# MA1522 Notes (AY24/25 Sem1)

Michael Yang

September 3, 2024

# Chapter 1: Linear Systems

## 1.1 Introduction to Linear Systems

**Definition**

A **linear equation** with $n$ variables in **standard form** has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Here $a_1, a_2, \ldots, a_n$ are known constants, called the **coefficients**, $b$ is called the **constant** and $x_1, x_2, x_n$ are variables.

The linear equation is homogeneous if $b = 0$, i.e.

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

A **system of linear equations**, or a **linear system** consists of a finite number of linear equations. In general, a linear system with $m$ variables and $n$ equations in standard form is written as

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array}\right.$$

The linear system is **homogeneous** if $b_1 = b_2 = \cdots = b_m = 0$, that is, all the linear equations are homogeneous.

Given a linear system, we say that

$$x_1 = c_1, x_2 = c_2, \ldots, x_n = c_n$$

is a **solution** to the **linear system** if the equations are simultaneously satisfied after making the substitution.

The **general solution** to a linear system captures all possible solutions to the linear system.

A linear system is said to be **inconsistent** if it does not have any solutions. It is **consistent** otherwise, that is, a linear system is consistent if it has at least one solution.

## 1.2 Solving a Linear System and Row-Echelon Form

**Definition**

(Augmented Matrix)

A linear system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array}\right.$$

can be expressed uniquely as an **augmented matrix**

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

**Definition**

In the augmented matrix, a **zero row** is a row with all entries 0. A **leading entry** of a row is the first nonzero entry of the row counting from the left.

An augmented matrix is in **row-echelon form (REF)** if

1. If **zero rows** exists, they are at the bottom of the matrix.
2. The **leading entries** are **further to the right** as we move down the rows.

An augmented matrix in **REF** has the form

$$\left(\begin{array}{cccccccc|c} * & & & & & & & & * \\ 0 & \dots & 0 & * & & & & & * \\ 0 & \dots & 0 & 0 & \dots & 0 & * & & * \\ \vdots & & & & & & & \vdots & \vdots \\ 0 & \dots & & & & & \dots & 0 & 0 \end{array}\right)$$

In the **row-echelon form**, a **pivot column** is a column containing a **leading entry**. Otherwise, it is called a **non-pivot column**.

The augmented matrix is in **reduced row-echelon form (RREF)** if further

1. The leading entries are 1.
2. In each pivot column, all entries except the leading entry is 0.

A matrix in **RREF** has the form

$$\left(\begin{array}{cccccccccc|c} 0 & \dots & 1 & & * & 0 & & * & 0 & * & * \\ 0 & \dots & 0 & \dots & 0 & 1 & & * & 0 & * & * \\ 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 1 & * & * \\ 0 & \dots & 0 & \dots & & 0 & & & 0 & 0 & 0 \\ \vdots & & & & & \vdots & & & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & & 0 & & \dots & 0 & 0 & 0 \end{array}\right)$$

**Solutions from REF and RREF**

- If the augmented matrix is in row-echelon form, we perform **back substitution** to obtain the solutions.
- If the augmented matrix is in reduced row-echelon form, we will read off the solutions directly.

**Number of solutions from row-echelon form.**

- **No solution**: a row of zero before the bar (coefficient matrix) and a non zero number after the bar.
- Unique solution: all columns of coefficient matrix are pivot columns (not possible if there is more variables than equations)
- **Infinitely many solutions**: when there is a non-pivot column in the augmented matrix before the bar

### 1.3 Elementary Row Operations

**Definition**

There are 3 types of **elementary row operations**.

1. Exchanging 2 rows, $R_i \leftrightarrow R_j$;
2. Adding a multiple of a row to another, $R_i + cR_j, c \in \mathbb{R}$ and $i \neq j$;
3. Multiplying a row by a nonzero constant, $aR_j, a \neq 0$.

Two **(augmented) matrices** are **row equivalent** if one can be obtained from the other by **elementary row operations**.

**Theorem**

Two linear systems have the **same solutions** if their **augmented matrices** are **row equivalent**.

**Definition**

(Reverse of elementary row operations)

Every elementary row operation has a reverse elementary row operation. The reverse of the row operations are given as such.

1. The reverse of exchanging 2 rows, $R_i \leftrightarrow R_j$, is itself.
2. The reverse of adding a multiple of a row to another, $R_i + cR_j$ is subtracting the multiple of that row, $R_i - cR_j$.
3. The reverse of multiplying a row by a nonzero constant, $aR_j$ is the multiplication of the reciprocal of the constant, $\frac{1}{a}R_j$.

### 1.4 Row Reduction, Gaussian and Gauss-Jordan Elimination

**Definition**

**Gaussian elimination**

**Step 1:** Locate the leftmost column that does not consist entirely of zeros.

**Step 2**: Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.

**Step 3**: For each row below the top row, add a suitable multiple of the top row to it so that the entry below the leading entry of the top row becomes zero.

**Step 4**: Now cover the top row in the augmented matrix and begin again with Step 1 applied to the submatrix that remains. Continue this way until the entire matrix is in row-echelon form.
The (augmented) matrix is now is row-echelon form. The result of step 1 to 4 reduces to (augmented) matrix to a row-echelon form. The process up to step 4 is called **Gaussian Elimination**.

**Gauss-Jordan elimination**
If we continue to perform the next 2 steps, the entire process is called **Gauss-Jordan Elimination**.

**Step 5**: Multiply a suitable constant to each row so that all the leading entries become 1.

**Step 6**: Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading entries.

The (augmented) matrix is now in reduced row-echelon form.

### 1.5 More on Linear Systems

**Summary to solving linear systems**

1. Write the linear system in its standard form.
2. Form the augmented matrix of the linear system.
3. Reduce the augmented matrix to either a row-echelon form or reduced row echelon form. May use Gaussian/Gauss-Jordan elimination.
4. Check if the system is consistent
   - If the last column is a pivot column, the system is inconsistent.
   - Otherwise, the system is consistent. If there are any non-pivot columns in the left hand side of the augmented matrix, assign the corresponding variables as parameters, $s, t$, or $s_1, s_2, \ldots, s_k$.
5. If the system is in reduced row-echelon form, read off the solutions directly.
6. If the system is in row-echelon form only, do back substitution, starting from the lowest nonzero row.
7. Write down the (general) solution to the system.

# Chapter 2: Matrix Algebra

## 2.1 Definition and Special types of Matrices

**Definition**

A (real-valued) **matrix** is a rectangular array of (real) numbers

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{m\times n} = (a_{ij})_{i=1}^{m}{}_{j=1}^{n}$$

where $a_{ij} \in \mathbb{R}$ are real numbers. The size of the matrix is said to be $m \times n$ (read as $m$ by $n$), where $m$ is the number of rows and $n$ is the number of columns.

The numbers in the array are called **entries**. The $(i, j)$-entry, $a_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n$, is the number in the $i$-th row $j$-th column.

**Definition**

(Special Types of Matrices)

**Vectors**: A $n \times 1$ matrix is called a (column) **vector**, and a $1 \times n$ matrix is called a (row) **vector**.

**Zero matrices**: All entries equal 0, denoted as $\mathbf{0}_{m\times n}$. **Not necessarily a square matrix**.

**Square matrices**: Number of rows = number of columns.

$$\mathbf{A} = (a_{ij})_n = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{m2} & \dots & a_{nn} \end{pmatrix}$$

A size $n \times n$ matrix is a square matrix of **order** $n$.

The entries $a_{ii}, i = 1, 2, \dots, n$, (explicitly, $a_{11}$, $a_{22}$, $\dots$, $a_{nn}$) are called the **diagonal entries** of the **(square) matrix**.

**Diagonal matrix**: $\mathbf{D} = (a_{ij})_n$, $a_{ij} = 0$ for $i \neq j$. Denote as $\mathbf{D} =$

$$diag(d_1, d_2, \dots, d_n) = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$$

**Scalar matrix**: $\mathbf{C} = (a_{ij})$, $a_{ij} = \begin{cases} c & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, $\mathbf{C} =$

$$diag(c, c, \dots, c) = \begin{pmatrix} c & 0 & \dots & 0 \\ 0 & c & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c \end{pmatrix}$$

**Identity matrix**: $\mathbf{I} = (a_{ij})$, $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, $\mathbf{I}_n =$

$$diag(1,1,\ldots,1) = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

A scalar matrix can also be denoted as $\mathbf{C} = c\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

**Upper triangular**: $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for $i > j$:

$$\begin{pmatrix} * & * & \ldots & * \\ 0 & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & * \end{pmatrix}$$

**Strictly upper triangular**: $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for $i \geq j$:

$$\begin{pmatrix} 0 & * & \ldots & * \\ 0 & 0 & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

**Lower triangular**: $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for $i < j$:

$$\begin{pmatrix} * & 0 & \ldots & 0 \\ * & * & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \ldots & * \end{pmatrix}$$

**Strictly lower triangular**: $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for $i \leq j$:

$$\begin{pmatrix} 0 & 0 & \ldots & 0 \\ * & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \ldots & 0 \end{pmatrix}$$

**Symmetric matrices**: $\mathbf{A} = (a_{ij})$, $a_{ij} = a_{ji}$:

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{m2} & \ldots & a_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{pmatrix}$$

## 2.2 Matrix Algebra

**Definition**

Scalar multiplication: $c\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{pmatrix}$

Matrix addition:

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

**Theorem**

(Properties of Matrix Addition and Scalar Multiplication)

For matrices $\mathbf{A} = (a_{ij})_{m\times n}, \mathbf{B} = (b_{ij})_{m\times n}, \mathbf{C} = (c_{ij})_{m\times n}$ and real numbers $a, b \in \mathbb{R}$,

1. (Commutative) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. (Associative) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
3. (Additive identity) $\mathbf{0}_{m\times n} + \mathbf{A} = \mathbf{A}$
4. (Additive inverse) $(\mathbf{A}) + (\text{-}\mathbf{A}) = \mathbf{0}_{m\times n}$
5. (Distributive law) $a(\mathbf{A}+\mathbf{B}) = a\mathbf{A} + a\mathbf{B}$
6. (Scalar addition) $(a+b)\mathbf{A} = a\mathbf{A} + b\mathbf{A}$
7. (Associative) $(ab)\mathbf{A} = a(b\mathbf{A})$
8. If $a\mathbf{A} = \mathbf{0}_{m\times n}$, then either $a = 0$ or $\mathbf{A} = \mathbf{0}$.

**Definition**

**Matrix multiplication**

$\mathbf{AB} = \mathbf{A} = (a_{ij})_{m\times p}\mathbf{B} = (b_{ij})_{p\times n} = (\sum_{k=1}^{p} a_{ik}b_{kj})_{m\times n}$

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mp} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mp} \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}b_{11}+a_{12}b_{21}+\cdots+a_{1p}b_{p1} & a_{11}b_{12}+a_{12}b_{22}+\cdots+a_{1p}b_{p2} & \dots & a_{1p}b_{1n}+a_{12}b_{2n}+\cdots+a_{1p}b_{pn} \\ a_{21}b_{11}+a_{22}b_{21}+\cdots+a_{2p}b_{p1} & a_{21}b_{22}+a_{22}b_{22}+\cdots+a_{2p}b_{p2} & \dots & a_{21}b_{1n}+a_{22}b_{2n}+\cdots+a_{2p}b_{pn} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11}+a_{m2}b_{21}+\cdots+a_{mp}b_{p1} & a_{m1}b_{12}+a_{m2}b_{22}+\cdots+a_{mp}b_{p2} & a_{m2} & a_{m1}b_{1n}+a_{m2}b_{2n}+\cdots+a_{mp}b_{pn} \end{pmatrix}$$

Caution: Matrix multiplication is not commutative $\mathbf{AB} \neq \mathbf{BA}$ in general.

**Definition**

If we multiply $\mathbf{A}$ to the left of $\mathbf{B}$, we are **pre-multiplying $\mathbf{A}$ to $\mathbf{B}$**.

If we multiply $\mathbf{A}$ to the right of $\mathbf{B}$, we are **post-multiplying $\mathbf{A}$ to $\mathbf{B}$**.

**Theorem**

(Properties of Matrix Multiplication)

1. (Associative) $(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$
2. (Left distributive law) $\mathbf{A(B+C)=AB+AC}$
3. (Right distributive law) $\mathbf{(A+B)C} = \mathbf{AC} + \mathbf{BC}$
4. (Commute with scalar multiplication) $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$
5. (Multiplicative identity) For any $m \times n$ matrix $\mathbf{A}$, $\mathbf{I}_n\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$
6. (Nonzero Zero divisor) There exists $\mathbf{A} \neq \mathbf{0}_{m\times p}$ and $\mathbf{B} \neq \mathbf{0}_{p\times n}$ such that $\mathbf{AB} = \mathbf{0}_{m\times n}$
7. (Zero matrix) For any $m \times n$ matrix $\mathbf{A}$, $\mathbf{A}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$ and $\mathbf{0}_{p\times m}\mathbf{A} = \mathbf{0}_{p\times n}$

**Definition**

Define the power of square matrices inductively as such.

1. $\mathbf{A}^0 = \mathbf{I}$
2. $\mathbf{A}^n = \mathbf{A}\mathbf{A}^{n-1}$, for $n \geq 1$.

### 2.3 Linear System and Matrix Equation

**Definition**

(Matrix Equation)

A linear system in standard form

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}\right.$$

can be expressed as a **matrix equation**

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \mathbf{A}\mathbf{x} = \mathbf{b}$$

Here $\mathbf{A} = (a_{ij})_{m\times n}$ is called the **coefficient matrix**, $\mathbf{x} = (x_i)_{n\times 1}$ the **variable vector**, and $\mathbf{b} = (b_i)_{m\times 1}$ the **constant vector**.

It can also be expressed as a **vector equation**:

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

Here $\mathbf{a}_i$ is called the **coefficient vector** for variable $x_i$, for $i = 1, \dots, n$.

**Properties of Homogeneous Linear System**

A homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ is **always consistent**, since the **zero vector** is a solution, $\mathbf{A0} = \mathbf{0}$.

**Definition**

The zero vector is called the **trivial solution**. If $x \neq 0$ is a nonzero solution to the homogeneous system, it is called a **nontrivial solution**.

**Theorem**

A **homogeneous linear system $\mathbf{Ax} = \mathbf{0}$** has **infinitely many solutions** if and only if it has a **nontrivial solution**.

**Lemma**

Let $\mathbf{v}$ be a particular solution $\mathbf{Ax} = \mathbf{b}$, and $\mathbf{u}$ be a particular solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$ with the same coefficient matrix $\mathbf{A}$. Then $\mathbf{v} + \mathbf{u}$ is also a solution to $\mathbf{Ax} = \mathbf{b}$.

**Lemma**

Suppose $\mathbf{v}_1$ and $\mathbf{v}_2$ are solutions to the linear system $\mathbf{Ax} = \mathbf{b}$. Then $\mathbf{v}_1 - \mathbf{v}_2$ is a solution to the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ with the same coefficient matrix.

**Definition**

A $p \times q$ **submatrix** of an $m \times n$ matrix $\mathbf{A}$, $p \leq m$, $q \leq n$, is formed by taking a $p \times q$ block of the entries of the matrix $\mathbf{A}$.

**Theorem**

(Block Multiplication)

Let $\mathbf{A}$ be an $m \times p$ matrix and $\mathbf{B}$ a $p \times n$ matrix. Let $\mathbf{A}_1$ be a $(m_2 - m_1 + 1) \times p$ submatrix of $\mathbf{A}$ obtained by taking rows $m_1$ to $m_2$, and $\mathbf{b}_1$ a $p \times (n_2 - n_1 + 1)$ submatrix of $\mathbf{B}$ obtained by taking columns $n_1$ to $n_2$. Then the product $\mathbf{A}_1\mathbf{B}_1$ is a $(m_2 - m_1 + 1) \times (n_2 - n_1 + 1)$ submatrix of $\mathbf{AB}$ obtained by taking rows $m_1$ to $m_2$ and columns $n_1$ to $n_2$.

In particular, let $\mathbf{b}_i$ be the $j$-th column of $\mathbf{B}$. Then

$$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_n) = (\mathbf{Ab_1} \quad \mathbf{Ab_2} \quad \ldots \quad \mathbf{Ab_n}).$$

Also, if $\mathbf{a}_i$ is the $i$-th row of B, then

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} \mathbf{B} = \begin{pmatrix} \mathbf{a}_1\mathbf{B} \\ \mathbf{a}_2\mathbf{B} \\ \vdots \\ \mathbf{a}_m\mathbf{B} \end{pmatrix}$$

### 2.4 Inverse of Matrices

**Definition**

A $n \times n$ square matrix $\mathbf{A}$ is **invertible** if there exists a **square** matrix $\mathbf{B}$ of the same size such that $\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}$.

A matrix is said to be **non-invertible** otherwise.

A non-invertible square matrix is called a **singular matrix**.

**Theorem**

(Uniqueness of inverse)

If $\mathbf{B}$ and $\mathbf{C}$ are both inverses of a square matrix $\mathbf{A}$, then $\mathbf{B} = \mathbf{C}$.

**Definition**

Since the inverse is **unique**, we can denote the **inverse** of an **invertible** matrix $\mathbf{A}$ by $\mathbf{A}^{-1}$ and call it the **inverse** of $\mathbf{A}$. That is, $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is its (unique) inverse if

$$\mathbf{AA^{-1}} = \mathbf{I}_n = \mathbf{A^{-1}A}.$$

**Theorem**

(Inverse of 2 by 2 square matrices)

A $2 \times 2$ square matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. In this case, the inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

**Theorem**

(Cancellation Law for Matrices)

Let $\mathbf{A}$ be an **invertible** matrix of order $n$.

1. (Left cancellation) If $\mathbf{B}$ and $\mathbf{C}$ are $n \times m$ matrices with $\mathbf{AB} = \mathbf{AC}$, then $\mathbf{B} = \mathbf{C}$.
2. (Right cancellation) If $\mathbf{B}$ and $\mathbf{C}$ are $m \times n$ matrices with $\mathbf{BA} = \mathbf{CA}$, then $\mathbf{B} = \mathbf{C}$.

Caution: If $\mathbf{AB} = \mathbf{CA}$, we cannot conclude that $\mathbf{B} = \mathbf{C}$.

**Theorem**

Suppose $\mathbf{A}$ is an $n \times n$ invertible square matrix. Then for any $n \times 1$ vector $\mathbf{b}$, $\mathbf{Ax} = \mathbf{b}$ has a unique solution.

**Corollary**

Suppose $\mathbf{A}$ is **invertible**. Then the **trivial solution** is the **only solution** to the homogeneous system $\mathbf{Ax} = \mathbf{0}$.

**Algorithm to compute inverse**

Suppose $\mathbf{A}$ is an invertible $n \times n$ matrix. By uniqueness of the inverse, there must be a unique solution to

$$\mathbf{AX} = \mathbf{I}.$$

By block multiplication, we are solving the augmented matrix

$$(\mathbf{A} \mid \mathbf{I}) \xrightarrow{RREF} (\mathbf{I} \mid \mathbf{A}^{-1}).$$

**Theorem**

(Properties of inverses)

Let $\mathbf{A}$ be an **invertible matrix** of order $n$.

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
2. For any **nonzero** real number $a \in \mathbb{R}$, $(a\mathbf{A})$ is **invertible** with **inverse** $(a\mathbf{A})^{-1} = \frac{1}{a}\mathbf{A}^{-1}$.
3. $\mathbf{A}^T$ is **invertible** with **inverse** $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$. (that is, the inverse of the transpose is the transpose of the inverse).
4. If $\mathbf{B}$ is an **invertible** matrix of order n, then $(\mathbf{AB})$ is **invertible** with **inverse** $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

By (4): Product of invertible matrices is invertible.

If $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$ are **invertible** matrices of the same size, then the product $\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_k$ is **invertible** with $(\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1} \dots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1}$.

**Definition**

The **negative power** of an **invertible** matrix is defined to be

$$\mathbf{A}^{-n} = (\mathbf{A}^{-1})^n$$

for any $n > 0$.

### 2.5 Elementary Matrices

**Definition**

A square matrix of order $n$ $\mathbf{E}$ is called an **elementary matrix** if it can be obtained from the identity matrix $I_n$ by performing a **single elementary row operation** where

$$\mathbf{I}_n \xrightarrow{r} \mathbf{E}$$

is an elementary row operation.

The **elementary row operation** performed to obtain $\mathbf{E}$ is said to be the **row operation** corresponding to the elementary matrix.

**Theorem**

(Elementary matrix and elementary row operation)

Let $\mathbf{A}$ be an $n \times m$ matrix and $\mathbf{E}$ be the **elementary matrix corresponding** to the **elementary row operation** $r$.

Then the product $\mathbf{EA}$ is the **resultant** of performing the row operation $r$ on $\mathbf{A}$,

$$\mathbf{A} \xrightarrow{r} \mathbf{EA}.$$

Suppose now $\mathbf{B}$ is row equivalent to $\mathbf{A}$,

$$\mathbf{A} \xrightarrow{r_1}\xrightarrow{r_2} \dots \xrightarrow{r_k} \mathbf{B}.$$

Let $\mathbf{E}_i$ be the elementary matrix corresponding to the row operation $r_i$, for $i = 1, 2, \dots, k$. Then

$$\mathbf{B} = \mathbf{E}_k \dots \mathbf{E}_2\mathbf{E}_1\mathbf{A}.$$

**Theorem**

Two $n \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$ are **row equivalent** if and only if there exists **elementary matrices** $\mathbf{E}_1, \mathbf{E}_2 \dots, \mathbf{E}_k$ such that $\mathbf{B} = \mathbf{E}_k \dots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$.

**Theorem**

(Inverse of elementary matrices)

Every elementary matrix $\mathbf{E}$ is **invertible**. The inverse $\mathbf{E}^{-1}$ is the elementary row operation corresponding to the **reverse** of the original corresponding row operation.

1. $\mathbf{I}_n \xrightarrow{R_i+cR_j} \mathbf{E} \xrightarrow{R_i-cR_j} \mathbf{I}_n \implies \mathbf{E}: R_i + cR_j, \mathbf{E}^{-1}: R_i - cR_j.$

2. $\mathbf{I}_n \xrightarrow{R_i \leftrightarrow R_j} \mathbf{E} \xrightarrow{R_i \leftrightarrow R_j} \mathbf{I}_n \implies \mathbf{E}: R_i \leftrightarrow R_j, \mathbf{E}^{-1}: R_i \leftrightarrow R_j.$

3. $\mathbf{I}_n \xrightarrow{cR_i} \mathbf{E} \xrightarrow{\frac{1}{c}R_i} \mathbf{I}_n \implies \mathbf{E}: cR_i, \mathbf{E}^{-1}: \frac{1}{c}R_i.$

### 2.6 Equivalent Statements for Invertibility

**Theorem**

If $\mathbf{A} = \mathbf{E}_k \dots \mathbf{E}_2\mathbf{E}_1$ is a product of elementary matrices, then $\mathbf{A}$ is invertible.

**Corollary**

If the reduce row-echelon form of $\mathbf{A}$ is the identity matrix, then $\mathbf{A}$ is invertible.

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if it's reduced row-echelon form is the identity matrix.

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if it is a product of elementary matrices.

**Definition**

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

1. A $m \times n$ matrix $\mathbf{B}$ is said to be a left inverse of $\mathbf{A}$ if $\mathbf{BA} = \mathbf{I}_m$, where $\mathbf{I}_m$ is the $m \times m$ identity matrix.

2. A $m \times n$ matrix $\mathbf{B}$ is said to be a right inverse of $\mathbf{A}$ if $\mathbf{AB} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

$\mathbf{B}$ is a left inverse of $\mathbf{A}$ if and only if $\mathbf{A}$ is a right inverse of $\mathbf{B}$.

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if it has a left inverse.

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if it has a right inverse.

**Theorem**

A square matrix $\mathbf{A}$ is invertible if and only if $\mathbf{Ax} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$.

**Algorithm for Finding Inverse**

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

Step 1: Form the $n \times 2n$ (augmented matrix) $(\ \mathbf{A} \mid \mathbf{I}_n)$.

Step 2: Reduce the matrix $(\ \mathbf{A} \mid \mathbf{I}_n) \to (\ \mathbf{R} \mid \mathbf{B})$ to its REF or RREF.

Step 3: If RREF $\mathbf{R} \neq \mathbf{I}$ or REF has a zero row, then $\mathbf{A}$ is not invertible. If RREF $\mathbf{R} = \mathbf{I}$ or REF has no zero row, $\mathbf{A}$ is invertible with inverse $\mathbf{A}^{-1} = \mathbf{B}$.

### 2.7 LU Factorization

**Definition**

A square matrix $\mathbf{L}$ is a **unit lower triangular** matrix if $\mathbf{L}$ is a lower triangular matrix with 1 in the diagonal entries.

An **LU factorization** of a $m \times n$ matrix $\mathbf{A}$ is the decomposition

$$\mathbf{A} = \mathbf{LU},$$

where $\mathbf{L}$ is a unit lower triangular matrix, and $\mathbf{U}$ is a row-echelon form of $\mathbf{A}$.

If such LU factorization exits for $\mathbf{A}$, we say that $\mathbf{A}$ is LU factorizable.

**Lemma**

(Product of unit lower triangular matrix is unit lower triangular)

Let $\mathbf{A}$ and $\mathbf{B}$ be unit lower triangular matrices of the same size. Then $\mathbf{AB}$ is a unit lower triangular matrix too.

**Lemma**

If $\mathbf{E}$ is an elementary matrix corresponding to the operation $R_i + cR_j$ for $i > j$ for some real number $c$, then $\mathbf{E}$ is a lower triangular matrix.

**Algorithm to LU factorization**

Suppose $\mathbf{A} \xrightarrow{r_1, r_2, \ldots, r_k} \mathbf{U}$, where each row operation $r_l$ is of the form $R_i + cR_j$ for some $i > j$ and real number $c$, and $\mathbf{U}$ is an row-echelon form of $\mathbf{A}$. Let $\mathbf{E}_i$ be the elementary matrix corresponding for $r_i$, for $r = 1, 2, \ldots, k$. Then

$$\mathbf{E}_k \ldots \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{U} \implies \mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \ldots \mathbf{E}_k^{-1}\mathbf{U} = \mathbf{LU}$$

where $\mathbf{L} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1}\dots\mathbf{E}_k^{-1}$. Then

$$\mathbf{A} = \mathbf{LU} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ * & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & 1 \end{pmatrix} \begin{pmatrix} * & & & & \dots & * \\ 0 & \dots & 0 & * & \dots & * \\ \vdots & & & & & \vdots \\ 0 & \dots & & & \dots & * \end{pmatrix}$$

**Solving Linear System using LU Factorization**

Let $\mathbf{A} = \mathbf{LU}$ be a LU-factorization. Consider the linear system $\mathbf{Ax} = \mathbf{b}$.

- Since $\mathbf{L}$ is a unit lower, can solve $\mathbf{Ly} = \mathbf{b}$ by substitution starting from top row.
- Since $\mathbf{U}$ is in row-echelon form, can solve $\mathbf{Ux} = \mathbf{y}$ by back subsitution.

**Definition**

A $n \times n$ matrix $\mathbf{P}$ is a **permutation matrix** if every rows and columns has a 1 in only one entry, and 0 everywhere else. Equivalently, $\mathbf{P}$ is a permutation matrix if and only if $\mathbf{P}$ is the product of elementary matrices corresponding to row swaps.

## 2.8 Determinant by Cofactor Expansion

**Definition**

The **determinant** of $\mathbf{A}$ is defined to be

$$\begin{aligned} \det(\mathbf{A}) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in} = \sum_{k=1}^{n} a_{ik}A_{ik} \qquad (1) \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \dots + a_{nj}A_{nj} = \sum_{k=1}^{n} a_{kj}A_{kj} \qquad (2) \end{aligned}$$

where

$$A_{ij} = (-1)^{i+j}\det(\mathbf{M}_{ij})$$

is the $(i, j)$-**cofactor** of $\mathbf{A}$, and $\mathbf{M}_{ij}$, the $(i, j)$ **matrix minor** of $\mathbf{A}$, obtained from $\mathbf{A}$ by deleting the $i$-th row and $j$-th column.

This is called the **cofactor expansion** along $\begin{cases} \text{row} & i \quad (1) \\ \text{column} & j \quad (2) \end{cases}$.

The determinant of $\mathbf{A}$ is also denoted as $\det(\mathbf{A}) = |\mathbf{A}|$.

**Theorem**

(Determinant is invariant under transpose)

The determinant of a square matrix $\mathbf{A}$ is equal to the determinant of its transpose,

$$\det(\mathbf{A}) = \det(\mathbf{A}^T).$$

**Corollary**

The **determinant** of a **triangular matrix** is the **multiplication of the diagonal entries**. That is, if $\mathbf{A} = (a_{ij})_n$ is a **triangular matrix**, then

$$\det(\mathbf{A}) = a_{11}a_{22} = \dots a_{nn} = \prod_{k=1}^{n} a_{ii}.$$

### 2.9 Determinant by Reduction

**Theorem**

Suppose $\mathbf{B}$ is obtained from $\mathbf{A}$ by a **single elementary row operation**, $\mathbf{A} \xrightarrow{r} \mathbf{B}$.
Then the **determinant** of $\mathbf{B}$ is obtained from the **determinant** of $\mathbf{A}$ as such.

- If $r = R_i + aR_j$, then $\det(\mathbf{B}) = \det(\mathbf{A})$;
- If $r = cR_i$, then $\det(\mathbf{B}) = c\,\det(\mathbf{A})$;
- If $r = R_i \leftrightarrow R_j$, then $\det(\mathbf{B}) = -\det(\mathbf{A})$.

**Corollary**

The determinant of an elementary matrix $\mathbf{E}$ is given as such.

- If $\mathbf{E}$ corresponds to $R_i + aR_j$, then $\det(\mathbf{E}) = 1$.
- If $\mathbf{E}$ corresponds to $cR_i$, then $\det(\mathbf{E}) = c$.
- If $\mathbf{E}$ corresponds to $R_i \leftrightarrow R_j$, then $\det(\mathbf{E}) = -1$.

**Theorem**

Let $\mathbf{A}$ and $\mathbf{R}$ be square matrices such that

$$\mathbf{R} = \mathbf{E}_k \dots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$$

for some elementary matrices $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_k$. Then

$$\det(\mathbf{R}) = \det(\mathbf{E}_k)\dots\det(\mathbf{E}_2)\det(\mathbf{E}_1)\det(\mathbf{A}).$$

**Corollary**

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

Suppose $\mathbf{A} \xrightarrow{r_1}\xrightarrow{r_2} \dots \xrightarrow{r_k} \mathbf{R} = \begin{pmatrix} d_1 & * & \dots & * \\ 0 & d_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$, where $\mathbf{R}$ is the reduced row-echelon form of $\mathbf{A}$. Let $\mathbf{E}_1$ be the elementary matrix corresponding to the elementary row operation $r_i$, for $i = 1, \dots, k$. Then

$$\det(\mathbf{A}) = \frac{d_1 d_2 \dots d_n}{\det(\mathbf{E}_k)\dots\det(\mathbf{E}_2)\det(\mathbf{E}_1)}.$$

### 2.10 Properties of Determinant

**Theorem**

A **square matrix $\mathbf{A}$** is **invertible** if and only if $\det(\mathbf{A}) \neq 0$.

**Theorem**

(Determinant of product is the product of determinant)

Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices of the same size. Then

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

By induction, for square matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ of the same size,

$$\det(\mathbf{A}_1\mathbf{A}_2\ldots\mathbf{A}_k) = \det(\mathbf{A}_1)\det(\mathbf{A}_2)\ldots\det(\mathbf{A}_k).$$

**Theorem**

(Determinant of inverse is the inverse of determinant)

If $\mathbf{A}$ is **invertible**, then

$$\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$$

**Theorem**

(Determinant of scalar multiplication)

For any square matrix $\mathbf{A}$ of order $n$ and scalar $c$,

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}).$$

**Definition**

Let $\mathbf{A}$ be a $n \times n$ square matrix. The **adjoint** of $\mathbf{A}$, denoted as $\mathbf{adj(A)}$, is the $n \times n$ square matrix whose $(i, j)$ entry is the $(j, i)$-cofactor of $\mathbf{A}$,

$$\mathbf{adj}(\mathbf{A}) = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}^T = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{pmatrix}$$

**Theorem**

(Adjoint formula)

Let $\mathbf{A}$ be a **square** matrix and $\mathbf{adj(A)}$ be its **adjoint**. Then

$$\mathbf{A}(\mathbf{adj}(\mathbf{A})) = \det(\mathbf{A})\mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.

**Corollary**

(Adjoint formula for inverse)

Let $\mathbf{A}$ be an **invertible** matrix. Then the **inverse** of $\mathbf{A}$ is given by

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathbf{adj}(\mathbf{A})$$

**Note**

For any square matrix $\mathbf{A}$, $\mathbf{A}^2 \neq -\mathbf{I}$ (False)
Take $\mathbf{A}$ = [0 -1; 1 0]

For any square matrix $\mathbf{A}$, $\mathbf{A} = \mathbf{0}$ iff $\mathbf{adj(A) = 0}$ (False)
Take $\mathbf{A}$ = [1 1 1; 1 1 1; 1 1 1]

# Chapter 3: Euclidean Vector Spaces

## 3.1 Euclidian Vector Spaces

**Definition**

A (real) $n$-**vector** is a collection of $n$ ordered real numbers,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, where\ v_i \in \mathbb{R}\ for\ i = 1, \dots, n.$$

The real number $v_i$ is called the $i$-th coordinate of the vector $\mathbf{v}$. The **Euclidean** $n$**-space**, denoted $\mathbb{R}^n$, is the collection of all $n$-vectors

$$\mathbb{R}^n = \left\{ v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \middle| v_i \in \mathbb{R}\ for\ i = 1, \dots, n. \right\}$$

**Properties of Vector Addition and Scalar Multiplication**

Since vectors are matrices (column vectors are $n \times 1$ matrices and row vectors are $1 \times n$ matrices), the properties of matrix addition and scalar multiplication holds for vectors. For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and scalars $a, b \in \mathbb{R}$ ,

1. The sum $\mathbf{u} + \mathbf{v}$ is a vector in $\mathbb{R}^n$
2. (Commutative) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. (Associative) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
4. (Zero vector) $\mathbf{0} + \mathbf{v} = \mathbf{v}$.
5. The negative $-\mathbf{v}$ is a vector in $\mathbb{R}^n$ such that $\mathbf{v} - \mathbf{v} = \mathbf{0}$.
6. (Scalar multiple) $a\mathbf{v}$ is a vector in $\mathbb{R}^n$.
7. (Distribution) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
8. (Distribution) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
9. (Associativity of scalar multiplication) $(ab)\mathbf{u} = a(b\mathbf{u})$.
10. If $a\mathbf{u} = \mathbf{0}$, then either $a = 0$ or $\mathbf{u} = \mathbf{0}$.

**Definition**

A **linear combination** of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \in \mathbb{R}^n$ is $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$ for some $c_1, c_2, \dots, c_k \in \mathbb{R}$.

**Definition**

A set $V$ equipped with **addition** and **scalar multiplication** is said to be a **vector space** over $\mathbb{R}$ if it satisfies the following axioms.

1. For any vectors $\mathbf{u}, \mathbf{v}$ in $V$, the sum $\mathbf{u} + \mathbf{v}$ is in $V$.

2. (Commutative) For any vectors $\mathbf{u}, \mathbf{v}$ in $V$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. (Associative) For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $V$, $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
4. (Zero vector) There is a vector $\mathbf{0}$ in $V$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all vectors $\mathbf{v}$ in $V$.
5. (Negative) For any vector $\mathbf{u}$ in $V$, there exists a vector $-\mathbf{u}$ in $V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
6. For any scalar $a$ in $\mathbb{R}$ and vector $\mathbf{v}$ in $V$, $a\mathbf{v}$ is a vector in $V$.
7. (Distribution) For any scalar $a$ in $\mathbb{R}$ and vector $\mathbf{u}, \mathbf{v}$ in $V$, $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
8. (Distribution) For any scalars $a, b$ in $\mathbb{R}$ and vector $\mathbf{u}$ in $V$, $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
9. (Associativity of scalar multiplication) For any scalars $a, b$ in $\mathbb{R}$ and vector $\mathbf{u}$ in $V$, $a(b\mathbf{u}) = (ab)\mathbf{u}$.
10. For any vector $\mathbf{u}$ in $V$, $1\mathbf{u} = \mathbf{u}$.

### 3.2 Dot Product, Norm, Distance

**Definition**

The **inner product** (or **dot product**) of vectors $\mathbf{u} = (u_i)$ and $\mathbf{v} = (v_i)$ in $\mathbb{R}^n$ is defined to be

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2, + \cdots + u_n v_n.$$

Define the **norm** of a vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u} = (u_i)$, to be the square root of the inner product of $\mathbf{u}$ with itself, and is denoted as $||\mathbf{u}||$,

$$||u|| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

This is also known as the **length** or **magnitude** of the vector.

**Theorem**

(Properties of inner product and norm)

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be vectors and $a, b, c \in \mathbb{R}$ be real numbers.

1. Inner product is **symmetric**,
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}.$$
2. Inner product **commutes** with scalar multiple,
$$c\mathbf{u} \cdot \mathbf{v} = (c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}).$$
3. Inner product is **distributive**,
$$\mathbf{u} \cdot (a\mathbf{v} + b\mathbf{w}) = a\mathbf{u} \cdot \mathbf{v} + b\mathbf{u} \cdot \mathbf{w}$$
4. Inner product is **positive definite**, $\mathbf{u} \cdot \mathbf{u} \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.
5. $||c\mathbf{u}|| = |c| \; ||\mathbf{u}||$.

**Definition**

A vector $\mathbf{u} \in \mathbb{R}^n$ is a **unit vector** if its norm is 1,

$$||\mathbf{u}|| = 1$$

**Normalizing a vector**
Let $\mathbf{u}$ be a nonzero vector $\mathbf{u} \neq \mathbf{0}$. By multiplying by the reciprocal of the norm, we get a unit vector,

$$\mathbf{u} \longrightarrow \frac{\mathbf{u}}{||\mathbf{u}||}$$

This is called **normalizing u**.

**Definition**

The **distance** between two vectors $\mathbf{u}$ and $\mathbf{v}$, denoted as $d(\mathbf{u}, \mathbf{v})$, is defined to be

$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||.$$

Define the **angle** $\theta$ between two **nonzero** vectors, $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ to be such that

$$cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \; ||\mathbf{v}||}$$

### 3.3 Linear Combinations and Linear Spans

**Definition**

A linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \in \mathbb{R}^n$ is

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k, \text{ for some } c_1, c_2, \dots c_k \in \mathbb{R}.$$

The scalars $c_1, c_2, \dots c_k$ are called **coefficients**.

Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. The **span** (or **linear span**) of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ is the subset of $\mathbb{R}^n$ containing all the linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$,

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\} = \{c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \mid c_1, c_2, \dots, c_k \in \mathbb{R}\}.$$

**Algorithm to Check for Linear Combination**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

1. Form the $n \times k$ matrix $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k)$ whose columns are the vectors in $S$.
2. Then a vector $\mathbf{v}$ in $\mathbb{R}^n$ is in $span\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ if and only if the system $\mathbf{Ax} = \mathbf{v}$ is consistent.
3. If the system is consistent, then the solutions to the system are the possible coefficients of the linear combination. That is, if $\mathbf{u} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is a solution to $\mathbf{Ax} = \mathbf{v}$, then

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

**Algorithm to Check if $span(S) = \mathbb{R}^n$**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

1. Form the $n \times k$ matrix $\mathbf{A} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ whose columns are the vectors in $S$.
2. Then $span(S) = \mathbb{R}^n$ if and only if the system $\mathbf{Ax} = \mathbf{v}$ is consistent for all $\mathbf{v}$.
3. This is equivalent to the reduced row-echelon form of $\mathbf{A}$ having no zero rows.

**Properties of linear span**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a finite set of vector. The span of $S$, $span(S)$ has the following properties.

1. The span of $S$ **contains the origin**,
$$\mathbf{0} \in span(S).$$
2. The span of $S$ is **closed under vector addition**, for any $\mathbf{u}, \mathbf{v} \in span(S)$,
$$\mathbf{u} + \mathbf{v} \in span(S)$$
3. The span $S$ is **closed under scalar multiplication**, for any $\mathbf{u} \in span(S)$ and real number $\alpha \in \mathbb{R}$,
$$\alpha\mathbf{u} \in span(S).$$

   Properties (ii) and (iii) can be combined together into one property (ii'): The span is **closed under linear combinations**, that is, if $\mathbf{u}, \mathbf{v}$ are vectors in $span(S)$ and $\alpha, \beta$ are any scalars, then the linear combination $\alpha\mathbf{u} + \beta\mathbf{v}$ is a vector in $span(S)$.

**Theorem**

(Linear span is closed under linear combinations)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. For any vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ in $span(S)$, the span of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ is a subset of $span(S)$,
$$span\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \subseteq span(S).$$

**Algorithm to check for Set Relations between Spans**

Suppose we are given 2 sets of vectors $T = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$.

1. By the corollary, if $\mathbf{v}_i \in span(S)$ for $i = 1, \dots, m$, we can conclude that $span(T) \subseteq span(S)$.
2. Recall that to check if $\mathbf{v}_i \in span(S)$, we check that the system $\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_k & \mathbf{v}_i \end{array}\right)$ is consistent for all $i = 1, \dots, m$.
3. There are in total $m$ such linear systems to check. However, since they have the same coefficient matrix, we may combine and check them together, that is, check that
$$\left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_m \end{array}\right)$$
   is consistent.

**Theorem**

(Algorithm to check for set relations between spans)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be sets of vectors in $\mathbb{R}^n$. Then $span(T) \subseteq span(S)$ if and only if $\left(\ \mathbf{u}_1\ \mathbf{u}_2\ \dots\ \mathbf{u}_k \mid \mathbf{v}_1\ \mid \mathbf{v}_2 \mid\ \dots\ \mid \mathbf{v}_m\ \right)$ is consistent.

### 3.4 Subspaces

**Definition**

The set of solutions to a linear system $\mathbf{Ax} = \mathbf{b}$ can be expressed **implicitly** as

$$V = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{Au} = \mathbf{b}\}$$

or **explicitly** as

$$V = \{\mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k \mid s_1, s_2, \dots, s_k \in \mathbb{R}\},$$

where $\mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k,\ s_1, s_2, \dots, s_k \in \mathbb{R}$ is the general solution.

**Definition**

A subset $V$ of $\mathbb{R}^n$ is a **subspace** if it satisfies the following properties.

1. $V$ **contains the zero vector**, $\mathbf{0} \in V$.
2. $V$ is **closed under scalar multiplication**. For any vector $v$ in $V$ and scalar $\alpha$, the vector $\alpha\mathbf{v}$ is in $V$.
3. $V$ is **closed under addition**. For any vectors **u, v** in $V$, the sum $\mathbf{u} + \mathbf{v}$ is in $V$.

Property (i) can be replaced with property (i'): $V$ is **nonempty**.

Properties (ii) and (iii) is equivalent to property (ii'): $V$ is **closed under linear combination**. For any **u, v** in $V$, and scalars $\alpha, \beta$, the linear combination $\alpha\mathbf{u} + \beta\mathbf{v}$ is in $V$.

**Theorem**

(Solution set of a homogeneous system is a subspace)

The solution set $V = \{\mathbf{u} \mid \mathbf{Au} = \mathbf{b}\}$ to a linear system $\mathbf{Ax} = \mathbf{b}$ is a **subspace** if and only if $\mathbf{b} = \mathbf{0}$, that is, the system is **homogeneous**.

**Definition**

The **solution set** to a **homogeneous system** is call a **solution space.**

**Theorem**

(Subspaces are equivalent to linear spans)

A subset $V \subseteq \mathbb{R}^n$ is a subspace if and only if it is a linear span, $V = span(S)$, for some finite set $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$.

**Check if a set is a subspace**

To show that a set $V$ is a subspace, we can either

- find a spanning set, that is, find a set $S$ such that $V = span(S)$, or
- show that $V$ satisfies the 3 conditions of being a subspace.

To show that a subset $V$ is not a subspace, we can either

- show that it does not contain the zero vector, $\mathbf{0} \notin V$,
- find a vector $\mathbf{v} \in V$ and a scalar $\alpha \in \mathbb{R}$ such that $\alpha\mathbf{v} \notin V$, or
- find vectors $\mathbf{u}, \mathbf{v} \in V$ such that the sum is not in $V$, $\mathbf{u} + \mathbf{v} \notin V$.

**Theorem**

(Affine spaces)

The solution set $W = \{ \mathbf{w} \mid \mathbf{Aw} = \mathbf{b} \}$ of a non-homogeneous linear system $\mathbf{Ax} = \mathbf{b}$, $\mathbf{b} \neq 0$, is given by

$$\mathbf{u} + V := \{ \mathbf{u} + \mathbf{v} \mid \mathbf{v} \in V \}$$

where $V = \{ \mathbf{v} \mid \mathbf{Av} = \mathbf{0} \}$ is the solution space to the associated homogeneous system and $\mathbf{u}$ is a particular solution, $\mathbf{Au} = \mathbf{b}$.

That is, vectors in $\mathbf{u} + V$ are of the form $\mathbf{u} + \mathbf{v}$ for some $\mathbf{v}$ in $V$.

### 3.5 Linear Independence

**Definition**

A set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is **linearly independent** if the **only coefficients** $c_1, c_2, \ldots, c_k \in \mathbb{R}$ satisfying the equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$$

are $c_1 = c_2 = \cdots = c_k = 0$. Otherwise, we say that the set is **linearly dependent**.

**Algorithm to Check for Linear Independence**

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

- $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is **linearly independent** if and only if the **homogeneous system** $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \end{pmatrix}\mathbf{x} = \mathbf{0}$ has only the **trivial solution**.
- The homogeneous system has only the **trivial solution** if and only if the **reduced row-echelon form** of $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \end{pmatrix}$ has **no non-pivot column**.

**Theorem**

(Solution set of a homogeneous system is a subspace)

A subset $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of $\mathbb{R}^n$ is **linearly independent** if and only if the **reduced row-echelon form** of $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \end{pmatrix}$ has **no non-pivot columns**.

### 3.6 Basis and Coordinates

**Definition**

Let $V$ be a subspace of $\mathbb{R}^n$. A set $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a **basis** for $V$ if

- $span(S) = V$ and
- $S$ is linearly independent.

**Theorem**

Suppose $S$ $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a basis for $V$. Then every vector $\mathbf{v}$ in the subspace $V$ **can be written** as a linear combination of vectors in $S$ **uniquely**.

**Theorem**

(Basis for Solution Set of Homogeneous System)

Let $V = \{\mathbf{u} | \mathbf{A}\mathbf{u} = \mathbf{0}\}$ be the solution space to some homogeneous system. Suppose

$$s_1\mathbf{u}_1 + s_1\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k, s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is a general solution to the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$.
Then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a basis for the subspace $V = \{\mathbf{u} | \mathbf{A}\mathbf{u} = \mathbf{0}\}$.

**Theorem**

Basis for the zero space $\{\mathbf{0}\}$ of $\mathbb{R}^n$ is the **empty set** $\{\}$ or $\emptyset$.

**Theorem**

A $n \times n$ square matrix $\mathbf{A}$ is invertible if and only if the columns are linearly independent.

**Theorem**

A $n \times n$ square matrix $\mathbf{A}$ is invertible if and only if the columns spans $\mathbb{R}^n$.

**Corollary**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a subset of $\mathbb{R}^n$ containing $n$ vectors. Then $S$ is linearly independent if and only if $S$ spans $\mathbb{R}^n$.

**Corollary**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a subset of $\mathbb{R}^n$ and $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \end{pmatrix}$ be the matrix whose columns are vectors in $S$. Then $S$ is a **basis** for $\mathbb{R}^n$ if and only if $k = n$ and $\mathbf{A}$ is an **invertible matrix**.

**Theorem**

A $n \times n$ **square** matrix $\mathbf{A}$ is invertible if and only if the **columns** of $\mathbf{A}$ form a **basis** for $\mathbb{R}^n$.

**Theorem**

A $n \times n$ **square** matrix $\mathbf{A}$ is invertible if and only if the **rows** of $\mathbf{A}$ form a **basis** for $\mathbb{R}^n$.

**Theorem**

(Equivalent Statements for Invertibility)

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

1. $\mathbf{A}$ is **invertible**.
2. $\mathbf{A}^T$ is **invertible**.
3. $\mathbf{A}$ has a **left-inverse**, that is, there is a matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}$.
4. $\mathbf{A}$ has a **right-inverse**, that is, there is a matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{I}$.
5. The **reduced row-echelon form** of $\mathbf{A}$ is the **identity matrix**.
6. $\mathbf{A}$ can be expressed as a **product** of **elementary matrices**.
7. The **homogeneous system** $\mathbf{Ax} = \mathbf{0}$ has **only the trivial solution**.
8. For **any** $\mathbf{b}$, the system $\mathbf{Ax} = \mathbf{b}$ is **consistent**.
9. The **determinant** of $\mathbf{A}$ is **nonzero**, $\det(\mathbf{A}) \neq 0$.
10. The **columns/rows** of $\mathbf{A}$ are **linearly independent** for $\mathbb{R}^n$.
11. The **columns/rows** of $\mathbf{A}$ **spans** $\mathbb{R}^n$.

**Definition**

(Coordinates Relative to a Basis)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a basis for a subspace $V$ of $\mathbb{R}^n$.
Then given any vector $\mathbf{v} \in V$, we can write $\mathbf{v}$ unique as

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

The coordinates of $\mathbf{v}$ relative to the basis $S$ is defined to be the vector

$$[\mathbf{v}]_s = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

**Algorithm for Computing Relative Coordinate**

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u_k}\}$ be a basis for a subspace $V$ of $\mathbb{R}^n$.

For $\mathbf{v} \in V$, find real numbers $c_1, c_2, \dots, c_k \in \mathbb{R}$ such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{v}.$$

That is, we are solving for

$$(\mathbf{u}_1\ \mathbf{u}_2\ \dots\ \mathbf{u}_k \mid \mathbf{v}).$$

**Theorem**

Let $V$ be a subspace of $\mathbb{R}^n$ and $B$ a basis for $V$.

1. For any vectors $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} = \mathbf{v}$ if and only if $[\mathbf{u}]_B = [\mathbf{v}]_B$.

2. For any $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in V$,

$$[c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m]_B = c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + \cdots + c_m[\mathbf{v}_m]_B.$$

### 3.7 Dimensions

**Theorem**

Let $V$ be a subspace of $\mathbb{R}^n$ and $B$ a basis for $V$. Suppose $B$ contains $k$ vectors, $|B| = k$. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ be vectors in $V$. Then

1. $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ is linearly independent (respectively, dependent) if and only if $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, \dots, [\mathbf{v}_m]_B$ is linearly independent (respectively, dependent) in $\mathbb{R}^k$; and

2. $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ spans $V$ if and only if $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, \dots, [\mathbf{v}_m]_B$ spans $\mathbb{R}^k$.

**Corollary**

Let $V$ be a subspace of $\mathbb{R}^n$ and $V$ a basis for $B$. Suppose $B$ contains $k$ vectors, $|B| = k$.

1. If $S = \{\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_m\}$ is a subset of $V$ with $m > k$, then $S$ is **linearly dependent**.

2. If $S = \{\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_m\}$ is a subset of $V$ with $m < k$, then $S$ **cannot span** $V$.

**Corollary**

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ are bases for a subspace $V \subseteq \mathbb{R}^n$. Then $k = m$.

**Definition**

Let $V$ be a subspace of $\mathbb{R}^n$. The **dimension** of $V$, denoted by $\dim(V)$, is defined to be the **number of vectors** in any **basis** of $V$.

**Theorem**

(Dimension of solution space)

Let $\mathbf{A}$ be a $m \times n$ matrix. The **number of non-pivot columns** in the reduced row-echelon form of $A$ is the **dimension** of the solution space

$$V = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{u} = \mathbf{0}\}.$$

**Theorem**

(Spanning Set Theorem)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a subset of vectors in $\mathbb{R}^n$, and let $V = span(S)$. Suppose $V$ is not the zero space, $V \neq \{\mathbf{0}\}$. Then there must be a subset of $S$ that is a basis for $V$.

**Theorem**

(Linear Independence Theorem)

Let $V$ be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ a linearly independent subset of $V$, $S \subseteq V$. Then there must be a set $T$ containing $S$, $S \subseteq T$ such that $T$ is a basis for $V$.

**Theorem**

Let $U$ and $V$ be **subspaces** of $\mathbb{R}^n$.

1. If $U \subseteq V$, then $dim(U) \leq dim(V)$.
2. If $U \subseteq V$, and $U \neq V$, then $dim(U) < dim(V)$ That is, $U \subseteq V$, then $dim(U) \leq dim(V)$ with **equality** if and only if $U = V$.

**Theorem**

(B1)
Let $V$ be a $k$-dimensional subspace of $\mathbb{R}^n$, $dim(V) = k$. Suppose $S \subseteq V$ is a **linearly independent** subset containing $k$ vectors, $|S| = k$. Then $S$ is a **basis** for $V$.

In summary,

1. $|S| = dim(V)$
2. $S \subseteq V$
3. $S$ is linearly independent

**Theorem**

(B2)
Let $V$ be a $k$ dimensional subspace of $\mathbb{R}^n$, $dim(V) = k$. Suppose $S$ is a set containing $k$ vectors, $|S| = k$, such that $V \subseteq span(S)$. Then $S$ is a **basis** for $V$.

In summary,

1. $|S| = dim(V)$
2. $V \subseteq span(S)$

### 3.8 Transition Matrices

**Definition**

Let $V$ be a subspace of $\mathbb{R}^n$. Suppose $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are **basis** for the subspace $V$. Define the **transition matrix** from $T$ **to** $S$ to be

$$\mathbf{P} = \begin{pmatrix} [\mathbf{v}_1]_S & [\mathbf{v}_2]_S & \ldots & [\mathbf{v}_k]_S \end{pmatrix},$$

the matrix whose columns are the coordinates of the vectors in $T$ relative to the basis $S$.

**Theorem**

(Transition Matrix)

Let $V$ be a subspace of $\mathbb{R}^n$. Suppose $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are **bases** for the subspace $V$. Let $\mathbf{P}$ be the transition matrix from $T$ **to** $S$. Then for any vector $w$ in $V$,

$$[\mathbf{w}]_S = \mathbf{P}[\mathbf{w}]_T.$$

**Algorithm to find Transition Matrix**

Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be basis for a subspace $V$ in $\mathbb{R}^n$. To find $\mathbf{P}$, the transition matrix from $T$ to $S$,

$$(\text{“}S''|\text{“}T'') = \left(\begin{array}{cccc|cccc} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \ldots \mathbf{v}_k \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{c|c} \mathbf{I}_k & \mathbf{P} \\ \mathbf{0}_{(n-k)\times k} & \mathbf{0}_{(n-k)\times k} \end{array}\right)$$

**Theorem**

(Inverse of Transition Matrix)

Suppose $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are **bases** for a subspace $V$ of $\mathbb{R}^n$. Let $\mathbf{P}$ be the **transition matrix from** $T$ **to** $S$. Then $\mathbf{P}^{-1}$ is the **transition matrix from** $S$ **to** $T$.

# 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.

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

**Definition**

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

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

**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 $\mathbf{Ax} = \mathbf{b}$ 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

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

**Note**

If $\mathbf{A}$ and $\mathbf{B}$ are order $n$ square matrices and $\mathbf{AB} = \mathbf{0}$, then rank($\mathbf{A}$) + rank($\mathbf{B}$) $\leq n$.

$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_n) = \mathbf{0} \Rightarrow \text{Col}(\mathbf{B}) \in \text{Null}(\mathbf{A}) \Rightarrow \text{nullity}(\mathbf{A}) \geq \text{rank}(\mathbf{B}) \Rightarrow n - \text{rank}(\mathbf{A}) \geq \text{rank}(\mathbf{B}) \Rightarrow \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \leq n$

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

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

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 $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ 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 $i = 1, \ldots, k$.

**Theorem**

Let $V$ be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ 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 $\mathbf{A} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \ldots \quad \mathbf{u}_k)$;

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

**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 $\mathbf{v} \in V$,

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

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

**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 $[\mathbf{v}]_S$, $\parallel [\mathbf{v}]_S \parallel$?

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

**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.

**Note**

Let $\mathbf{A}$ be an orthogonal matrix of order $n$ and **u, v** be any two vectors in $\mathbb{R}^n$. Then

1. $\parallel \mathbf{u} \parallel = \parallel \mathbf{Au} \parallel$
Proof: $(\mathbf{Au}) \cdot (\mathbf{Av}) = \mathbf{u}^T\mathbf{A}^T\mathbf{Av} = \mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$. Hence $\parallel \mathbf{Au} \parallel = \sqrt{(\mathbf{Au}) \cdot (\mathbf{Au})} = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \parallel \mathbf{u} \parallel$

2. $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{Au}, \mathbf{Av})$
Proof: $d(\mathbf{Au}, \mathbf{Av}) = \parallel \mathbf{Au} - \mathbf{Av} \parallel = \parallel \mathbf{A}(\mathbf{u} - \mathbf{v}) \parallel = \parallel \mathbf{u} - \mathbf{v} \parallel = d(\mathbf{u}, \mathbf{v})$

3. the angle between $\mathbf{u}$ and $\mathbf{v}$ is equal to the angle between $\mathbf{Au}$ and $\mathbf{Av}$
Proof: $\frac{\mathbf{Au} \cdot \mathbf{Av}}{||\mathbf{Au}||\,||\mathbf{Av}||} = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||\,||\mathbf{v}||}$. Taking cosine on both sides will give the same angle.

**Note**

Let $\mathbf{A}$ be an orthogonal matrix of order $n$. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and define $T = \{\mathbf{Au}_1, \mathbf{Au}_2, \dots, \mathbf{Au}_k\}$.

If $S$ is orthogonal, then $T$ is orthogonal.
Proof: If $\mathbf{A}$ is an orthogonal matrix, then $\mathbf{Au} \cdot \mathbf{Av} = \mathbf{u}^T\mathbf{A}^T\mathbf{Av} = \mathbf{u} \cdot \mathbf{v}$. So for any two vectors in $T$, $\mathbf{Au}_i \cdot \mathbf{Au}_j = \mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$. So $T$ is an orthogonal set.

If $S$ is orthonormal, then $T$ is orthonormal.
Proof: $\mathbf{Au}_i \cdot \mathbf{Au}_j = \mathbf{u}_i \cdot \mathbf{u}_j$. So if $i = j$, $\mathbf{Au}_i \cdot \mathbf{Au}_j = \mathbf{u}_i \cdot \mathbf{u}_j = 1$

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

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

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

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

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

## 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

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

Since the set $S = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ 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

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

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 $i = 1, 2, \ldots, n$.

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 $\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_n)$ to obtain an **orthonormal set** $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$.
2. Set $\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$.
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 $\mathbf{BA} = \mathbf{I}_n$

## 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 u**, that is, for any solution $\mathbf{u}$ of $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$., 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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ 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 $\mathbf{Ax} = \mathbf{b}$ (**T**)
- **Au - 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 $\lambda$ 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 $\lambda$. 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 $n$. $\lambda$ is an **eigenvalue** of $\mathbf{A}$ if and only if $\lambda$ 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 $\lambda$ be an eigenvalue of $\mathbf{A}$. The **algebraic multiplicity** of $\lambda$ is the **largest integer** $r_\lambda$ such that

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

for some polynomial $p(x)$. Alternatively, $r_\lambda$ is the **positive integer** such that in the above equation, $\lambda$ 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

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

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

Then the **algebraic multiplicity** of $\lambda_i$ 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 $\lambda$ of $\mathbf{A}$ is

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

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

**Note**

Let $\mathbf{A}$ be a matrix such that $\mathbf{A}^T = -\mathbf{A}$.

(a) If $\lambda$ is an eigenvalue of $\mathbf{A}$, then $\lambda = 0$.

$\mathbf{A}^T\mathbf{v} = -\mathbf{A}\mathbf{v} \Rightarrow$ (Take transpose of both sides) $\mathbf{v}^T\mathbf{A} = -\mathbf{v}^T\mathbf{A}^T \Rightarrow$ (Post multiply both sides by $\mathbf{v}$) $\mathbf{v}^T\mathbf{A}\mathbf{v} = -\mathbf{v}^T\mathbf{A}^T\mathbf{v} \Rightarrow \mathbf{v}^T(\lambda\mathbf{v}) = -\mathbf{v}^T(\lambda\mathbf{v}) \Rightarrow \lambda \parallel \mathbf{v} \parallel^2 = -\lambda \parallel \mathbf{v} \parallel^2 \Rightarrow \lambda = 0$

(b) $\mathbf{A}$ is diagonalizable iff $\mathbf{A} = \mathbf{0}$.

($\Rightarrow$) If $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \mathbf{P}\mathbf{0}\mathbf{P}^{-1} = \mathbf{0}$ since eigenvalues of $\mathbf{A}$ can only be 0.
($\Leftarrow$) If $\mathbf{A} = \mathbf{0}$, then $\mathbf{A} = \mathbf{I}\mathbf{0}\mathbf{I}^{-1}$, so $\mathbf{A}$ is diagonalizable.

**Note**

$\lambda$ is an eigenvalue of $\mathbf{A}$ iff it is an eigenvalue of $\mathbf{A}^T$.
Proof: $\det(x\mathbf{I} - \mathbf{A}) = \det((x\mathbf{I} - \mathbf{A})^T) = \det((x\mathbf{I})^T - \mathbf{A}^T) = \det(x\mathbf{I} - \mathbf{A}^T)$

If $\mathbf{A}$ is diagonalizable, then $\mathbf{A}^T$ is diagonalizable.
Proof: $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \Rightarrow \mathbf{A}^T = (\mathbf{P}\mathbf{D}\mathbf{P}^{-1})^T = (\mathbf{P}^{-1})^T\mathbf{D}^T\mathbf{P}^T = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1}$

If $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ associated to eigenvalue $\lambda$, then $\mathbf{v}$ is an eigenvector of $\mathbf{A}^k$ associated to eigenvalue

$\lambda^k$ for any $k \in \mathbb{Z}^+$.
Proof: $\mathbf{A}^k\mathbf{v} = \mathbf{A}^{k-1}\mathbf{A}\mathbf{v} = \lambda\mathbf{A}^{k-2}\mathbf{A}\mathbf{v} = \cdots = \lambda^{k-1}\mathbf{A}\mathbf{v} = \lambda^k\mathbf{v}$. Since $\mathbf{v} \neq \mathbf{0}$, $\lambda^k$ is an eigenvalue of $\mathbf{A}^k$.

If $\mathbf{A}$ is invertible, $\mathbf{v}$ is an eigenvector of $\mathbf{A}^k$ associated to eigenvalue $\lambda^k$ for any negative integer $k$.
Proof: $\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \Rightarrow$ (premulitply by inverse) $\mathbf{A}^{-1}\mathbf{A}\mathbf{v} = \mathbf{A}^{-1}\lambda\mathbf{v} \Rightarrow \mathbf{v} = \lambda\mathbf{A}^{-1}\mathbf{v} \Rightarrow \frac{1}{\lambda}\mathbf{v} = \mathbf{A}^{-1}\mathbf{v}$, so $\lambda^{-1}$ is an eigenvalue of $\mathbf{A}^{-1}$. Then for any negative $k$, let $k = -m$, $\mathbf{A}^k\mathbf{v} = (\mathbf{A}^{-1})^m\mathbf{v} = \lambda^{-m}\mathbf{v}$

A square matrix is *nilpontent* iff there is a positive integer $k$ such that $\mathbf{A}^k = \mathbf{0}$. If $\mathbf{A}$ is nilpotent, then 0 is the only eigenvalue.
Proof: If $\mathbf{A}^k = \mathbf{0}$ then $\mathbf{A}^k\mathbf{v} = \mathbf{0}\mathbf{v} = \mathbf{0}$. Since $\mathbf{v}$ is a non-zero vector, the only eigenvalue is 0.

Let $\mathbf{A}_{n\times n}$ have only one eigenvalue $\lambda$ with algebraic multiplicity $n$. Then $\mathbf{A}$ is diagonalizable iff $\mathbf{A}$ is a scalar matrix, $\mathbf{A} = \lambda\mathbf{I}$.
Proof: If $\mathbf{A}$ is diagonalizable, then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, and the entries of $\mathbf{D}$ are all $\lambda$, i.e. $\mathbf{D} = \lambda\mathbf{I}$. Hence $\mathbf{A} = \mathbf{P}\lambda\mathbf{I}\mathbf{P}^{-1} = \lambda\mathbf{P}\mathbf{I}\mathbf{P}^{-1} = \lambda\mathbf{I}$.

The only diagonalizable nilpotent matrix is the zero matrix.
Proof: Let $\mathbf{A}$ be a nilpotent matrix. Then 0 is the only eigenvalue. If $\mathbf{A}$ is diagonalizable, then $\mathbf{A} = \mathbf{P}\mathrm{diag}(0, 0, \ldots, 0)\mathbf{P}^{-1} = \mathbf{0}$.

**Note**

If $\mathbf{A}$ and $\mathbf{B}$ are similar matrices, i.e. $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$ for some invertible $\mathbf{P}$, then the characteristic polynomials of $\mathbf{A}$ and $\mathbf{B}$ are equal.

Proof: $\det(x\mathbf{I} - \mathbf{A}) = \det(\mathbf{P})\det(\mathbf{P})^{-1}\det(x\mathbf{I} - \mathbf{A}) = \det(\mathbf{P})\det(x\mathbf{I} - \mathbf{A})\det(\mathbf{P})^{-1} = \det(\mathbf{P}(x\mathbf{I} - \mathbf{A})\mathbf{P}^{-1}) = \det(x\mathbf{P}\mathbf{I}\mathbf{P}^{-1} - \mathbf{P}\mathbf{A}\mathbf{P}^{-1}) = \det(x\mathbf{I} - \mathbf{B})$

**Note**

If the characteristic polynomial of $\mathbf{A}$ and $\mathbf{B}$ are equal, we cannot conclude that $\mathbf{A}$ and $\mathbf{B}$ are similar matrices.

E.g. [1 1; 0 1] and [1 0; 0 1] have the same characteristic polynomial $(x-1)^2$ but for any invertible matrix $\mathbf{P}$, $\mathbf{P}\mathbf{I}\mathbf{P}^{-1} = \mathbf{I} \neq$[1 1; 0 1]

**Note**

If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices with the same determinant, it is not true that their characteristic polynomials are equal.
E.g. A=[1 0; 0 0] and B=[2 0; 0 0] have the same determinant 0, but their characteristic polynomials are different.

**Note**

If $\mathbf{A}$ is a $2 \times 2$ matrix, its characteristic polynomial is $x^2 - tr(\mathbf{A}) + \det(\mathbf{A})$ where $tr(\mathbf{A})$ is the sum of diagonal entries of $\mathbf{A}$.

Proof: Write $\mathbf{A}$=[a b; c d]. Then $tr(\mathbf{A}) = a + d$ and $\det(\mathbf{A}) = ad - bc$. $\det(x\mathbf{I} - \mathbf{A}) = (x - a)(x - d) - bc = x^2 - (a + d)x + ad - bc = x^2 - tr(\mathbf{A})x + \det(\mathbf{A})$

If $\mathbf{A}$ is an $n \times n$ matrix and $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, then $a_0 = p(0) = \det(0\mathbf{I} - \mathbf{A}) = \det(-\mathbf{A}) = (-1)^n \det(\mathbf{A})$

**Note**

If $\mathbf{AB} = \mathbf{BA}$, and $\mathbf{v}$ is an eigenvector of $\mathbf{B}$, then $\mathbf{Av}$ is also an eigenvector of $\mathbf{B}$.

Proof: $\mathbf{Bv} = \lambda\mathbf{v} \Rightarrow \mathbf{ABv} = \mathbf{B}(\mathbf{Av}) = \lambda(\mathbf{Av})$

### 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 $\mu_i$ 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 $\lambda_1$ and $\lambda_2$ 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 $\lambda_1$, and $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is a **linearly independent** subset of of eigenspace associated to eigenvalue $\lambda_2$. 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** $\lambda$ of a square matrix $\mathbf{A}$ is **no greater** than the **alge-**

**braic 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 $r_{\lambda_i}$ is the **algebraic multiplicity** of $\lambda_i$, 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 $\lambda_i$,

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

**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 $\lambda$ 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, 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 $\lambda_i$ of $\mathbf{A}$, $i = 1, \dots, k$, find a basis $S_{\lambda_i}$ for the eigenspace, that is, find a basis $S_{\lambda_i}$ 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 $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ 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 $\mathbf{D} = diag(\mu_1, \mu_2, \dots, \mu_n)$, where $\mu_i$ is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, \dots, n$,

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

Then

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

**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**)

**Note**

If $\mathbf{A}$ is diagonalizable, then there exists a $\mathbf{B}$ such that $\mathbf{B}^3 = \mathbf{A}$.

Proof: $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. Let $\mathbf{Q}$ be the diagonal matrix $\mathbf{D}$ but with all the entries cube rooted. Then $\mathbf{A} = \mathbf{P}\mathbf{Q}^3\mathbf{P}^{-1} = \mathbf{B}^3$, $\mathbf{B} = \mathbf{P}\mathbf{Q}\mathbf{P}^{-1}$

### 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 **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 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ 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 $\mu_i$ 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 $\lambda_1$ and $\lambda_2$ are **distinct eigenvalues** of a **symmetric matrix A**, $\lambda_2 \neq \lambda_2$, and $\mathbf{v}_i$ is an eigenvector associated to eigenvalue $\lambda_i$, 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 $\lambda_i$ of **A**, $i = 1, \dots, k$, find a basis $S_{\lambda_i}$ for the eigenspace, that is, find a basis $S_{\lambda_i}$ 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 $S_{\lambda_i}$ of the eigenspace $E_{\lambda_i}$ to obtain an orthonormal basis $T_{\lambda_i}$. 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 $\mathbf{D} = diag(\mu_1, \mu_2, \dots, \mu_n)$, where $\mu_i$ 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

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

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

**Corollary**

(Powers of diagonalizable matrices)

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

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

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** $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$, together with a **stochastic matrix P** such that
$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \ldots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, \ldots$$

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

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

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 **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,

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

**Theorem**

Suppose

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

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

**Algorithm to Computing Equilibrium vector**

Let **P** be a $n \times n$ stochastic matrix.

1. Find an eigenvector **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 **v** is $\frac{u_i}{\sum_{k=1}^{n} u_k}$ and hence, the sum of the coordinates of **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

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

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)

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

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

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

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

$$\mathbf{\Sigma} = \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0}_{r\times(n-r)} \\ 0 & 0 & \dots & \sigma_r & \\ & \mathbf{0}_{(m-r)\times r} & & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right)$$

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 $\mu_i$. Set

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{pmatrix}.$$

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 $\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_r \end{pmatrix}$.

   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

$$\begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{pmatrix}^T \mathbf{x} = \mathbf{0}$$

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

$$\mathbf{U} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m).$$

5. Then

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

**Note**

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

where $\mu_i, i = 1, \dots, n$ is the eigenvalues of $\mathbf{A}^T\mathbf{A}$; that is $\mathbf{A}^T\mathbf{A} = \mathbf{P\Sigma}^T\mathbf{\Sigma P}^T$, where $\mathbf{P} = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n)$ and $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 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 $\alpha, \beta$ ,

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

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 $\alpha$,
$$T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$$
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 $\mathbf{u}_1, \mathbf{u}_2 \dots, \mathbf{u}_k$ in $\mathbb{R}^n$ and scalars $c_1, c_2, \dots, c_k$,

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

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 $\alpha$ 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

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

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

**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 $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ 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

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

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 $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ 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 rows, $rank(\mathbf{A}) = m$.
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

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