis S.

Moreover, if $\mathbf{P}$ is the transition matrix from the standard basis E of $\mathbb{R}^n$ to basis S, then the standard matrix $\mathbf{A}$ of T is given by

$$\mathbf{A} = [T]_S\mathbf{P}$$

7.2 Range and Kernel of Linear Transformation

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The **range** of T is

$$\mathrm{R}(T) = T(\mathbb{R}^n) = \{\ \mathbf{v} \in \mathbb{R}^m \mid \mathbf{v} = T(\mathbf{u}) \text{ for some } \mathbf{u} \in \mathbb{R}^n\ \}$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The **range** of T is a subspace.

Let $\mathbf{A}$ be the standard matrix of T. Then the range of T is the column space of $\mathbf{A}$,

$$\mathrm{R}(T) = \{\ \mathbf{v} = T(\mathbf{u}) \mid \mathbf{u} \in \mathbb{R}^n\ \} = \{\ \mathbf{v} = \mathbf{A}\mathbf{u} \mid \mathbf{u} \in \mathbb{R}^n\ \} = \mathrm{Col}(\mathbf{A})$$

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The **rank** of T is the dimension of the range of T

$$\mathrm{rank}(T) = \dim(\mathrm{R}(T))$$

Let $\mathbf{A}$ be the standard matrix of T. Then the rank of T is equal to the rank of $\mathbf{A}$,

$$\mathrm{rank}(T) = \dim(\mathrm{R}(T)) = \dim(\mathrm{Col}(\mathbf{A})) = \mathrm{rank}(\mathbf{A})$$

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The set of all vectors in $\mathbb{R}^n$ that maps to the zero vector $\mathbf{0}$ by T is called the **kernel** of T, and is denoted as

$$\text{Ker}(T) = \{ \ \mathbf{u} \in \mathbb{R}^n \mid T(\mathbf{u}) = \mathbf{0} \ \}$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The **kernel** of T is a subspace.

Let $\mathbf{A}$ be the standard matrix of T. Then the kernel of T is the nullspace of $\mathbf{A}$,

$$\ker(T) = \{ \ \mathbf{u} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{u} = T(\mathbf{u}) = \mathbf{0} \ \} = \text{Null}(\mathbf{A}).$$

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. The **nullity** of T is the **dimension** of the kernel of T,

$$\text{nullity}(T) = \dim(\text{Ker}(T))$$

Let $\mathbf{A}$ be the standard matrix of T. Then the nullity of T is equal to the nullity of $\mathbf{A}$,

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

Definition

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **injective**, or **one-to-one** if for every vector $\mathbf{v}$ in the range of T, $\mathbf{v} \in \text{R}(T)$, there is a **unique** $\mathbf{u}$ in $\mathbb{R}^n$ such that $T(\mathbf{u}) = \mathbf{v}$.

Alternatively, T is injective if whenever $T(\mathbf{u}_1) = T(\mathbf{u}_2)$, then $\mathbf{u}_1 = \mathbf{u}_2$.

Theorem

A **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ is **injective** if and only if the kernel is **trivial**, $\ker(T) = \{\mathbf{0}\}$.

Let $\mathbf{A}$ be the standard matrix of T. Then T is injective if and only if the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

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, $rank(\mathbf{A}) = n$.
2. The rows of $\mathbf{A}$ spans $\mathbb{R}^n$, 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, $Null(\mathbf{A}) = \{\mathbf{0}\}$.
5. $\mathbf{A}^T\mathbf{A}$ is an invertible matrix of order n.

6. $\mathbf{A}$ has a left inverse.

7. The linear transformation $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$, $T_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ defined by $\mathbf{A}$ is injective.

Note

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If T is injective, then necessarily $n \leq m$.

Definition

A **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ is **surjective** or **onto** if for every $\mathbf{v}$ in the codomain $\mathbb{R}^m$, **there exists** a $\mathbf{u}$ in the domain $\mathbb{R}^n$ such that $T(\mathbf{u}) = \mathbf{v}$. Or equivalently, T is surjective if the range is the codomain, $\mathrm{R}(T) = \mathbb{R}^m$.

Let $\mathbf{A}$ be the standard matrix of T. Then T is surjective if and only if the column space of $\mathbf{A}$ is equal to $\mathbb{R}^m$. This means that the rank of $\mathbf{A}$ is equal to the number of rows.

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, $rank(\mathbf{A}) = m$.

2. The columns of $\mathbf{A}$ spans $\mathbb{R}^m$, Col($\mathbf{A}$)= $\mathbb{R}^m$.

3. The rows of $\mathbf{A}$ are linearly independent.

4. The linear system $\mathbf{Ax} = \mathbf{b}$ 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.

7. The linear transformation $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$, $T_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ defined by $\mathbf{A}$ is surjective.

Note

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If T is surjective, then necessarily $n \geq m$.

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. Then T is both **injective** and **surjective** if and only if $n = m$ and the matrixrepresentation of T is **invertible**.

Theorem

(Equivalent statements for invertibility)

15. The **linear transformation** $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$, $T_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ defined by $\mathbf{A}$ is **injective**.

15. The **linear transformation** $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$, $T_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v}$ defined by $\mathbf{A}$ is **surjective**.

Note

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is **bijective** if it is both **injective** and **surjective**. $T : \mathbb{R}^n \to \mathbb{R}^m$ is bijective if and only if there is a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{and} \quad S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n$$