Lecture 1 (Groups & Fields)

Definition: 1. Let G be a non empty set. A function $* : G \times G \longrightarrow G$ is called a **binary operation** on G.

Definition: 2. A non empty set G together with a binary operation $*$ is called a **group**, denoted as $(G, *)$, if it satisfies the following three properties:

1. $a * (b * c) = (a * b) * c \quad \forall \quad a, b, c \in G$ (Associativity);
2. there exists a unique element $e \in G$ such that $a * e = e * a = a \quad \forall a \in G$. The element e is called the identity element of G (Existence of identity);
3. for each $a \in G, \quad \exists \quad b \in G$ such that $a * b = b * a = e$. The element b is called the inverse of a and is denoted as a^{-1} (Existence of inverse).

In addition, if a group $(G, *)$ satisfies $a * b = b * a \quad \forall \quad a, b \in G$, then G is called a **commutative or an abelian group**.

Examples:

1. The set of real numbers $\mathbb{R}$, set of rational numbers $\mathbb{Q}$, set of integers $\mathbb{Z}$ form a group under usual addition.
2. The set of all $m \times n$ matrices with real entries $M_{m\times n}(\mathbb{R})$ forms a group under matrix addition.
3. Let $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$. Then $(Q^*, *)$ is a group under the usual multiplication. Similarly, $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ and $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ are groups under usual multiplication.
4. **Permutation/Symmetric Groups:** Let $S_n = \{\sigma \mid \sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ is a bijection$\}$. Then, S_n has $n!$ elements and forms a group with respect to composition of functions.

   Let $\sigma \in S_n$. Then,

   (a) σ can be written as $\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}$.

   (b) σ is one-one. Hence, $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\} = \{1, 2, \ldots, n\}$ and thus, $\sigma(1)$ has n choices, $\sigma(2)$ has $n-1$ and so on. Therefore, S_n has $n!$ elements.

   (c) $\sigma_1 \circ \sigma_2 \in S_n$ for any $\sigma_1, \sigma_2 \in S_n$. Thus, the operation $\circ$ on S_n is binary.

   (d) the associativity holds as $\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$ for all permutations $\sigma_1, \sigma_2, \sigma_3 \in S_n$. (Check yourself!)

   (e) the permutation $\sigma_0 \in S_n$ given by $\sigma_0(i) = i$ for $1 \leq i \leq n$ is the identity element of S_n.

   (f) for each $\sigma \in S_n$, σ^{-1} given by $\sigma^{-1}(m) = l$ if $\sigma(l) = m$ is the inverse element of σ in S_n. (Exercise: Show that σ^{-1} is well-defined and a bijection.)

Here, we discuss a few properties and results on permutation groups, which we will use later to define determinant function.

Proposition: 3. *Fix a positive integer n. Then, the group S_n satisfies the following:*

1. *Let $\tau \in S_n$. Then $\{\tau \circ \sigma : \sigma \in S_n\} = S_n$.*

2. *$S_n = \{\sigma^{-1} : \sigma \in S_n\}$.*

Proof. Part 1: Note that $\{\tau \circ \sigma : \sigma \in S_n\} \subseteq S_n$. Thus, $\{\tau \circ \sigma : \sigma \in S_n\} \neq S_n$ if and only if $\tau \circ \sigma_1 = \tau \circ \sigma_2$ for some $\sigma_1 \neq \sigma_2 \in S_n$, which is not possible. (Justify it!)

Part 2: Note that $\{\sigma^{-1} : \sigma \in S_n\} \subseteq S_n$ and equality does not hold only when $\sigma_1^{-1} = \sigma_2^{-1}$, where $\sigma_1 \neq \sigma_2 \in S_n$. But we know that $(\sigma^{-1})^{-1} = \sigma$ and get a contradiction.

Definition: 4 (Cyclic Notation)**.** *Let $\sigma \in S_n$. Suppose there exist r, $2 \leq r \leq n$ and $i_1, i_2, \ldots, i_r$ such that $\sigma(i_1) = i_2, \sigma(i_2) = i_3, \ldots, \sigma(i_r) = i_1$ and $\sigma(j) = j$ for all $j \neq i_1, i_2, \ldots, i_r$. Then, we represent such a permutation by $\sigma = (i_1\, i_2\, \ldots\, i_r)$ and call it an r-cycle.*

For Example, $\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 1 & 4 & 6 \end{pmatrix} = (1\,3\,5\,4)$ and $\sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 & 5 \end{pmatrix} = (2\,3)$.

Remark: 1. *1. Every permutation is either a cycle or product of disjoint cycles. For example,* $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 6 & 3 & 5 & 7 & 1 & 4 & 9 & 8 \end{pmatrix} = (1\,2\,6)(4\,5\,7)(8\,9)$.

2. A cycle of length 2 is called **transposition**.

3. For any cycle $(i_1\, i_2\, \ldots\, i_r)$, $(i_1\, i_2\, \ldots\, i_r) = (i_1\, i_r)(i_1\, i_{r-1}) \cdots (i_1\, i_2)$.

4. Every permutation is a product of transpositions. For example, $(1\,2\,3) = (1\,3)(1\,2)$ and $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 6 & 3 & 5 & 7 & 1 & 4 & 9 & 8 \end{pmatrix} = (1\,2\,6)(4\,5\,7)(8\,9) = (1\,6)(1\,2)(4\,7)(4\,5)(8\,9)$.

Definition: 5. *A permutation $\sigma \in S_n$ is called an* **even permutation** *if it can be written as product of even number of transpositions or it is the identity permutation and it is called an* **odd permutation** *if it can be written as a product of odd number of transpositions.*

Remark: 2. *1. A decomposition of a permutation into a product of transposition need not be unique. (Look for examples!)*

2. A permutation is either always even or always odd, that is, if a permutation can be expressed as a product of an even number of transpositions, then every decomposition of that permutation into transpositions must have an even number of transpositions.

Definition: 6. *A function sgn: $S_n \to \{1, -1\}$, called the* **signature of a permutation**, *by*

$$sgn(\sigma) = \begin{cases} 1 & \text{if } f \text{ is an even permutation} \\ -1 & \text{if } f \text{ is an odd permutation} \end{cases}$$

Remark: 3. *1. If σ and τ are both even or both odd permutations, then $\sigma \circ \tau$ and $\tau \circ \sigma$ are both even. Whereas, if one of them is odd and the other even then $\sigma \circ \tau$ and $\tau \circ \sigma$ are both odd.*

2. *The identity permutation* σ_0 *is an even permutation and hence* $sgn(\sigma_0) = 1$.

3. *A transposition is an odd permutation and hence its signature is -1.*

4. $sgn(\sigma \circ \tau) = sgn(\sigma)sgn(\tau)$.

Definition: 7. Let $\mathbb{F}$ be a non-empty set with two binary operations addition denoted as $+$ and multiplication denoted as $\cdot$. Then $\mathbb{F}$ is called a **field**, denoted as $(\mathbb{F}, +, \cdot)$, if

1. $\mathbb{F}$ is an abelian group under addition $+$;

2. $\mathbb{F}^* = \mathbb{F} \setminus \{e\}$ is an abelian group under multiplication ., where e denotes the additive identity of $\mathbb{F}$;

3. $a \cdot (b + c) = a \cdot b + a \cdot c \quad \forall \quad a, b, c \in \mathbb{F}$.

Definition: 8. Let $\mathbb{F}$ be a field and $\mathbb{F}_1 \subseteq \mathbb{F}$. Then $\mathbb{F}_1$ is said to be a **subfield** of $\mathbb{F}$ if $\mathbb{F}_1$ is itself a field under the same binary operations defined on $\mathbb{F}$.

Examples:

1. The set of complex numbers $\mathbb{C}$ forms a field under usual addition and multiplication of complex numbers.

2. The sets $\mathbb{R}$ and $\mathbb{Q}$ form a field under usual addition and multiplication.

3. The set of integers $\mathbb{Z}$ does not form a field under usual addition and multiplication.

4. $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and $\mathbb{R}$ is a subfield of $\mathbb{C}$.

Note: The elements of a field are also called scalars.

Lecture 2
System of Linear Equations

Definition 1. An equation of the form $a_1x_1 + \ldots + a_nx_n = b$, where $b, a_1, a_2, \ldots, a_n$ are constants, is called a **linear equation** in n unknowns. If the constants $a_1, \ldots, a_n$ and b are from a set X, the equation is called a linear equation over the set X.

Throughout this course, we deal with linear equations over the field $\mathbb{R}$ or $\mathbb{C}$.

System of linear equations: Let $\mathbb{F}$ be a field and $a_{ij}, b_j \in \mathbb{F}$, for $1 \leq i \leq m$, and $1 \leq j \leq n$. The following system is called a system of m linear equations in n unknowns over $\mathbb{F}$.

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m \end{array} \qquad (1)$$

If $b_j = 0$ for all $1 \leq j \leq m$, the system is called a **homogeneous system of linear equations**, otherwise it is called a **non-homogeneous system of linear equations**. The above system can be written as $\sum_{j=1}^{n} a_{ij}x_j = b_i$, for $1 \leq i \leq m$. An n-tuple $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ is called a **solution** of this system if it satisfies each of the equations of the system.

First we consider a system having only one equation

$$2x + 3y + 4z = 5.$$

Both $(1, 1, 0)$ and $(-1, 1, 1)$ satisfy this equation. In fact, for any real numbers x and y and we can find z by substituting the values of x and y in the equation. Geometrically, the collection of all the solutions of the equation $2x + 3y + 4z = 5$ is a plane in $\mathbb{R}^3$.

Now, we consider a system of two linear equations:

$$\begin{aligned} 2x + 3y + 4z &= 5 \\ x + y + z &= 2 \end{aligned}$$

A solution of this system is a solution of the first equation which is also a solution of the second equation. If A_i $(i = 1, 2)$ is the set of solutions of the i-th equation, then the set of solutions of the system is $A_1 \cap A_2$. Here, we know that for each i, A_i is a plane in $\mathbb{R}^3$. Thus, the solution of system is the intersection of two planes. In $\mathbb{R}^3$, the intersection of two planes is either an empty set (plane are parallel) or a line or a plane (the planes are identical). For this system, the solution set is $A_1 \cap A_2 = \{(1,1,0) + z(1,-2,1) : z \in \mathbb{R}\}$ which represents a line in $\mathbb{R}^3$. (Check it yourself!)

Remark 2. *1. A non-homogeneous system of 2 linear equations in 3 unknowns over $\mathbb{R}$ has either no solution or infinitely many solutions.*

2. A homogeneous system of 2 linear equations in 3 unknowns over $\mathbb{R}$ always has infinitely many solutions.

Now consider the following two systems of linear equations:

$$\begin{aligned} 2x + 3y + 4z &= 5 \\ x + y + z &= 2 \\ y + z &= 1 \end{aligned} \qquad \begin{aligned} 2x + 3y + 4z &= 5 \\ x + y + z &= 2 \\ x + 2y + 3z &= 2 \end{aligned} \qquad \begin{aligned} 2x + 3y + 4z &= 5 \\ x + y + z &= 2 \\ x + 2y + 3z &= 3 \end{aligned}$$

The first system has the unique solution, that is, $(1,1,0)$, the second system has no solution and the third system has more than one solution, in fact, infinitely many solutions.

Question 3. *When System (1) has no solution or a unique solution or infinitely many solutions?*

System (1) can be described by the following matrix equation:

$$Ax = b,$$

where $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in M_{m\times n}(\mathbb{F})$, $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in M_{n\times 1}(\mathbb{F})$ and $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in M_{m\times 1}(\mathbb{F})$.

The matrix A is called the **coefficient matrix**, x is the matrix (or column) of unknowns and b is the matrix (or column) of constants.

The matrix $(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right) \in M_{m\times(n+1)}(\mathbb{F})$ obtained by attaching the column b with A, is called the **augmented matrix** of the system.

Some properties of a system of linear equations

Let $Ax = b$ ($\sum_{j=0}^{n} a_{ij}x_j = b_i$ for $1 \le i \le m$) be a non-homogeneous system of linear equations and $Ax = 0$ ($\sum_{j=0}^{n} a_{ij}x_j = 0$ for $1 \le i \le m$) be the associated homogeneous system. Let S and S_h denote the solution sets of the systems $Ax = b$ and $Ax = 0$ respectively. The addition of two elements in $\mathbb{F}^n$ is given by

$$x + y = (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

and the scalar multiplication is given by

$$\alpha x = \alpha(x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

Then we have the following statements.

P1: $x = (x_1, \ldots, x_n) \in S$ and $y = (y_1, \ldots, y_n) \in S_h \Rightarrow x + \alpha y \in S$ for all $\alpha \in \mathbb{R}$.

Proof: The i-th component of $A(x + \alpha y)$ is $\sum_{j=1}^{n} a_{ij}(x_j + \alpha y_j) = \sum_{j=1}^{n} a_{ij}x_j + \alpha \sum_{j=1}^{n} a_{ij}y_j = \sum_{j=1}^{n} a_{ij}x_j = b_i$ for $1 \le i \le m$. Therefore, $A(x + \alpha y) = b$ so that $x + \alpha y \in S$. □

P2: Let $x \in S$ and $x + S_h := \{x + y : y \in S_h\}$. Then $S = x + S_h$.

Proof: **P1**$\Rightarrow x + S_h \subseteq S$. Also, for all $z \in S$, $z = x + (z - x) \in x + S_h \Rightarrow S \subseteq x + S_h$. □

P3: If the system $Ax = b$ has more than one solution, then it has infinitely many solutions.

Proof: Let x, y be two solutions of $Ax = b$. Then it is easy to see that $\alpha x + (1 - \alpha)y$ is again a solution for each $\alpha \in \mathbb{R}$. □

P4: If $Ax = 0$ has a non zero solution, then it has infinitely many solutions. (Do it yourself!)

Exercise 4. *Classify the following systems in the categories:*

1) The system has no solution 2) Exactly one solution 3) More than one solution

1. $x_1 + x_2 + x_3 = 3$, $x_1 + 2x_2 + 3x_3 = 6$, $x_2 + 2x_3 = 1$.

2. $x_1 + x_2 + x_3 = 3$, $x_1 + 2x_2 + 3x_3 = 6$, $x_1 + x_2 + 2x_3 = 4$.

3. $x_1 + x_2 + x_3 = 3$, $x_1 + 2x_2 + 3x_3 = 6$, $x_2 + 2x_3 = 3$.

Definition 5. An equation $d_1x_1 + d_2x_2 + \ldots + d_nx_n - e = 0$ is called a **linear combination** of the equations Eq_i if it can be written as $c_1\mathrm{Eq}_1 + c_2\mathrm{Eq}_2 + \cdots + c_n\mathrm{Eq}_n$, where $\mathrm{Eq}_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n - b_i$ and $c_i \in \mathbb{F}$ for $1 \leq i \leq n$.

Remark 6. *If $(x_1, x_2, \ldots, x_n)$ is a solution of System (1), then it is a solution of $d_1x_1 + d_2x_2 + \ldots + d_nx_n - e = 0$. But converse need not be true. For instance, consider the following systems:*

$$\begin{aligned} x + y + z &= 1 \\ x + y &= 1 \\ x - z &= 1 \end{aligned} \qquad\qquad \begin{aligned} x + y + z &= 1 \\ 2x + y - z &= 2 \end{aligned}$$

Then latter one is obtained from former system. We see that $(-1, 3, -1)$ is solution of the latter one but not of the former one.

Definition 7. Two systems, say S_1 and S_2 of linear equations, are called equivalent if each equation of S_1 is a linear combination of the equations of S_2 and vice versa.

Theorem 8. The solution sets of equivalent systems of linear equations are identical.

Lecture 3
Elementary Matrices & Row Reduced Echelon Form

Definition 1. Elementary row/column operations: *Let A be an $m \times n$ matrix and $R_1, \ldots, R_m$ denote the rows of A and $C_1, \ldots, C_n$ denote the columns of A. Then an elementary row(column) operation is a map from $M_{m\times n}(\mathbb{F})$ to itself which is any one of the following three types:*

1. *Multiplying the i-th row(column) by a nonzero scalar $\lambda \in \mathbb{F} \setminus \{0\}$ denoted by $R_i \to \lambda R_i$ ($C_i \to \lambda C_i$).*
2. *Interchanging the i-th row(column) and the j-th row(column) denoted by $R_i \leftrightarrow R_j$ ($C_i \leftrightarrow C_j$).*
3. *For $i \neq j$, replacing the i-th row(column) by the sum of the i-th row(column) and μ multiple of the j-th row(column) denoted by $R_i \to R_i + \mu R_j$ ($C_i \to C_i + \mu C_j$).*

A **row operation** is a map from $M_{m\times n}(\mathbb{F})$ to itself which is a composition of finitely many elementary row operations.

Remark 2. 1. *Every elementary row operation is invertible.*

1. The inverse of $R_i \to \lambda R_i$ is $R_i \to \frac{1}{\lambda} R_i$;

2. The inverse of $R_i \leftrightarrow R_j$ is $R_i \leftrightarrow R_j$ (self inverse, i.e., inverse of itself);

3. The inverse of $R_i \to R_i + \mu R_j$ is $R_i \to R_i - \mu R_j$.

2. *Let ρ be an elementary row operation and $A \in M_{m\times n}(\mathbb{F})$. Then $\rho(A) = \rho(I_m)\,A$, where I_m is the $m \times m$ identity matrix.*

ρ(A) = ρ(Im)A = Ei . A

Definition 3. Elementary Matrix: *Let I_m denote the $m \times m$ identity matrix. A matrix obtained by performing an elementary row operation on I_m is called an elementary matrix. Therefore, there are three types of elementary matrices:*

1. $E_i(\lambda)$, obtained by multiplying the i-th row by a nonzero scalar $\lambda \in \mathbb{F} \setminus \{0\}$ of I_m.

2. E_{ij}, obtained by interchanging the i-th row and the j-th row of I_m.

3. $E_{ij}(\mu)$, obtained by replacing the i-th row by the sum of the i-th row and μ multiple of the j-th row of I_m.

Example 4.

$$M_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad M_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad M_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

In the above, M_1, M_2 are elementary matrices but M_3 is not.

Remark 5. *1. Performing an elementary row operation on a matrix A is same as pre multiplication of the respective elementary matrix to A.*
2. Performing an elementary column operation on a matrix A is same as post multiplication of the respective elementary matrix to A.

Definition 6. Row-equivalent matrices: *Let A and B be two $m \times n$ matrices over a field $\mathbb{F}$. Then B is said to be row-equivalent to A if B is obtained from A by performing a finite sequence of elementary row operations.*

Theorem 7. If A and B are row equivalent matrices, the homogeneous systems of linear equations $Ax = 0$ and $Bx = 0$ have exactly the same solutions.

Proof: It is given that A and B are row equivalent, that is, there exist elementary row operations, $\rho_1, \rho_2, \ldots, \rho_k$, such that $B = \rho_k \circ \cdots \circ \rho_2 \circ \rho_1(A)$, equivalently,

$$A = A_0 \to A_1 \to \cdots \to A_k = B.$$

It is enough to show that $A_j x = 0$ and $A_{j+1}x = 0$ have the same solutions. In words, elementary operations do not make any change to the solution set. Let $k = 1$. Then $B = \rho(A) \Rightarrow B = \rho(I_m)A$. If $Ax = 0$, then $Bx = \rho(I_m)Ax = \rho(I_m)0 = 0$. Similarly, if $Bx = 0$, then $Ax = \rho^{-1}(B)x \Rightarrow Ax = \rho^{-1}(I_m)Bx = 0$.

Definition 8. Row-equivalent systems: *The systems of linear equations $Ax = b$ and $Cx = d$ are said to be row equivalent if their respective augmented matrices, $(A|b)$ and $(C|d)$ are row equivalent.*

Theorem 9. *Let $Ax = b$ and $Cx = d$ be two row equivalent linear systems. Then they have the same solution set.*

Proof: Let $E_1, E_2, \ldots, E_k$ be the elementary matrices such that $E_1E_2 \cdots E_k(A|b) = (C|d)$. Suppose y is a solution of $Ax = b$, i.e., $Ay = b$. Then $Cy = E_1E_2 \cdots E_kAy = E_1E_2 \cdots E_kb = d$. Similarly, we can prove that any solution of $Cx = d$ is a solution of $Ax = b$..

Definition 10. Row echelon form: *A form of a matrix satisfying the following properties is called row echelon matrix.*

1. *Every zero-row of A (row which has all its entries 0) occurs below every non-zero row (which has a non-zero entry);*

2. *Suppose the matrix has r nonzero rows (and remaining $m-r$ rows are zero). If the leading coefficient (the first non-zero entry) of i-th row $(1 \leq i \leq r)$ occurs in the k_i-th column, then $k_1 < k_2 < \cdots < k_r$,*

that is, the leading coefficient of each row after the first is positioned to the right of the leading coefficient of the previous row.

Definition 11. Row reduced echelon form: *A form of a matrix satisfying the following properties is called row reduced echelon form (in short RRE) or reduced echelon form:*

1. *The matrix is in row echelon form;*
2. *The leading coefficient of each row is 1;*
3. *All other elements in a column that contains a leading coefficient are zero.*

Remark 12. 1. *The process of computing row echelon form of a matrix by performing row operations is called* **Gaussian elimination**.

2. *The process of computing row-reduced echelon form of a matrix by applying row operations is called* **Gaussian-Jordan elimination**.

3. *Every matrix is row equivalent to a row-reduced echelon matrix. In fact, row-reduced echelon form of a matrix is unique.*

Example 13. *Find the RRE form of* $\begin{pmatrix} 0 & 0 & 4 & 1 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{pmatrix}$.

Solution:

$$\begin{pmatrix} 0 & 0 & 4 & 1 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{pmatrix} \overset{R_3 \leftrightarrow R_4}{\sim} \begin{pmatrix} 0 & 0 & 4 & 1 \\ 0 & 3 & 0 & 1 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_1 \leftrightarrow R_2}{\sim} \begin{pmatrix} 0 & 3 & 0 & 1 \\ 0 & 0 & 4 & 1 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_1 \to \frac{1}{3} R_1}{\sim} \begin{pmatrix} 0 & 1 & 0 & 1/3 \\ 0 & 0 & 4 & 1 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\overset{R_3 \to R_3 - 4R_1}{\sim} \begin{pmatrix} 0 & 1 & 0 & 1/3 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 2 & -4/3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_2 \to \frac{1}{4} R_2}{\sim} \begin{pmatrix} 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 1/4 \\ 0 & 0 & 2 & -4/3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_3 \to R_3 - 2R_2}{\sim} \begin{pmatrix} 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 1/4 \\ 0 & 0 & 0 & -11/6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\overset{R_3 \to \frac{-6}{11} R_3}{\sim} \begin{pmatrix} 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 1/4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_1 \to R_1 - R_3,\, R_2 \to R_2 - \frac{1}{4} R_3}{\sim} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Remark 14. *Consider a system of m linear equations in n unknowns $Ax = 0$. Let R be the RRE form of A with r non-zero rows. The number of leading columns (column which contains a leading coefficient)* *is r. We call the variables associated with leading columns* **leading variables** *or* **dependent variables**. *The variables other than the dependent variables are called* **free variables** *or* **independent variables**. *Note that, either there is no equation in which the free variable appears or it appears with at least one another variable.*

For instance, if we consider a homogeneous system of linear equation $Ax = 0$, where A is as in Example 13. *Then the columns 2,3 and 4 are leading columns and hence, x_2, x_3 and x_4 are leading variables or dependent variable, and x_1 is an independent (free) variable which is not appearing in any of the reduced equations.*

Now consider a homogeneous system corresponding to the matrix $\begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & 2 & 1 & 4 \\ 1 & 2 & -1 & 2 \end{pmatrix}$. *Then the RRE form of A is* $\begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. *Here, x_1, x_3 are leading or dependent variables and x_2, x_4 are free or independent variables.*

Theorem 15. *If A is an $m \times n$ matrix and $m < n$, then the homogeneous system of linear equations, $Ax = 0$, has a non trivial solution.*

Proof: Let R be the row-reduced echelon form of A. Then the systems $Ax = 0$ and $Rx = 0$ have the same solution set. If r is the number of non-zero rows in R, then $r \leq m$ so that $r < n$, equivalently, $n - r > 0$. Thus, there exists at least one independent variable (free variable) for the system $Rx = 0$, and hence $Rx = 0$ has a non-trivial solution. $\square$

Theorem 16. *Let $A \in M_{n\times n}(\mathbb{F})$. The matrix A is row equivalent to the $n \times n$ identity matrix if and only if the system of equations $Ax = 0$ has only the trivial solution.*

Proof: If A is row equivalent to the $n \times n$ identity matrix I_n, then $Ax = 0$ and $I_n x = 0$ have only the trivial solution. Conversely, suppose $Ax = 0$ has only the trivial solution and R is the RRE form of A. Let r be the number of non-zero rows of R. Then $r \leq n$. Note that $Rx = 0$ has only the trivial solution (as $Ax = 0$ and $Rx = 0$ are row equivalent) so that $r \geq n$. Hence, $r = n$ and $R = I_n$. $\square$

Lecture 4
Invertible Matrix & Gauss-Jordan Method

Definition 1. Invertible Matrix: *A square matrix M is said to be* **invertible** *if there exists a matrix N of the same order such that $MN = NM = I$. The matrix N is called inverse of M and is denoted as M^{-1}.*

Theorem 2. Let A and B be two $n \times n$ matrices then: (a) if A is invertible, then so is A^{-1} with $(A^{-1})^{-1} = A$; (b) if both A and B are invertible, then so is AB with $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem 3. An elementary matrix is invertible.

Proof: Let E be an elementary matrix corresponding to the elementary row operation ρ. If ρ' is the inverse operation of ρ and $E' = \rho'(I)$, then $EE' = \rho(I)\rho'(I) = \rho(\rho'(I)) = (\rho \circ \rho')(I) = I$ and $E'E = \rho'(I)\rho(I) = \rho'(\rho(I)) = (\rho' \circ \rho)(I) = I$ so that E is invertible. □

Theorem 4. Let A be an $m \times n$ matrix. Then by applying a sequence of row and column operations A can be reduced to the form

$$\begin{bmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$$

which is called the **normal form** of the matrix, equivalently, there exist elementary row matrices $E_1, \ldots, E_s$, and elementary column matrics $F_1, \ldots, F_k$ such that

$$E_1 \cdots E_s A F_1 \cdots F_k = \begin{bmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix}.$$

Theorem 5. Let A be an $n \times n$ matrix. Then A is invertible if and only if A is a product of elementary matrices.

Proof: If A is an invertible matrix then there exist elementary matrices $E_1, \ldots, E_s, F_1, \ldots, F_k$ such that

$$E_1 \cdots E_s A F_1 \cdots F_k = \begin{bmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix} = I_n.$$

Ei(λ) = Fi(λ)
Eij = Fji
Eij(λ) = Fji(λ)

Therefore, $A = E_s^{-1} \ldots E_1^{-1} I_n \; F_k^{-1} \ldots F_1^{-1}$. Note that an elementary column matrix is one of the elementary row matrices. Further, inverse of an elementary matrix is again an elementary matrix. Hence, A is a product of elementary matrices. Converse follows from the fact that the product of invertible matrices is invertible. □

Theorem 6. Let A be an $n \times n$ matrix. Then A is invertible if and only if A can be reduced to the identity matrix I_n by performing a finite sequence of elementary row operations on A.

Proof: If A is invertible then by above theorem $A = E_k \cdots E_1$ for some $k \in \mathbb{N}$, equivalently $E_1^{-1} \cdots E_k^{-1} A = I$. Thus A can be reduced to identity matrix. Conversely, if A can be reduced to the identity matrix I_n by performing a finite sequence of elementary row operations on A. Then there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $E_k \cdots E_1 A = I$, then $A = E_1^{-1} \cdots E_s^{-1}$. Therefore, A is invertible as product of invertible matrices is invertible.

Gauss-Jordan Method for finding inverse: Let A be an invertible matrix. Then there exist elementary matrices E_1, E_2, ..., E_k such that $I = E_k E_{k-1} \ldots E_1 A$ which is equivalent to $A^{-1} = E_k E_{k-1} \ldots E_1 I$. This shows that sequence of elementary operations which reduces A to the identity matrix I, also reduces I to A^{-1} by performing in the same order.

(A | I) ---> (I | A^-1)

Example 7. *Find inverse of* $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix}$ *by using Gauss-Jordan method.*

$$(A|I) = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right) \overset{R_2 \to R_2 - R_1, R_3 \to R_3 - R_1}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array}\right)$$

$$\overset{R_3 \to R_3 - R_2, R_1 \to R_1 - R_2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 0 & -1 & 1 \end{array}\right) \overset{R_3 \to R_3/2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1/2 & 1/2 \end{array}\right)$$

$$\overset{R_1 \to R_1 - R_3}{\sim} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1/2 & -1/2 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1/2 & 1/2 \end{array}\right) = (I \mid A^{-1})$$

Therefore, $A^{-1} = \begin{pmatrix} 2 & -1/2 & -1/2 \\ -1 & 1 & 0 \\ 0 & -1/2 & 1/2 \end{pmatrix}$.

Gauss-Jordan elimination method for finding solutions of a system of linear equations Let $AX = B$ be a system of linear equations. Now consider the augmented matrix $(A|B)$. Apply finite number of elementary row operations to get the form $(A'|B')$. Here $(A'|B')$ is row reduced echelon form of the matrix $(A|B)$. Thus $(A'|B')$ is row equivalent to $(A|B)$, therefore $AX = B$ and $A'X = B'$ are equivalent systems and hence they have the same solution.

Example 2: Solve the following system of linear equations

$$x + 3y + z = 9$$
$$x + y - z = 1$$
$$3x + 11y + 5z = 35.$$

Solution: $(A|B) = \left(\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 5 & 35 \end{array}\right) \overset{R_2 \to R_2 - R_1, R_3 \to R_3 - 3R_1}{\sim} \left(\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 2 & 8 \end{array}\right)$

$\overset{R_3 \to R_3 - R_2}{\sim} \left(\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 0 & 0 \end{array}\right) \overset{R_2 \to -R_2/2}{\sim} \left(\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right) \overset{R_1 \to R_1 - 3R_2}{\sim} \left(\begin{array}{ccc|c} 1 & 0 & -2 & -3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right) = (A'|B').$

The equivalent system is

$$x - 2z = -3$$
$$y + z = 4.$$

The solution set is $\{(2z - 3, 4 - z, z) : z \in \mathbb{R}\}$.

Definition 8. *A system of linear equation $Ax = b$ is said to be consistent if it has at least one solution (unique or infinitely many) and the system is called inconsistent if it has no solution.*

Theorem 9. *Consider a system of linear equation $Ax = b$, where $A \in M_{m\times n}(\mathbb{R})$. Suppose R and $(R|b')$ are the RRE forms of A and $(A|b)$ respectively. Let r and r' be the number of non-zero rows in R and $(R|b)$. Then*

1. *if $r \neq r'$, the system is inconsistent.*
2. *if $r = r' = n$, the system the unique solution.*
3. *if $r = r' < n$, the system has infinitely many solutions.*

Proof. Case 1: Note that $r' \geq r$. If $r \neq r'$, then $(R|b')_{r+1,n+1} = 1$ whereas $(R|b')_{r+1,j} = 0$ for all $j < n+1$. Suppose the system $Ax = b$ is consistent and y is one of its solutions. Then y is a solution of $Rx = b'$ (row-equivalent systems). The $r+1$-th equation of $Rx = b'$ gives that $0 = 1$, which is absurd, hence the system has no solution, that is, the system is inconsistent.

~~b"?~~ ans. b" are all the non-zero elements of b'

$\boxed{\text{Case 2:}}$ If $r = r' = n$, then $(R|b') = \left(\begin{array}{c|c} I_n & b''_{n\times 1} \\ 0_{m-n\times n} & 0_{m-n\times 1} \end{array}\right)$. Therefore, $x = b''$ is the only solution of the system $Ax = b$.

Case 3: If $r = r' < n$, then $(R|b') = \left(\begin{array}{c|c} R'_{r\times n} & b''_{r\times 1} \\ 0_{m-r\times n} & 0_{m-r\times 1} \end{array}\right)$ so that the system $Rx = b'$ is equivalent to the system $R'x = b''$ for which the number of equations is less than the number of variables. Thus, $R'x = b''$ has infinitely many solutions and so $Rx = b'$ as well as $Ax = b$. $\square$

Example 3: Find $a, b \in \mathbb{R}$ such that the following system of equations (i) is consistent, and (ii) is inconsistent (iii) has a unique solution (iv) has infinitely many solutions.

$$x + ay = 1, 2x + y = b.$$

The augmented matrix of the system is $\left(\begin{array}{cc|c} 1 & a & 1 \\ 2 & 1 & b \end{array}\right)$. Thus,

$$\left(\begin{array}{cc|c} 1 & a & 1 \\ 2 & 1 & b \end{array}\right) \overset{R_2 \to R_2 - 2R_1}{\sim} \left(\begin{array}{cc|c} 1 & a & 1 \\ 0 & 1-2a & b-2 \end{array}\right)$$

Case 1: If $1 - 2a = 0$ and $b - 2 \neq 0$. Then, the RRE form is $\left(\begin{array}{cc|c} 1 & a & 1 - \frac{1}{b-2} \\ 0 & 0 & 1 \end{array}\right)$. Thus, $r = 1$ and $r' = 2$. Therefore, the system has no solution (system is inconsistent).

Case 2: If $1 - 2a = 0$ and $b - 2 = 0$. Then, the RRE form is $\left(\begin{array}{cc|c} 1 & a & 1 - \frac{1}{b-2} \\ 0 & 0 & 0 \end{array}\right)$. Thus, $r = r' = 1 < 2$. Therefore, the system has infinitely many solutions.

Case 3: If $1 - 2a \neq 0$ and $b \in \mathbb{R}$. Then, the RRE form is $\left(\begin{array}{cc|c} 1 & 0 & 1 - a\frac{b-2}{1-2a} \\ 0 & 1 & \frac{b-2}{1-2a} \end{array}\right)$. Thus, $r = r' = 2$. Therefore, the system has unique solution.

Hence,

(i) the system is consistent when either $a \neq 1/2$, and $b \in \mathbb{R}$ or $a = 1/2$ and $b = 2$.

(ii) the system is inconsistent when $a = 1/2$ and $b \neq 2$.

(iii) the system has a unique solution if $a \neq 1/2$ and $b \in \mathbb{R}$.

(iv) the system has infinitely many solutions if $a = 1/2$ and $b = 2$.

Lecture 5
Determinant Function & Its Properties

Definition 1. *Let $A = (a_{ij})$ be an $n \times n$ matrix and S_n denote the set of all permutation on $S = \{1, 2, \dots, n\}$. Then determinant is a function from $M_n(\mathbb{F})$ to $\mathbb{F}$, denoted by $\det(A)$ or $|A|$, and given by*

$$\boxed{\det(A) = |A| = \sum_{\sigma \in S_n} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.}$$

Let $n = 2$ and $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Then $S_n = \{(1), (1\,2)\}$. Set $\sigma_1 = (1)$ and $\sigma_2 = (1\,2)$. Then

$$\begin{aligned}
\det(A) &= sign(\sigma_1) a_{1\sigma_1(1)} a_{2\sigma_1(2)} + sign(\sigma_2) a_{1\sigma_2(1)} a_{2\sigma_2(2)} \\
&= a_{11}a_{22} - a_{12}a_{21}.
\end{aligned}$$

Properties of Determinant

P1: Let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ matrices. Then if B is obtained from A by interchanging two rows of A, then $|A| = -|B|$.

Proof: Let B is obtained by interchanging k-th row and r-th row of A. Then $B = (b_{ij})$ such that $b_{kj} = a_{rj}$, $b_{rj} = a_{kj}$, and $b_{ij} = a_{ij}$ for $j = 1, 2, \dots, n$ and $i \neq k, r$.

Let $\tau = (k, r)$. Then $S_n = \{\sigma \circ \tau \,:\, \sigma \in S_n\}$. Therefore,

$$\begin{aligned}
|B| &= \sum_{\sigma \circ \tau} sign(\sigma \circ \tau) b_{1\sigma\circ\tau(1)} \cdots b_{k\sigma\circ\tau(k)} \cdots b_{r\sigma\circ\tau(r)} \cdots b_{n\sigma\circ\tau(n)} \\
&= \sum_{\sigma} sign(\sigma) sign(\tau) b_{1\sigma(1)} \cdots b_{k\sigma(r)} \cdots b_{r\sigma(k)} \cdots b_{n\sigma(n)} \\
&= -\sum_{\sigma} sign(\sigma) a_{1\sigma(1)} \cdots a_{r\sigma(r)} \cdots a_{k\sigma(k)} \cdots a_{n\sigma(n)} && (\text{since } sign(\tau) = -1) \\
&= -|A|
\end{aligned}$$

P2: If two rows of A are identical, then $|A| = 0$.

Proof: Let $R_1, R_2, \dots, R_n$ denote the rows of A. It is given that $R_k = R_j$ for some $j \neq k$. Let B be the matrix obtained by interchanging j-th row and k-th row of A. Then $|B| = -|A|$, but $A = B$. Therefore, $|A| = 0$.

P3: If B is obtained by multiplying a row of A by a constant c, then $|B| = c|A|$.

Proof: Let $B = (b_{ij})$ is obtained by multiplying a constant c to the k-th row of A. Then $b_{kj} = ca_{kj}$ and $b_{ij} = a_{ij}$ for $i \neq k$. Then

$$\begin{aligned}
|B| &= \sum_{\sigma} sign(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{k\sigma(k)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots ca_{k\sigma(k)} \cdots a_{n\sigma(n)} \\
&= c \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{k\sigma(k)} \cdots a_{n\sigma(n)} \\
&= c|A|
\end{aligned}$$

P4: Let A, B and C be $n \times n$ matrices which differ only in the k-th row, and $c_{kj} = a_{kj} + b_{kj}$ $\forall j$, then $|C| = |A| + |B|$.

Proof:

$$\begin{aligned}
|C| &= \sum_{\sigma} sign(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots c_{k\sigma(k)} \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma} sign(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots (a_{k\sigma(k)} + b_{k\sigma(k)}) \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma} sign(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots a_{k\sigma(k)} \cdots c_{n\sigma(n)} + \sum_{\sigma} sign(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots b_{k\sigma(k)} \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{k\sigma(k)} \cdots a_{n\sigma(n)} + \sum_{\sigma} sign(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{k\sigma(k)} \cdots b_{n\sigma(n)} \\
&= |A| + |B|
\end{aligned}$$

P5: If B is obtained by adding λ times the r-th row of A to its k-th row, then $|A| = |B|$.

Proof: Here, $b_{kj} = \lambda a_{rj} + a_{kj}$, $b_{ij} = a_{ij}$ for $i \neq k$ and $j = 1, 2, \ldots, n$. Then

$$\begin{aligned}
|B| &= \sum_{\sigma} sign(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{k\sigma(k)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots (\lambda a_{r\sigma(k)} + a_{k\sigma(k)}) \cdots a_{n\sigma(n)} \\
&= \lambda \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{r\sigma(k)} \cdots a_{n\sigma(n)} + \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots + a_{k\sigma(k)} \cdots a_{n\sigma(n)} \\
&= 0 + |A| = |A|
\end{aligned}$$

P6: Let E be an elementary matrix. Then $|E| \neq 0$.

P7: If E is an elementary matrix, then $|EA| = |E||A|$. (Prove it yourself!)

P8: A is invertible $\Leftrightarrow$ $|A| \neq 0$. (Prove it yourself!)

P9: Let A, B be $n \times n$ matrices. Then $|AB| = |A||B|$.

Proof: Suppose A is not invertible. Then $|A| = 0$. Let $|AB| \neq 0$ so that AB is invertible. Therefore, the system $ABx = 0$ has only trivial solution. But $Ax = 0$ has a non trivial solution, say y. If B is invertible, then $Bx = y$ has a unique solution, say x^*. Note that $x^* \neq 0$ and $ABx^* = 0$ so that $ABx = 0$ has a non-trivial solution which contradicts our assumption and hence, $|AB| = 0$. Now if $|B| = 0$, the system $Bx = 0$ has a non-trivial solution so that $ABx = 0$ has a non-trivial solution which again gives a contradiction. Therefore, $|AB| = 0$.

Suppose A is invertible. Then $A = E_1 \dots E_s$. This implies

$$\begin{aligned} |AB| &= |(E_1 \dots E_s B)| \\ &= |E_1||E_2| \dots |E_s||B| \\ &= |E_1 \dots E_s||B| \\ &= |A||B|. \end{aligned}$$

P10: $|A| = |A^t|$, where A^t denotes the transpose of A.

Remark 2. *The properties P1-P5 are also valid for column operations.*

Cramer's Rule for solving system of linear equations

Let $Ax = b$ be a system of n linear equations in n unknowns such that $|A| \neq 0$. Then the system $Ax = b$ has a unique solution given by

$$x_j = \frac{|C_j|}{|A|}, \ j = 1, 2, \dots, n$$

where C_j is the matrix obtained from A by replacing the j-th column of A with the column matrix $b = (b_1, b_2, \dots, b_n)^t$.

Proof: If $|A| \neq 0$, then A is invertible and $x = A^{-1}b$ is the unique solution of $Ax = b$. Define a matrix

$$X_j = \begin{pmatrix} 1 & 0 & \cdots & x_1 & \cdots & 0 \\ 0 & 1 & \cdots & x_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x_n & \cdots & 1 \end{pmatrix}.$$

$AX_j = C_j$

$\hookrightarrow$ jth

Note that the matrix $|X_j| = x_j$ (apply properties of determinant function). Therefore,

$$x_j = |X_j| = |I_n X_j| = |A^{-1} A X_j| = \frac{|AX_j|}{|A|} = \frac{|C_j|}{A_j} \,\forall j = 1, 2, \ldots, n.$$

Lecture 6

Vector Space and Its Properties

Definition 1. Let $\mathbb{F}$ be a field with binary operations $+$ (addition) and $\cdot$ (multiplication). A non empty set V is called a vector space over the field $\mathbb{F}$ if there exist two operations, called vector addition $\oplus$ and scalar multiplication $\odot$,

$$\oplus : V \times V \longrightarrow V \qquad \text{and} \qquad \odot : \mathbb{F} \times V \longrightarrow V,$$

such that the following conditions are satisfied.

1. Vector addition is associative, i.e., $v_1 \oplus (v_2 \oplus v_3) = (v_1 \oplus v_2) \oplus v_3$ for all $v_1, v_2, v_3 \in V$;
2. There is a unique vector $0 \in V$, called the zero vector, such that $v \oplus 0 = v = 0 \oplus v$ for all $v \in V$;
3. For each vector $v \in V$ there is a unique vector $-v \in V$ such that $v \oplus (-v) = 0$;
4. Vector addition is commutative, i.e., $v_1 \oplus v_2 = v_2 \oplus v_1$ for all $v_1, v_2 \in V$;
5. $\alpha \odot (v_1 \oplus v_2) = \alpha \odot v_1 \oplus \alpha \odot v_2$ for all $v_1, v_2 \in V$ and $\alpha \in \mathbb{F}$;
6. $(\alpha + \beta) \odot v = \alpha \odot v \oplus \beta \odot v$ for all $v \in V$ and $\alpha, \beta \in \mathbb{F}$;
7. $(\alpha \cdot \beta) \odot v = \alpha \odot (\beta \odot v)$ for all $v \in V$ and $\alpha, \beta \in \mathbb{F}$;
8. $1 \odot v = v$, where 1 is the multiplicative identity of the field $\mathbb{F}$.

zero vector can be non-zero.
eg: V(R+):
addition: u+v = uv
scalar multi: c.u=u^c
u+v=uv => v=1 is additive identity.

If V is a vector space over the field $\mathbb{F}$, we denote it by $V(\mathbb{F})$. The elements of V are called **vectors** and elements of $\mathbb{F}$ are called **scalars**.

Example 2. *1. $\mathbb{R}(\mathbb{R})$, $\mathbb{C}(\mathbb{C})$ and $\mathbb{C}(\mathbb{R})$ are vector spaces under their usual addition and scalar multiplication.*

2. Let $V = \mathbb{F}^n = \{(x_1, \ldots, x_n) \,|\, x_1, \ldots, x_n \in \mathbb{F}\}$. Then V forms a vector space over $\mathbb{F}$ under the following operations:

$$(x_1, \ldots, x_n) \oplus (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

$$\alpha \odot (x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

for all $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in V$ and $\alpha \in \mathbb{F}$.

3. The set of all $m \times n$ matrices $M_{m\times n}(\mathbb{F})$ with entries from the field $\mathbb{F}$ is a vector space over the field $\mathbb{F}$ under the following operations:

$$(a_{ij}) \oplus (b_{ij}) = (a_{ij} + b_{ij}), \text{ and } \alpha \odot (a_{ij}) = (\alpha a_{ij}),$$

for all $A = (a_{ij}), B = (b_{ij}) \in M_{m\times n}(\mathbb{F})$ *and* $\alpha \in \mathbb{F}$.

4. Let X *be a non-empty set. Let* V *be the set of all the functions from* X *to* $\mathbb{R}$. *Then* V *forms a vector space over* $\mathbb{R}$ *under the following operations:* $(f \oplus g)(x) = f(x) + g(x)$ *and* $(\alpha \odot f)(x) = \alpha f(x)$, *for all* $x \in X$, $f, g \in V$, *and* $\alpha \in \mathbb{R}$.

5. Let $P_n = \{a_0 + a_1 x + \ldots + a_n x^n \mid a_0, a_1, \ldots, a_n \in \mathbb{F}\}$. *The set* P_n *forms a vector space over* $\mathbb{F}$ *under the following operations:*

$$(a_0 + a_1 x + \cdots + a_n x^n) \oplus (b_0 + b_1 x + \ldots + b_n x^n) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

$$\alpha \odot (a_0 + a_1 x + \cdots + a_n x^n) = (\alpha a_0 + \alpha a_1 x + \cdots + \alpha a_n x^n)$$

for all $(a_0 + a_1 x + \cdots + a_n x^n), (b_0 + b_1 x + \cdots + b_n x^n) \in P_n$ *and* $\alpha \in \mathbb{F}$.

6. $\mathbb{R}^2$ *over* $\mathbb{R}$ *is not a vector space with respect to the following operations*

(here additive identity = (-1,-1))

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1 + x_2 + 1, y_1 + y_2 + 1)$$
$$\alpha \odot (x, y) = (\alpha x, \alpha y),$$

where $(x_1, y_1), (x_2, y_2), (x, y) \in \mathbb{R}^2$ *and* $\alpha \in \mathbb{R}$. *To see this, we need to find which property is not satisfied. Let* $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ *and* $\alpha \in \mathbb{R}$. *Then*

$$\begin{aligned}\alpha \odot ((x_1, y_1) \oplus (x_2, y_2)) &= \alpha \odot (x_1 + x_2 + 1, y_1 + y_2 + 1)\\ &= (\alpha(x_1 + x_2 + 1), \alpha(y_1 + y_2 + 1))\\ &= (\alpha x_1 + \alpha x_2 + \alpha, \alpha y_1 + \alpha y_2 + \alpha)\\ &\neq \alpha \odot (x_1, y_1) \oplus \alpha \odot (x_2, y_2)\end{aligned}$$

Take $\alpha = 2$ *and* $(x_1, y_1) = (1, 1) = (x_2, y_2)$.

Remark 3. If $\mathbb{F}_1$ is a subfield of $\mathbb{F}$, then $\mathbb{F}(\mathbb{F}_1)$ forms a vector space but converse is not true. For example, $\mathbb{C}(\mathbb{R})$ is a vector space but $\mathbb{R}(\mathbb{C})$ is not a vector space.

Note: If there is no confusion between the operations on a vector space and the operations on the field, we simply write $\oplus$ by $+$ and $\odot$ by $\cdot$.

Theorem 4. Let V be a vector space over $\mathbb{F}$. Then

1. $0 \cdot \boldsymbol{v} = \mathbf{0}$, where 0 and $\mathbf{0}$ are additive identity of $\mathbb{F}$ and V respectively, and $\boldsymbol{v} \in V$.

2. $\alpha \cdot \mathbf{0} = \mathbf{0} \quad \forall \alpha \in \mathbb{F}$.

3. $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$.

4. if $\alpha \in \mathbb{F}$ and $\boldsymbol{v} \in V$ such that $\alpha \cdot \boldsymbol{v} = \mathbf{0}$, then either $\alpha = 0$ or $\boldsymbol{v} = \mathbf{0}$.

Proof: For the first statement, we write $0 = 0 + 0$ so that

$$\begin{aligned} 0 \cdot \boldsymbol{v} &= (0+0) \cdot \boldsymbol{v} \\ 0 \cdot \boldsymbol{v} &= 0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} \quad \text{(Condition 6.)} \\ 0 \cdot \boldsymbol{v} + (-0 \cdot \boldsymbol{v}) &= 0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} + (-0 \cdot \boldsymbol{v}) \quad \text{(using additive inverse)} \\ 0 \cdot \boldsymbol{v} + (-0 \cdot \boldsymbol{v}) &= 0 \cdot \boldsymbol{v} + (0 \cdot \boldsymbol{v} + (-0 \cdot \boldsymbol{v})) \quad \text{(using additive inverse and additive associativity)} \\ \mathbf{0} &= 0 \cdot \boldsymbol{v} + \mathbf{0} = 0 \cdot \boldsymbol{v}. \end{aligned}$$

For the second statement, write $\mathbf{0} = \mathbf{0} + \mathbf{0}$ so that

$$\begin{aligned} \alpha \cdot \mathbf{0} &= \alpha \cdot (\mathbf{0} + \mathbf{0}) \\ \alpha \cdot \mathbf{0} &= \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0} \quad \text{(Condition 5.)} \\ \alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0}) &= \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0}) \quad \text{(using additive inverse)} \\ \alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0}) &= \alpha \cdot \mathbf{0} + (\alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0})) \quad \text{(using additive inverse and additive associativity)} \\ \mathbf{0} &= \alpha \cdot \mathbf{0} + \mathbf{0} = \alpha \cdot \mathbf{0}. \end{aligned}$$

For the third statement, we write $0 = (-1) + 1$ so that

$$\begin{aligned} 0 \cdot \boldsymbol{v} &= ((-1) + (1)) \cdot \boldsymbol{v} \\ 0 \cdot \boldsymbol{v} &= (-1) \cdot \boldsymbol{v} + 1 \cdot \boldsymbol{v} \quad \text{(using Condition 5.)} \\ \mathbf{0} &= (-1) \cdot \boldsymbol{v} + \boldsymbol{v} \quad \text{(using the first statement and Condition 8.)} \\ \mathbf{0} + (-\boldsymbol{v}) &= (-1) \cdot \boldsymbol{v} + (\boldsymbol{v} + (-\boldsymbol{v})) \quad \text{(using additive inverse and associativity)} \\ -\boldsymbol{v} &= (-1) \cdot \boldsymbol{v} + \mathbf{0} = (-1) \cdot \boldsymbol{v}. \end{aligned}$$

Prove the fourth statement yourself.

Definition 5. Let V be a vector space over the field $\mathbb{F}$. A subspace of V is a non-empty subset W of V which is itself a vector space over $\mathbb{F}$ with the operations of vector addition and scalar multiplication on V.

Example The subsets $\{\mathbf{0}\}$ and V are subspaces of a vector space V. These subspaces are called trivial subspaces of V.

Theorem 6. Let V be a vector space over the field $\mathbb{F}$ and $W \subseteq V$. Then W is subspace of V if and only if $\alpha w_1 + \beta w_2 \in W$, for all $w_1, w_2 \in W$ and $\alpha, \beta \in \mathbb{F}$.

(µw1 = w1', тw2 = w2' => w1' + w2' belongs to W by definition of vector space)

Proof: Direct part follows from the definition of subspace. Conversely, if $\alpha = 1$ and $\beta = 1$, then we see that $w_1 + w_2 \in W$ $\forall w_1, w_2 \in W$, also if $\beta = 0$, then $\alpha w_1 \in V$ $\forall \alpha \in \mathbb{F}$ and $w_1 \in W$. Thus, W is closed under vector addition and scalar multiplication. Further, let $\alpha = \beta = -1$ and $w_1 = w_2$. Then $0 \in W$, *i.e.*, zero vector of V lies in W. The rest of the properties trivially true as the elements are from vector space V. Thus, W is a vector space over $\mathbb{F}$. $\square$

Example 7. 1. A line passing through origin is a subspace of $\mathbb{R}^2$ over $\mathbb{R}$.

2. Let A be an $m \times n$ matrix over $\mathbb{F}$. Then the set of all $n \times 1$ (column) matrices x over $\mathbb{F}$ such that $Ax = 0$ is a subspace of the space of all $n \times 1$ matrices over $\mathbb{F}$ or $\mathbb{F}^n$. To see this we need to show that $A(\alpha x + y) = 0$, when $Ax = 0, Ay = 0$, and α is an arbitrary scalar in $\mathbb{F}$.

3. The solution set of a system of non-homogeneous linear equations is not a subspace of $\mathbb{F}^n$ over $\mathbb{F}$.
(0 vector does not belong to the solution set)

4. The collection of polynomial of degree less than or equal to n over $\mathbb{R}$ with the constant term 0 forms a subspace of the space of polynomials of degree less than or equal to n.

5. The collection of polynomial of degree n over $\mathbb{R}$ is not a subspace of the space of polynomials of degree less than or equal to n.

Theorem 8. Let W_1 and W_2 be subspaces of a vector space V over $\mathbb{F}$. Then $W_1 \cap W_2$ is a subspace of V.

Proof: Since W_1 and W_2 are subspaces, $0 \in W_1 \cap W_2$ so that $W_1 \cap W_2$ is a non-empty set. Let $w, w' \in W_1 \cap W_2$ and $\alpha, \beta \in \mathbb{F}$. Then $\alpha w + \beta w' \in W_1$ as W_1 is a subspace of V and $w, w' \in W_1$. Similarly, $\alpha w + \beta w' \in W_2$. Thus, $\alpha w + \beta w' \in W_1 \cap W_2$. By Theorem 6, $W_1 \cap W_2$ is a subspace of V.

Remark 9. The above theorem can be generalized for any number of subspaces. However, the union of two subspaces need not be a subspace. Let $V = \mathbb{R}^2$, $W = X$-axis and $W' = Y$-axis. Then $(1,0) \in W$ and $(0,1) \in W'$ but $(1,0) + (0,1) = (1,1) \notin W \cup W'$. The union of two subspaces is a subspace if one is contained in other.

Lecture 7

Linear Combination, Linear Span, Linear Dependence & Independence

Definition 1. *Let V be a vector space over a field $\mathbb{F}$. A vector $v \in V$ is said to be a* **linear combination of the vectors** $v_1, v_2, \ldots, v_k \in V$ *if there exist scalars* $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{F}$ *such that*

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k.$$

Example 2. *1. Consider the vector space $\mathbb{R}^2$ over $\mathbb{R}$. Let $v_1 = (1,0), v_2(0,1) \in \mathbb{R}^2$. and $(x,y) \in \mathbb{R}^2$. Then every vector (x,y) in $\mathbb{R}^2$ is a linear combination of v_1 and v_2 as $(x,y) = x(1,0) + y(0,1)$.*

2. Let $\mathbb{R}^3(\mathbb{R})$ and $(1,1,1), (1,1,-1) \in \mathbb{R}^3$. Then $(1,1,2)$ is a linear combination of $(1,1,1)$ and $(1,1,-1)$ as $(1,1,2) = \frac{-1}{2}(1,1,1) + \frac{3}{2}(1,1,-1)$. But $(1,-1,0)$ is not a linear combination of $(1,1,1)$ and $(1,1,-1)$. (Verify yourself!)

Definition 3. Let V be a vector space over the field $\mathbb{F}$ and $S \subseteq V$. Then a vector $v \in V$ is said to be a **linear combination of vectors in** S if there exist a positive integer k and scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$ in $\mathbb{F}$ such that $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$, where $v_i \in S$.

Example 4. *Consider the vector space $P(\mathbb{R})$ over $\mathbb{R}$. Let $S = \{1, x, x^2, x^3, \ldots\}$. Then every polynomial in $P(\mathbb{R})$ is a linear combination of vectors in S.*

Definition 5. Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$. Then **linear span of** S, denoted as $L(S)$ or $[S]$, is a subset of V defined as $L(S) = \{\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k \,|\, v_i \in S, \alpha_i \in \mathbb{F}\}$. For instance, $L(\{(1,0),(0,1)\}) = \mathbb{R}^2$ and $L(\{(1,1,1),(1,1,-1)\}) = \{((a,a,b)) \mid a, b \in \mathbb{R}\}$. ***if S is subset of V, then L(S) is also subset of V***

Theorem 6. *Let S be a non empty subset of a vector space V over $\mathbb{F}$. Then $L(S)$ is the smallest subspace containing S.*

Proof: Let $v \in S$. Then $1.v \in L(S)$ so that S is contained in $L(S)$. Next, we show that $L(S)$ is a subspace of V. Let $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$ and $v' = \beta_1 v_1' + \beta_2 v_2' + \ldots + \beta_l v_l'$ belong to $L(S)$. Then for any scalars γ, δ, $\gamma v + \delta v' = \gamma\alpha_1 v_1 + \gamma\alpha_2 v_2 + \ldots + \gamma\alpha_k v_k + \delta\beta_1 v_1' + \delta\beta_2 v_2' + \ldots + \delta\beta_l v_l' \in L(S)$. Thus $L(S)$ is a subspace of V.

Now to show that $L(S)$ is the smallest subspace containing S, it is enough to show that $L(S)$ is a subset of any subspace containing S. Let T be a subspace of V which contains S and $v \in L(S)$. Then $v = \sum_{i=1}^{k} \alpha_i v_i$ for $\alpha_i \in \mathbb{F}$ and $v_i \in S$. Note that $v_i \in S$ implies $v_i \in T$ and hence $v \in T$ as T is a subspace. (for all vi in T, v = sum(alphai*vi) belongs to T as T is a subspace => L(S) subset of T). □
(theorem 6 of prev lecture).

Definition 7. *Let S be a set of vectors in a vector space V over $\mathbb{F}$. The* **subspace spanned by** S*, denoted as $\langle S \rangle$, is defined to be the intersection of all subspaces of V which contain S.*

Theorem 8. $L(S) = \langle S \rangle$. (intersection of all subspaces containing S is L(S) as L(S) is the smallest subspace containing S).

Definition 9. **The sum** $S_1 + S_2$ **of two subsets** S_1, S_2 *of a vector space* V *over* $\mathbb{F}$ *is given by*

$$S_1 + S_2 = \{v_1 + v_2 \mid v_1 \in S_1, v_2 \in S_2\}.$$

Theorem 10. Let V be a vector space over $\mathbb{F}$ and U and W be two subspaces of V. Then

1. $U + W$ is a subspace of V;
2. $U + W = L(U \cup W)$.

Proof: Let $v, v' \in U + W$. The $v = u + w$ and$v' = u' + w'$ for some $u, u' \in U$ and $w, w' \in W$. Let $\alpha, \beta \in \mathbb{F}$. Then $\alpha v + \beta v' = (\alpha u + \beta u') + (\alpha w + \beta w') \in U + W$. Therefore, $U + W$ is a subspace of V.

(belongs to U) + (belongs to W) => belongs to (U+W)

(thm 6 with T=U+W, S = U U W)

Note that $U + W$ is a subspace of V containing $U \cup W$. Hence, $L(U \cup W) \subseteq U + W$. Now suppose $v \in U + W$. Then $v = u + w$, where $u \in U$ and $w \in W$. Note that $u, w \in U \cup W$ and hence, $u + w \in L(U \cup W)$. Therefore, $U + W \subseteq L(U \cup W)$.

Definition 11. Let V be a vector space over $\mathbb{F}$. A subset S of V is said to be **linearly dependent** (LD) if there exist distinct vectors $v_1, v_2, \ldots, v_n \in S$, and scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$, not all of which are 0, such that $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$.
A set which is not linearly dependent is called **linearly independent**.

Let $S = \{v_1, v_2, \ldots, v_k\}$. Then $v_1, v_2, \ldots, v_k$ are said to be linearly dependent if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{F}$, not all of which are 0, such that $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$.

The vectors $v_1, v_2, \ldots, v_k$ are not linearly dependent, that is, linearly independent if $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$ implies $\alpha_i = 0$ for all $i = 1, 2, \ldots, k$.

Example 12. *1. Consider the vector space* R^3 *over* $\mathbb{R}$. *The set* $S = \{(n, n, n) \mid n \in \mathbb{N}\}$ *is linearly dependent since* $(2, 2, 2), (3, 3, 3) \in$ *and* $3(2, 2, 2) - 2(3, 3, 3) = 0$ *so that* S *is linearly dependent.*

2. The $S = \{(1, 2, 3), (2, 3, 4), (1, 1, 2)\}$ *is linearly independent in* $\mathbb{R}^3(\mathbb{R})$. *To see this consider* $\alpha_1(1, 2, 3) + \alpha_2(2, 3, 4) + \alpha_3(1, 1, 2) = 0$. *Then* $(\alpha_1 + 2\alpha_2 + \alpha_3, 2\alpha_1 + 3\alpha_2 + \alpha_3, 3\alpha_1 + 4\alpha_2 + 2\alpha_3) = (0, 0, 0)$. *Thus,* $\alpha_1 + 2\alpha_2 + \alpha_3 = 0, 2\alpha_1 + 3\alpha_2 + \alpha_3 = 0, 3\alpha_1 + 4\alpha_2 + 2\alpha_3 = 0$. *By solving this system of linear equations, we see that* $\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0$ *is the only possible solution.*

3. Observe that $1.0 = 0$. *Thus, any subset of a vector space containing the zero vector is linearly dependent.* (cuz the coeff of the 0 vector can be non zero, making it lin. dependent)

4. The set $\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\} \subseteq \mathbb{R}^n$ *is linearly independent.*

5. Let $V = \{f \mid f : [-1, 1] \to \mathbb{R}\}$. *The set* $\{x, |x|\}$ *is linearly independent. To see this, consider the equation* $\alpha x + \beta|x| = 0$. *A function is zero if it is zero at every point of the domain. Thus,* $\alpha x + \beta|x| = 0$

for all $x \in [-1, 1]$. *If* $x = 1$ *we get* $\alpha + \beta = 0$ *and if* $x = -1$, $\alpha - \beta = 0$. *Solving these two equations we get* $\alpha = \beta = 0$. *Thus the set is linearly independent.*

Remark 13. Let V be a vector space over $\mathbb{F}$. Then

1. the set $\{v\}$ is L.D. if and only if $v = 0$;
2. a subset of a linearly independent set is also linearly independent;
3. a set containing a linearly dependent set is also linearly dependent.

(in the bigger set, the coeffs of extra elements = 0, and there will be atleast one non zero coeff in the original equation. so linearly dependent)

Lecture 8

Basis & Dimension

Definition 1. Let V be a vector space over a field $\mathbb{F}$. A subset S of V is said to be a **basis** of V if the following conditions are satisfied.

1. S is linearly independent set.

2. The linear span $L(S)$ is the vector space V, that is, $L(S) = V$.

Example 2. *1. Let $V = \mathbb{F}^n(\mathbb{F})$ and $B = \{e_1, e_2, \ldots, e_n\}$, where $e_i = (0, 0, \ldots, 1, \ldots, 0)$, (1 at the i-th component and 0 otherwise). The set B is a basis of $\mathbb{F}^n$ and is called* **the standard basis of** $\mathbb{F}^n(\mathbb{F})$.

2. The set $B = \{e_{ij} \in M_{m\times n} \mid i, j\text{-th entry of } e_{ij} \text{ is 1 and 0 otherwise}\}$ is the standard basis of $M_{m\times n}(\mathbb{F})$ over $\mathbb{F}$.

3. The set $B = \{1, x, x^2, x^3, \ldots, x^n\}$ is the standard basis of $P_n(\mathbb{R})$ over $\mathbb{R}$.

4. The set $B = \{1, x, x^2, x^3 \ldots\}$ is the standard basis of $\mathbb{R}[x]$ over $\mathbb{R}$.

Remark 3. *1. Every vector space has a basis.*

2. The basis of the zero space is the empty set $\emptyset$.

3. A basis of a vector space need not be unique, for instance, $\{(1, 1), (1, -1)\}$ is also a basis of $\mathbb{R}^2(\mathbb{R})$.

4. A vector space V is called a **finite dimensional vector space** *if it has a finite basis, otherwise it is called an* **infinite dimensional space**.

Theorem 4. Let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis of a vector space V over $\mathbb{F}$. If $B' = \{w_1, w_2, \ldots, w_m\}$, where $(m > n)$. Then B' is a linearly dependent set.

Proof: We will show that there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{F}$, not all of which are 0, such that $\alpha_1 w_1 + \alpha_2 w_2 + \ldots + \alpha_m w_m = 0$. Since B is a basis of V, we can write an element of B' as linear combination of elements of B over $\mathbb{F}$ so that

$$w_1 = a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n$$

$$w_2 = a_{21}v_1 + a_{22}v_2 + \ldots + a_{2n}v_n$$

$$\vdots$$

$$w_m = a_{m1}v_1 + a_{m2}v_2 + \ldots + a_{mn}v_n$$

Therefore, $\alpha_1(a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n) + \alpha_2(a_{21}v_1 + a_{22}v_2 + \ldots + a_{2n}v_n) + \cdots + \alpha_m(a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n) = 0$. Equivalently,

$(\alpha_1 a_{11} + \alpha_2 a_{21} + \ldots + \alpha_m a_{m1})v_1 + (\alpha_1 a_{12} + \alpha_2 a_{22} + \ldots + \alpha_m a_{m2})v_2 + \ldots + (\alpha_1 a_{1n} + \alpha_2 a_{2n} + \ldots + \alpha_m a_{mn})v_n = 0.$
Since $\{v_1, v_2, \ldots, v_n\}$ is a basis of V, we get

$$\begin{aligned}
\alpha_1 a_{11} + \alpha_2 a_{21} + \ldots + \alpha_m a_{m1} &= 0 \\
\alpha_1 a_{12} + \alpha_2 a_{22} + \ldots + \alpha_m a_{m2} &= 0 \\
&\vdots \\
\alpha_1 a_{1n} + \alpha_2 a_{2n} + \ldots + \alpha_m a_{mn} &= 0
\end{aligned}$$

(in RRE Form, number of columns is greater than number of rows. So, there will be atleast one free variable which can have any value. So, there will be infinite values of that free variable and hence, there will be non zero solution). (here solution is the value of the coeffs (alphas)).

Above is a homogeneous system of n equations in m unknowns with $m > n$, therefore the system has a non-zero solution. Thus, $\alpha_i \neq 0$ for some i so that B' is linearly dependent. $\square$

(non zero solution => atleast one variable of the solution is non zero)

Corollary 5. *Let $V(\mathbb{F})$ be a finite dimensional vector space. The any two bases of V have the same number of elements.*

Definition 6. Let $V(\mathbb{F})$ be a finite dimensional vector space. Then the number of elements in a basis of V is called **dimension** of V and it is denoted as $\dim(V)$.

Example 7. *1.* $\dim(\mathbb{F}^n(\mathbb{F})) = n$; *2.* $\dim(\mathbb{C}(\mathbb{R})) = 2$; *3.* $\dim(M_{m\times n}(\mathbb{F})) = mn$; *4.* $\dim P_n(\mathbb{R}) = n + 1$; *5.* $\mathbb{R}[x]$ *is an infinite dimensional space.*

Remark 8. *Let V be a finite-dimensional vector space and let $n = \dim V$. Then*

1. *any subset of V which contains more than n vectors is linearly dependent;* (basically theorem 4)
2. *no subset of V which contains fewer than n vectors can span V.*

Theorem 9. Let $\{v_1, \ldots, v_n\}$ be a basis for a vector space V. Then each vector in V can be expressed uniquely as a linear combination of the basis vectors.

Proof: Let $v \in V$ and $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{F}$ such that $v = \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} \beta_i v_i$. Then $\sum_{i=1}^{n} (\alpha_i - \beta_i)v_i = 0$, but v_i' s are linearly independent so that $\alpha_i - \beta_i = 0$ for each i. Therefore, each vector in V can be expressed uniquely as a linear combination of the basis vectors.

Extension Theorem

Theorem 10. Let $S = \{v_1, \ldots, v_n\}$ be a linearly independent subset of a vector space V. If $v \notin L(S)$, then $S \cup \{v\}$ is linearly independent.

Proof: Consider the set $S' = S \cup \{v\}$. Let $\alpha v + \alpha_1 v_1 + \ldots + \alpha_n v_n = 0$. It is enough to show that $\alpha = 0$. If $\alpha \neq 0$, then $v \in L(S)$, which is not true. Hence, $\alpha = 0$ so that S' is L.I. $\square$

then v can be written as linear combination of v1,v2...vn (div by alpha)

Theorem 11. Let V be an n dimensional vector space. Then

1. a linearly independent set of n vectors of V is a basis of V; (=> do not need to check that L(S) = V)
2. a set of n vectors of V which spans V is a basis of V. (=> do not need to check linear independence)

Proof: Let $S = \{v_1, v_2, \ldots, v_n\} \subset V$ be a linearly independent set. It is enough to show that $L(S) = V$. Suppose it is not true and $v \in V \setminus S$. V\L(S)?? Then the set $S \cup \{v\}$ is L.I. which contradicts the fact that $\dim V = n$. Thus $v \in L(S)$. Therefore S is a basis of V. (CHECK END OF PG)

Let $S = \{v_1, v_2, \ldots, v_n\} \subset V$ and $L(S) = V$. Suppose S is linearly dependent. Then there exist i such that v_i a is linear combination of rest of the vectors in S. Therefore, the set $S \setminus \{v_i\}$ spans V having $n-1$ vectors so that $\dim V \leq n-1$ which contradicts the fact that $\dim V = n$. $\square$

Example 12. *Find a basis and dimension of the solution space of the homogeneous system $Ax = 0$, where* $A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & -1 \\ 3 & 11 & 5 \end{pmatrix}$.

Solution: *The RRE form of A is* $\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. *The solution set is $\{(2z, -z, z) \mid z \in \mathbb{R}\}$. Any solution is a linear combination of $(2, -1, 1)$ and a singleton set with a non-zero element is linearly independent. Thus $\{(2, -1, 1)\}$ is a basis of the solution space of $Ax = 0$. Hence, the dimension of the solution space is 1.*

Thm 11 Part 1) Assume that dim(V) = n. Let S be a linearly independent set having cardinality n.
Proof by Contradiction: Assume that L(S) != V. Then we can have an element v in (V-L(S)), which implies v does not belong to L(S).
Now if v doesnt belong to L(S), and S is linearly independent, we have S union v is linearly independent (Thm 10.)
From Thm 4., if a set has more than dimV elements, the set is linearly dependent. So, (S union v) has <= n elements.
=> n+1 <= n => 1<=0 which is false. Hence our assumption was false. Hence, L(S) = V.

Thm 11 Part 2) Assume that dim(V) = n. Let S be a set having cardinality n such that L(S)=V.
Proof by Contradiction: Assume that S is linearly dependent. Then some vector of S can be written as a linear combination of the other vectors in S. So, this means that any linear combination of the first n vectors can be written as a linear combination of the first n-1 vectors (assuming last vector in S (= x) was the one causing the dependency.) This means that S\{x} spans V cuz x didnt make a difference in spanning anyway, and S\{x} forms a basis. This would mean that dim(V) (based on the new S\{x}) <= n-1 (<= n-1 since there can be multiple elements in S that are linearly dependent. So we have to remove all those elements). Hence our assumption was false. Hence, S is linearly independent. }

theorems related to dimensions

Lecture 9

Basis & Dimension of Direct Sum of Subspaces

Theorem 1. *If W is a subspace of a finite dimensional vector space V, every linearly independent subset of W is finite and it is a part of a basis for W.*

(basically, every linearly independent subset of W can be extended to a basis for W).

let subset be S. S union v is also linearly independent, for some v belonging to V\L(S). keep adding such v's to S till basis is obtained

We say that W is a proper subspace of a vector space V if $W \neq \{0\}$ and $W \neq V$.

Theorem 2. *If W is a proper subspace of a finite-dimensional vector space V, then W is finite-dimensional and* $\dim W < \dim V$.

can keep extending {w} (linearly independent) until we obtain B.

Proof: Since W is not the zero space, then $\exists w \in W$ such that $w \neq 0$. There is a basis B of W containing w. Note that B can have at most n vectors as V is n dimensional. Hence W is finite-dimensional, and $\dim W \leq \dim V$. Since W is a proper subspace, there is a vector v in V which is not in W. Adjoining v to B, we obtain a linearly independent subset of V. Thus $\dim W < \dim V$.

Theorem 3. If W_1 and W_2 are two subspaces of a finite dimensional vector space V, then $W_1 + W_2$ is finite dimensional and $\dim(W_1 + W_2) = \dim(W_1) + dim(W_2) - dim(W_1 \cap W_2)$.

Proof: Since $W_1 \cap W_2$ is a subspace of W_1 as well as of W_2, it is finite dimensional. If $B_0 = \{w_1, \ldots, w_k\}$ is a basis of $W_1 \cap W_2$, then B_0 can be extended to a basis for W_1 as well as of W_2. Let $B_1 = \{w_1, \ldots, w_k, v_1, \ldots, v_l\}$ and $B_2 = \{w_1, \ldots, w_k, u_1, \ldots, u_m\}$ be bases of W_1 and W_2 respectively. We claim that the set $B = B_0 \cup B_1 \cup B_2 = \{w_1, \ldots, w_k, v_1, \ldots, v_l, u_1, \ldots, u_m\}$ forms a basis of the subspace $W_1 + W_2$. Clearly, $L(B) = W_1 + W_2$. We need to show that B is a linearly independent set. Let $\sum_{i=1}^{k} \alpha_i w_i + \sum_{j=1}^{l} \beta_j v_j + \sum_{r=1}^{m} \gamma_r u_r = 0$, where $\alpha_i, \beta_i, \gamma_i \in \mathbb{F}$. Then

$$\sum_{i=1}^{k} \alpha_i w_i + \sum_{j=1}^{l} \beta_j v_j = -\sum_{r=1}^{m} \gamma_r u_r$$

so that $-\sum_{k=1}^{m} \gamma_k u_k = W_1 \cap W_2$ (as RHS is in W_2 and LHS is in W_1). Therefore,

$$-\sum_{r=1}^{m} \gamma_r u_r = \sum_{i=1}^{k} \delta_i w_i$$

so that $\sum_{r=1}^{m} \gamma_r u_r + \sum_{i=1}^{k} \delta_i w_i = 0$. But $\{w_1, \ldots, w_k, u_1, \ldots, u_m\}$ is a basis of W_2, therefore $\gamma_r = 0$ for $1 \leq r \leq m$. This further implies that $\alpha_i = \beta_j = 0$. Thus, the set B forms a basis for $W_1 + W_2$. $\square$

Corollary 4. *Let W_1, W_2 be subspaces of V. Then*

$$\dim W_1 + \dim W_2 - \dim V \leq \dim(W_1 \cap W_2) \leq \min\{\dim W_1, \dim W_2\}.$$

Definition 5. *Let W_1 and W_2 be subspaces of a vector space V. The vector space V is called the* **direct sum** *of W_1 and W_2, denoted as $W_1 \oplus W_2$, if every element $v \in V$ can be uniquely represented as $v = w_1 + w_2$, where $w_1 \in W_1$ and $w_2 \in W_2$.*

Theorem 6. *A vector space $V(\mathbb{F})$ is the direct sum of its subspaces W_1 and W_2 if and only if $V = W_1 + W_2$, and $W_1 \cap W_2 = \{0\}$.*

Proof: Let $V = W_1 \oplus W_2$. Since every elements $v \in V$, $v = w_1 + w_2$, where $w_1 \in W_1$ and $w_2 \in W_2$. Thus, $W_1 + W_2 = V$. Let $x \in W_1 \cap W_2$. Then $x = x + 0$ and $x = 0 + x$. But x must have a unique representation, therefore $x = 0$.

Conversely, let $V = W_1 + W_2$ and $W_1 \cap W_2 = \{0\}$. Suppose $v \in V$ has more than one representation, *i.e.*, $v = w_1 + w_2 = w_1' + w_2'$. This implies $w_1 - w_1' = w_2 - w_2' \in W_1 \cap W_2 = \{0\}$. Thus $w_1 = w_1'$ and $w_2 = w_2'$. This follows the proof. $\square$

Corollary 7. $\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2$.

Example 8. *Let $V = \mathbb{R}^2(\mathbb{R})$ and $W_1 = \{(x, 2x) \mid x \in \mathbb{R}\}$, $W_2 = \{(x, 3x) \mid x \in \mathbb{R}\}$ be subspaces of V. Then $V = W_1 \oplus W_2$.*

Note that, $(x, y) = (3x - y, 2(3x - y)) + (y - 2x, 3(y - 2x))$. Let $(x, y) \in W_1 \cap W_2$ then $(x, y) = (a, 2a) = (b, 3b)$ for some $a, b \in \mathbb{R}$. Then $(x, y) = (0, 0)$ so that $W_1 \cap W_2 = \{0\}$.

Lecture 10

Linear Transformation

Definition 1. Let V and W be vector spaces over field $\mathbb{F}$. A map $T : V \to W$ is said to be a linear map (or linear transformation) if for $\forall\ \alpha \in \mathbb{F}$ and $\forall\ v_1, v_2 \in V$ we have:

(i) $T(v_1 + v_2) = T(v_1) + T(v_2)$, (ii) $T(\alpha v) = \alpha T(v)$.

Example 2. *1. The map $T : V \to W$ defined by $T(v) = 0$ for all $v \in V$, is linear (the zero map).*

2. The map $T : V \to V$ defined by $T(v) = v$ for all $v \in V$, is linear (the identity map).

3. Let $m \leq n$. Then a map $T : \mathbb{R}^m \to \mathbb{R}^n$, defined by $T(x_1, x_2, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$, $(n - m)$ zeroes, is linear (the inclusion map).

4. Let $m \geq n$. Then a map $T : \mathbb{R}^m \to \mathbb{R}^n$ defined by $T(x_1, x_2, \ldots, x_m) = (x_1, x_2, \ldots, x_n)$, is linear (the projection map).

5. A map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1, x_2) = (x_1, -x_2)$, is linear (reflection along x-axis).

6. A map $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T_\theta(x, y) = (x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta)$, is linear (rotation about origin with angle θ.

(xcos - ysin, xsin + ycos)

7. Let A be a matrix of order $m \times n$. Then A defines a linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T_A(x) = Ax$.

8. Let $D : \mathbb{R}[x] \to \mathbb{R}[x]$ defined by $D(f(x)) = \frac{d}{dx} f(x)$. Then D is linear (differentiation map).

Proposition 3. Let $T : V \to W$ be a linear map. Then

(i) $T(0) = 0$; (ii) $T(-v) = -T(v)$; (iii) $T(v_1 - v_2) = T(v_1) - T(v_2)$.

Definition 4. Let $T : V \to W$ be a linear map. Then the null space (or kernel) of $T = \{v \in V : T(v) = 0\}$, denoted as $\ker(T)$ and Range space (or Image) of $T = \{T(v) : v \in V\}$ denoted as Range(T).

Example 5. *1. If $T : V \longrightarrow W$ is the zero map, then $ker(T) = V$ and $Range(T) = \{0\}$.*

2. If $T : V \longrightarrow V$ is the identity map, then $ker(T) = \{0\}$ and $Range(T) = V$.

3. If $T : P_n(\mathbb{R}) \longrightarrow P_n(\mathbb{R})$ defined by $T(f(x)) = \frac{d}{dx}(f(x))$, then $ker(T)$ contains all constant polynomials and $Range(T) = P_{n-1}(\mathbb{R})$.

Theorem 6. Let $T : V \to W$ be a linear map. Then $ker(T)$ and $Range(T)$ are subspaces of V and W respectively. (Prove it yourself!)

Definition 7. The dimension of null space Ker(T) is called the **nullity** of T and the dimension of the range space Range(T) of T is called the **rank** of T.

Theorem 8. *Let V be a finite-dimensional vector space over the field $\mathbb{F}$ and let $\{v_1, \ldots, v_n\}$ be a basis for V. Let W be a vector space over the same field $\mathbb{F}$ and let $w_1, w_2, \ldots, w_n$ be any vectors in W. Then there is precisely one linear transformation T from V to W such that $T(v_i) = w_i \ \forall i = 1, \ldots, n$, and it is given by $T(v) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$, where $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$.*

Theorem 9. Let $T : V \to W$ be a linear map and $B = \{v_1, v_2, \ldots, v_n\}$ be a basis for V. Then the T is completely determined by its images on basis elements and Range$(T) = L(\{T(v_1), T(v_2), \ldots, T(v_n)\})$.

Proof: Let $v \in V$. Then $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some $\alpha_i \in \mathbb{F}$ and $i = 1, \ldots, n$. The map T is linear, $T(v) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$, that is, image of any vector is a linear combination of images of basis vectors. Thus, Range$(T) = \{T(v) : v \in V\} = \{\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) : v_1, v_2, \ldots, v_n \in B, \ \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}\} = L(\{T(v_1), T(v_2), \ldots, T(v_n)\})$.

Corollary 10. (Riesz Representation Theorem) Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a linear map. Then there exist $a \in \mathbb{R}^n$ such that $T(x) = a^t x$.

Proof: Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Then $T(x) = T(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n x_i T(e_i)$. Let $T(e_i) = a_i$. Thus $T(x) = a^T x$, where $a = (a_1, \ldots, a_n)$. □

Example 11. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (x + y - z, x - y + z, y - z)$. The null space of T is $\{(x, y, z) : x + y - z = 0, x - y + z = 0, y - z = 0\}$ which is the solution space of a homogeneous system of linear equations. Thus, $\ker(T) = \{(x, y, z) : x = 0, y = z, z \in \mathbb{R}\} = \{(0, t, t) : t \in \mathbb{R}\} = L(\{(0, 1, 1)\})$. Thus basis of ker(T) is $\{(0, 1, 1)\}$ (as non-zero singleton is independent) so that Nullity$(T) = 1$. Range$(T) = L(\{T(e_1), T(e_2), T(e_3)\}) = L(\{(1, 1, 0), (1, -1, 1), (-1, 1, -1)\}) = L(\{(1, 1, 0), (1, -1, 1)\}) = \{\alpha(1, 1, 0) + \beta(1, -1, 1) \mid \alpha, \beta \in \mathbb{R}\} = \{(\alpha + \beta, \alpha - \beta, \beta) \mid \alpha, \beta \in \mathbb{R}\}$. Note that Range of T is linear span of $\{(1, 1, 0), (1, -1, 1)\}$ which is linearly independent so that Rank(T) is 2.

Lecture 11

Rank-Nullity theorem & Vector Space Isomorphism

Theorem 1. Rank-Nullity Theorem: Let V and W be vector spaces over the field $\mathbb{F}$ and let $T : V \to W$ be a linear map. If V is finite dimensional then, $nullity(T) + rank(T) = \dim(V)$.

Proof: Since $Ker(T)$ is a subspace of V, its dimension is finite, say n. Let $B = \{v_1, \ldots, v_n\}$ be a basis for $Ker(T)$. Then B can be enlarged to form a basis for V. Let $B' = \{v_1, \ldots, v_n, v_{n+1}, \ldots, v_m\}$ be a basis for V. Now claim that the set $S = \{T(v_{n+1}), \ldots, T(v_m)\}$ forms a basis for $Range(T)$. Let $v \in V$. Then $v = \alpha_1 v_1 + \ldots + \alpha_m v_m$, this implies $T(v) = \alpha_{n+1}T(v_{n+1}) + \ldots + \alpha_m T(v_m)$. Thus $L(S) = Range(T)$. To show that S is linearly independent, assume that $\alpha_{n+1}T(v_{n+1}) + \ldots + \alpha_m T(v_m) = 0$. Then $T(\alpha_{n+1}v_{n+1} + \ldots + \alpha_m v_m) = 0$ so that $\alpha_{n+1}v_{n+1} + \ldots + \alpha_m v_m \in Ker(T)$. Therefore, $\alpha_{n+1}v_{n+1} + \ldots + \alpha_m v_m = \beta_1 v_1 + \ldots + \beta_n v_n$ or $\sum_{i=1}^{n} \beta_i v_i + \sum_{i=n+1}^{m} \alpha_i v_i = 0$. But B' is a basis for V. Therefore, $\alpha_i = 0$ and hence, S is linearly independent. □

Recall that a function $f : X \to Y$ is invertible if there exits a function $g : Y \to X$ such that $f \circ g = I_Y$ and $g \circ f = I_X$. Furthermore, a function$f$ is invertible if and only if it is one-one and onto, and the inverse function g is given by $g(y) = f^{-1}(y)$.

Theorem 2. *Let $T : V \to W$ be a linear map. If T is invertible, then the inverse map T^{-1} is linear.*

Proof: Suppose $T : V \longrightarrow W$ is invertible. Then T is one-one and onto. Let T^{-1} denote the inverse of T. We want to show that $T^{-1}(\alpha w_1 + \beta w_2) = \alpha T^{-1}(w_1) + \beta T^{-1}(w_2)$. Let $T^{-1}(w_1) = v_1$ and $T^{-1}(w_2) = v_2$. Then $T(\alpha v_1 + \beta v_2) = \alpha w_1 + \beta w_2$. Since T is one-one, $T^{-1}(\alpha w_1 + \beta w_2) = \alpha v_1 + \beta v_2 = \alpha T^{-1}(w_1) + \beta T^{-1}(w_2)$.

Definition 3. A linear map $T : V \to W$ is said to be non-singular if Ker(T) $= \{0\}$.

Theorem 4. A linear map $T : V \to W$ is non-singular if and only if T is one-one.

Proof: Let T is non-singular. If $T(x) = T(y)$, then $T(x - y) = 0$. This implies $x - y \in Ker(T) = \{0\}$. So $x = y$. Conversely, let $x \in Ker(T)$. Then $T(x) = 0 = T(0)$, as T is one one. So $x = 0$. □

Theorem 5. Let V and W be finite-dimensional vector spaces over the field $\mathbb{F}$ such that $\dim V = \dim W$. If T is a linear transformation from V to W, the following are equivalent:

(i) T is invertible.

(ii) T is non-singular.

(iii) T is onto, that is, the range of T is W.

Definition 6. Let V and W be vector spaces over the field $\mathbb{F}$. An invertible linear transformation from V to W is called an isomorphism. If there exists an isomorphism from V to W, we say that V and W are isomorphic.

Exercise 1. *Show that isomorphism is an equivalence relation on finite dimensional vector spaces over the field* $\mathbb{F}$.

Example 7. *Show that* $\mathbb{R}^2(\mathbb{R})$ *and* $\mathbb{C}(\mathbb{R})$ *are isomorphic.*

Solution: Define $T : \mathbb{R}^2 \to \mathbb{C}$ as $T(x, y) = x + iy$. Then T is linear and $Ker(T) = \{(x, y) \in \mathbb{R}^2 \mid x + iy = 0 + 0i\} = \{(0, 0)\}$. Hence, T is one-one. Note that $\dim \mathbb{R}^2 = \dim \mathbb{C} = 2$ over $\mathbb{R}$. By rank-nullity theorem, the map is onto.

Definition 8. Let V be a vector space of dimension n. A basis B is called an ordered basis if there is an one to one map between B and the set $\{1, \ldots, n\}$. In simple words, a basis B with an ordering of the elements (of B) is called an ordered basis.

Definition 9. Let V be a vector space with an ordered basis $B = \{v_1, v_2, \ldots, v_n\}$ over the field $\mathbb{F}$. Then for any $v \in V$ there exists a unique $(a_1, a_2, \ldots, a_n) \in \mathbb{F}^n$ such that $v = a_1v_1 + a_2v_2 + \ldots + a_nv_n$. Then the column vector $(a_1, \ldots, a_n)^T$, denoted as $[v]_B$, is called the **coordinate vector** of v with respect to the basis B.

For example, in $\mathbb{F}^n$ the coordinate vector of $(x_1, x_2, \ldots, x_n)$ with respect to the standard basis $\{e_1, \ldots, e_n\}$ is $(x_1, x_2, \ldots, x_n)^T$. Consider $\mathbb{R}^2$ with the basis $B = \{(1, 1), (1, -1)\}$. Let $v = (x, y)$. Then $(x, y) = a_1(1, 1) + a_2(1, -1)$ if and only if $a_1 = \frac{x+y}{2}$ and $a_2 = \frac{x-y}{2}$. Hence, $[(x, y)]_B = (\frac{x+y}{2}, \frac{x-y}{2})^T = \begin{pmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{pmatrix}$. Consider another basis $B' = \{(1, 2), (2, 1)\}$. Then $[(x, y)]_{B'} = \begin{pmatrix} \frac{2y-x}{3} \\ \frac{2x-y}{3} \end{pmatrix}$. Thus, the coordinate vector of a vector depends on the basis and it changes with a change of basis.

Theorem 10. Let V be an n-dimensional vector space over $\mathbb{F}$. Then $V \cong \mathbb{F}^n$. (V is isomorphic to Fn)

Proof: Let $B = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis of $V(\mathbb{F})$. The map $T : V \to \mathbb{F}^n$ given by $T(v) = [v]_B$ is an isomorphism. First we show that T is linear. Let $v, v' \in V$ with $[v]_B = (a_1, a_2, \ldots, a_n)^T$ and $[v']_B = (b_1, b_2, \ldots, b_n)^T$. Then $\alpha v + \beta v' = (\alpha a_1 + \beta b_1)v_1 + \cdots + (\alpha a_n + \beta b_n)v_n$ so that $[(\alpha v + \beta v')]_B = (\alpha a_1 + \beta b_1, \ldots, \alpha a_n + \beta b_n)^T = \alpha(a_1, \ldots, a_n)^T + \beta(b_1, \ldots, b_n)^T = \alpha T(v) + \beta T(v')$. Now $ker(T) = \{v \mid T(v) = 0\} = \{v \mid [v]_B = 0\} = \{0\}$. Thus T is one-one and onto (rank-nullity theorem).

Corollary 11. Two finite-dimensional vector spaces V and W over the field $\mathbb{F}$ are isomorphic if and only if $\dim(V) = \dim(W)$.

Lecture 12

Matrix Representation of a Linear Transformation & Similar Matrices

Definition 1. Let $B_1 = \{v_1, v_2, \ldots, v_n\}$ and $B_2 = \{u_1, \ldots, u_n\}$ be ordered bases of a vector space V over $\mathbb{F}$. Then the matrix $P_{B_1 \mapsto B_2}$ having the i-th $(1 \leq i \leq n)$ column as the coordinate vector of v_i with respect to the basis B_2, that is,

$$P_{B_1 \mapsto B_2} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix},$$

where $[v_i]_{B_2} = (p_{1i}, p_{2i}, \ldots, p_{ni})^T$, is called the transition matrix from the basis B_1 to the basis B_2.

Theorem 2. Let $B_1 = \{v_1, \ldots, v_n\}$ and $B_2 = \{u_1, \ldots, u_n\}$ be ordered bases of a vector space V over $\mathbb{F}$. If v is a vector in V, then $[v]_{B_2} = P_{B_1 \mapsto B_2}[v]_{B_1}$, where $P_{B_1 \mapsto B_2}$ is the transition matrix from B_1 to B_2.

Proof: Let $[v]_{B_1} = (a_1, a_2, \ldots, a_n)^T$. Then $v = a_1 v_1 + \ldots + a_n v_n$. We know $[v_i]_{B_2} = P_i$, where $i = 1, \ldots, n$ and P_i is the i-th column of $P_{B_1 \mapsto B_2}$. Thus, $v_1 = p_{11}u_1 + \ldots + p_{n1}u_n$, $v_2 = p_{12}u_1 + \ldots + p_{n2}u_n$, $\ldots$, $v_n = p_{1n}u_1 + \ldots + p_{nn}u_n$, where p_{ij} is the (i,j)-th entry of $P_{B_1 \mapsto B_2}$. Putting these value in (1), we get $v = (a_1p_{11} + a_2p_{12} + \cdots + a_np_{p1n})u_1 + \cdots + (a_1p_{n1} + a_2p_{n1} + \cdots + a_np_{nn})u_n$. Therefore, $[v]_{B_2} = P_{B_1 \mapsto B_2}[v]_{B_1}$. $\square$

Theorem 3. A transition matrix is invertible.

Proof: Let V be a vector space over $\mathbb{F}$ and B_1 and B_2 are bases of V. For $v \in V$, $P_{B_1 \mapsto B_2}[v]_{B_1} = [v]_{B_2}$. Let $P_{B_2 \mapsto B_1}$ be the transition matrix from basis B_2 to B_1. Then $P_{B_1 \mapsto B_2}P_{B_1 \mapsto B_2}[v]_{B_1} = P_{B_1 \mapsto B_2}[v]_{B_2} = [v]_{B_1}$. Thus, $P_{B_1 \mapsto B_2}$ is invertible (using Exercise: If a square matrix has a left inverse, then it is invertible).

Example 4. *Let $B_1 = \{(1,1), (1,-1)\}$ and $B' = \{(1,2), (2,1)\}$ be bases of $\mathbb{R}^2(\mathbb{R})$. Then*

$$P_{B_1 \mapsto B_2} = \begin{pmatrix} \frac{1}{3} & -1 \\ \frac{1}{3} & 1 \end{pmatrix} \text{ and } P_{B_2 \mapsto B_1} = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{pmatrix}.$$

Let $(x,y) \in \mathbb{R}^2$. Then $[(x,y)]_{B_1} = (\frac{x+y}{2}, \frac{x-y}{2})^t$ and $[(x,y)]_{B_2} = (\frac{2y-x}{3}, \frac{2x-y}{3})^t$. Verify that $[(x,y)]_{B_2} = P_{B_1 \mapsto B_2}[(x,y)]_{B_1}$ and $[(x,y)]_{B_1} = P_{B_2 \mapsto B_1}[(x,y)]_{B_2}$. Also, $P_{B_1 \mapsto B_2}$ is inverse of $P_{B_2 \mapsto B_1}$.

Matrix representation of a linear transformation: Let V and W be vector spaces over $\mathbb{F}$ with ordered bases $B_V = \{v_1, v_2, \ldots, v_m\}$ and $B_W = \{w_1, w_2, \ldots, w_n\}$ respectively. Let $T : V \to W$ be a linear transformation. Then

$$T(v_j) = \sum_{i=1}^{n} \alpha_{ij} w_i \text{ for } (1 \leq j \leq m),$$

where $\alpha_{ij} \in \mathbb{F}$ and we get an $n \times m$ matrix M_T given by

$$M_T = \alpha_{ij},$$

that is, the i-th column of M_T is the coordinate vector $[T(v_i)]_{B_W}$. The matrix M_T is called the **matrix representation of T with respect to the bases B_V and B_W**. Since the coordinate vectors are unique, M_T is also unique. We also denote the matrix representation of T with respect to B_V and B_W by $[T]_{B_V}^{B_W}$. If T is an operator $(T : V \to V)$ and both the bases are identical, then we simply write $[T]_B$.

Theorem 5. Let $v \in V$. Then $[T(v)]_{B_W} = [T]_{B_V}^{B_W} [v]_{B_V}$.

Example 6. *1. Let the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x, y, z) = (2x + z, y + 3z)$ with $B = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\}$ and $B' = \{(2, 3), (3, 2)\}$. Then*

$$\begin{aligned}
T(1,1,0) &= (2,1) = \frac{-1}{5}(2,3) + \frac{4}{5}(3,2) \\
T(1,0,1) &= (3,3) = \frac{3}{5}(2,3) + \frac{3}{5}(3,2) \\
T(1,1,1) &= (3,4) = \frac{6}{5}(2,3) + \frac{1}{5}(3,2)
\end{aligned}$$

Thus, the matrix representation of T with respect to B and B' is

$$\begin{pmatrix} -\frac{1}{5} & \frac{3}{5} & \frac{6}{5} \\ \frac{4}{5} & \frac{3}{5} & \frac{1}{5} \end{pmatrix}.$$

2. Let $D : P_2 \longrightarrow P_1$ be the differential operator. Find the matrix representations of D from $B = \{1, x, x^2\}$ to $B^{'} = \{1, 1 + x\}$.

$$\begin{aligned}
D(1) &= 0 = 0(1) + 0(1+x) \\
D(x) &= 1 = 1(1) + 0(1+x) \\
D(x^2) &= 2x = -2(1) + 2(1+x) \\
[D]_B^{B'} &= \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 2 \end{pmatrix}.
\end{aligned}$$

Definition 7. Let $A, B \in M_n(\mathbb{F})$. Then A and B are said to be similar if there exist an invertible matrix $P \in M_n(\mathbb{F})$ such that $A = P^{-1}BP$.

Theorem 8. Let $V(\mathbb{F})$ be a vector space with ordered bases B and $B^{'}$. Let T be a linear operator on V (that is, $T : V \to V$). If $[T]_B = A$ and $[T]_{B'} = A^{'}$, then $A^{'} = P^{-1}AP$, where P is the transition matrix from $B^{'}$ to B.

Proof: Let $v \in V$. Then $[T(v)]_B = A[v]_B$ and $[T(v)]_{B'} = A^{'}[v]_{B'}$. We know that $P[v]_{B'} = [v]_B$

so that $P([T(v)]_{B'}) = [T(v)]_B = A[v]_B$. Also, $P([T(v)]_{B'}) = P(A'[v]_{B'}) = P(A'P^{-1}[v]_B)$. Therefore, $A[v]_B = PA'P^{-1}[v]_B \ \forall [v]_B$. Hence, $A = PA'P^{-1}$ or $A' = P^{-1}AP$. $\square$

Theorem 9. Let $V(\mathbb{F})$ and $W(\mathbb{F})$ be vector spaces. Suppose B_1, B_1' are bases for V and B_2, B_2' are ordered bases for W. Then for any linear map $T : V \longrightarrow W$,

$$[T]_{B_1'}^{B_2'} = Q[T]_{B_1}^{B_2}P,$$

where $P = P_{B_1' \to B_1}$ and $Q = Q_{B_2 \to B_2'}$.

Example 10. *Consider Example 6 (1), let* $B_1 = \{(1,0,0),(0,1,0),(0,0,1)\}$, $B_1' = \{(1,1,0),(1,0,1),(1,1,1)\}$ *and* $B_2 = \{(1,0),(0,1)\}$. *and* $B_2' = \{(2,3),(3,2)\}$. *Then*

$$[T]_{B_1}^{B_2} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 3 \end{pmatrix} \text{ and } [T]_{B_1'}^{B_2'} = \begin{pmatrix} -\frac{1}{5} & \frac{3}{5} & \frac{6}{5} \\ \frac{4}{5} & \frac{3}{5} & \frac{1}{5} \end{pmatrix}.$$

The transition matrices

$$P = P_{B_1' \mapsto B_1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } Q = Q_{B_2 \mapsto B_2'} = \begin{pmatrix} \frac{-2}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{-2}{5} \end{pmatrix}.$$

Verify that $[T]_{B_1'}^{B_2'} = Q[T]_{B_1}^{B_2}P$.

Lecture 13

Rank of a matrix & System of linear equations

Definition 1. Let $A \in M_{m\times n}(\mathbb{F})$. The **column space of** A is the linear span of columns of A, i.e., column space(A) $= L(\{(a_{11}, a_{21}, \ldots, a_{m1}), \ldots, (a_{1n},\ a_{2n},\ \ldots, a_{mn})\}) \subseteq \mathbb{F}^m$, and the **row space of** A is the linear span of the rows of A, i.e., the row space$(A) = L(\{(a_{11}, a_{12}, \ldots, a_{1n}), \ldots, (a_{m1}, a_{m2}, \ldots, a_{mn}\}) \subseteq \mathbb{F}^n$. The dimension of the column space of (A) is called the **column rank of** A and dimension of the row space of (A) is called the **row rank of** A.

Theorem 2. Let $A \in M_{m\times n}(\mathbb{F})$. Then Row rank$(A)$ = Column rank(A).

Proof: Let $R_1, R_2, \ldots, R_m$ be the rows of A. Then the i^{th} vector $R_i = (a_{i1}, a_{i2}, \ldots, a_{in})$. Suppose dimension of the row space of A is s and $\{v_1, v_2, \ldots, v_s\}$ is a basis of the row space of A. Then

$$
\begin{aligned}
R_1 &= c_{11}v_1 + c_{12}v_2 + \ldots + c_{1s}v_s \\
R_2 &= c_{21}v_1 + c_{22}v_2 + \ldots + c_{2s}v_s \\
&\vdots \\
R_m &= c_{m1}v_1 + c_{m2}v_2 + \ldots + c_{ms}v_s
\end{aligned}
$$

Let $v_j = (b_{j1}, b_{j2}, \ldots, b_{jn})$ for $1 \le j \le s$. Then $a_{1i} = c_{11}b_{1i} + c_{12}b_{2i} + c_{1s}b_{si}$, $a_{2i} = c_{21}b_{1i} + c_{22}b_{2i} + c_{2s}b_{si}$, $\ldots$, $a_{mi} = c_{m1}b_{1i} + c_{m2}b_{2i} + c_{ms}b_{si}$. This implies, $(a_{1i}, a_{2i}, \ldots, a_{mi}) = b_{1i}(c_{11}, c_{21}, \ldots, c_{m1}) + \ldots + b_{si}(c_{1s}, c_{2s}, c_{ms})$. Thus, each column vector is a linear combination of s vectors $\{(c_{11}, c_{21}, \ldots, c_{m1}), (c_{12}, c_{22}, \ldots, c_{m2}), \ldots, (c_{1s}, c_{2s}, \ldots, c_{ms})\}$. Therefore, dim(column space) $\le s =$ dim(row space). Similarly, we can show that dim(row space) $\le s =$ dim(column space). $\square$

Definition 3. The rank of a matrix A is the dimension of row space of A (or the dimension of column space of A).

Definition 4. *The nullity of a matrix A is the dimension of the solution space of $Ax = 0$.*

Theorem 5 (Rank-Nullity Theorem for a Matrix)**.** *Let $A \in M_{m\times n}(\mathbb{R})$. Then*

$$
rank(A) + nullity(A) = number\ of\ columns\ of\ A.
$$

Proof. Recall that there is a one to one correspondence between $L(\mathbb{R}^n, \mathbb{R}^m)$ and $M_{m\times n}(\mathbb{R})$. Consider the map ϕ such that $T \mapsto [T]_B^{B'}$, where B and B' be the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Then ϕ is linear one-one and onto. For onto, given a matrix A, take the linear transformation T_A given by $T_A(x) = Ax$. $\square$

Remark 6. *1. The rank of a matrix A is same as the number of non-zero rows in its RRE form.*

Proof. Let the number of non zero rows in the RRE form of A is r. Observe that a row obtained by applying an elementary row operation is nothing but a linear combination of rows of the matrix, and the rows in RRE form are LI. Therefore, the dimension of row space or rank of A is r. $\square$

Determinantal-Rank of a matrix

Let $A \in M_{m\times n}(\mathbb{R})$. Then A has determinantal-rank r if
1. every $k \times k$ submatrix of A has zero determinant, where $k > r$;
2.there exist an $r \times r$ submatrix with non-zero determinant.

Theorem 7. *Rank*(A)=*Determinantal Rank*(A).

Proof. Let rank$(A) = l$ and determinantal-rank$(A) = r$. We show that $r = l$. Since determinantal-rank$(A) = r$, there exists an $r \times r$ submatrix R with non-zero determinant so that rank$(R) = r$, equivalently, all rows of R are linearly independent. Then the corresponding r rows of matrix A are LI. Therefore, $r \leq rank(A)$.

Let B be a submatrix of A consisting of linearly independent rows of A. Let $rank(A) = l$. Then order of B is $l \times n$ and $rank(B)$ is l. Hence, B has l linearly independent columns. Consider an $l \times l$ submatrix B' of B (also a submatrix of A) having those l linearly independent columnsof B. Then $rank(B') = l$ so that $|B'| \neq 0$. Therefore, $l \leq r$. $\square$

Application of rank in system of linear equations

First we recall a result on system of linear equation:

Theorem 8. *Let $Ax = b$ be a non-homogeneous system of linear equations, and $Ax = 0$ be the associated homogeneous system. If $Ax = b$ is consistent and x_0 is a particular solution of $Ax = b$, then any solution of $Ax = b$ can be written as $x = x_h + x_0$, where x_h is a solution of $Ax = 0$.*

Let $A \in M_{m\times n}(\mathbb{R})$ and Rank$(A) = r$. Suppose $Ax = b$ is a non-homogeneous system of linear equations, and $Ax = 0$ is the associated homogeneous system. Then

1. $Ax = b$ is consistent if and only if Rank$(A \mid b) = r$.

   Solution: If $Ax = b$ is consistent, then $b \in$ Column Space(A) so that Rank$(A \mid b) = r$. Similarly, the other way.

2. Let $Ax = b$ be consistent. Then the solution is unique if and only if $r = n$.

   Solution: Let $Ax = b$ have a unique solution. Then $Ax = 0$ has a unique solution, *i.e.*, the zero solution. This implies $nullity(A) = 0$. Then by rank-nullity theorem, we have $n = rank(A)$ and vice-versa.

3. If $r = m$, then $Ax = b$ always has a solution for every $b \in \mathbb{R}^m$.

   Solution: If $r = m$, then the column space is $\mathbb{R}^m$. Thus each vector in $\mathbb{R}^m$ is a linear combination of columns of A. Hence, $Ax = b$ has a solution for all $b \in \mathbb{R}^m$.

4. If $r = m = n$ then $Ax = b$ always has a unique solution for all b and further $Ax = 0$ has only zero solution.

   Solution: Since $r = m$, the column space is $\mathbb{R}^m$. Therefore, $Ax = b$ always has a solution for all b. Further, $nullity(A) = 0$. Thus, $Ax = 0$ has only zero solution and hence, $Ax = b$ always has a unique solution all b.

5. If $r = m < n$, for any $b \in \mathbb{R}^m$, $Ax = b$ as well as $Ax = 0$ have infinitely many solutions.

   Solution: Since $r = m$, $Ax = b$ has a solution for all $b \in \mathbb{R}^m$. Note that, $\text{nullity}(A) = (n - r) > 0$. Therefore, $Ax = 0$ has infinitely many solutions and hence, $Ax = b$ has infinitely many solutions.

6. In case (i) $r < m = n$, (ii) $r < m < n$ and (iii) $r < n < m$, if $Ax = b$ has a solution then there are infinitely many solutions.

   Solution: Note that $\text{nullity}(A) = (n - r) > 0$. Hence $Ax = 0$ has infinitely many solutions. Now if $Ax = b$ has a solutions then it has infinitely many solutions.

7. If $r = n < m$, then $Ax = 0$ has only zero solution and if $Ax = b$ has a solution, the solution is unique.

   Solution: In this case, $\text{nullity}(A) = 0$, implies $Ax = 0$ has only trivial solution. If $Ax = b$ has a solution, then it is unique.

Example 9. *Let $T : P_2(\mathbb{R}) \Rightarrow \mathbb{R}^2$ given by $T(p(x)) = (p(0), p(1))$. Find rank(T), nullity(T), basis of ker(T) and basis range(T).*

Solution: *Let $B = \{1, x, x^2\}$ and $B' = \{e_1, e_2\}$. Then*

$$[T]_B^{B'} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} = A.$$

$RRE(A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$. Thus, $Rank(T) = Rank(A) = 2$ so that range of T is $\mathbb{R}^2$ and its basis is $\{e_1, e_2\}$. Further, $nullity(T) = nullity(A) = nullity(RRE(A)) = 3 - 2 = 1$. The solution space of $Ay = 0$ is $\{(0, a, -a) \mid a \in \mathbb{R}\}$. Note that $y = [v]_B$, therefore, the $ker(T) = \{ax + (-a)x^2 \mid a \in \mathbb{R}\}$. Hence, basis of kernel T is $\{x - x^2\}$.

Lecture 14 (Eigenvalue & Eigenvector)

Definition 1. Let V be a vector space over $\mathbb{F}$ and $T : V \to V$ be a linear transformation. Then

1. a scalar $\lambda \in \mathbb{F}$ is said to be an **eigenvalue** or **characteristic value of T** if there exists a non-zero vector $v \in V$ such that $Tv = \lambda v$.

2. a non-zero vector v satisfying $Tv = \lambda v$ is called **eigenvector** or **characteristic vector of T associated to the eigenvalue λ.**

3. The set $E_\lambda = \{v \in V : Tv = \lambda v\}$ is called the **eigenspace of T associated to the eigenvalue λ.**

Example 2. *Let V be a non-zero vector space over $\mathbb{F}$.*

1. If T is the zero operator, zero is the only eigenvalue of T.

2. For identity operator, one is the only eigenvalue.

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (0, x)$. Then $T(x, y) = \lambda(x, y) \Leftrightarrow (0, x) = (\lambda x, \lambda y) \Leftrightarrow (\lambda x = 0, y = \lambda y \Leftrightarrow \lambda = 0, x = 0, y \neq 0$. Thus, 0 is the eigenvalue of T and $(0, 1)$ is an eigenvector corresponding to 0.

4. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (y, -x)$. Then $T(x, y) = \lambda(x, y) \Leftrightarrow (y, -x) = (\lambda x, \lambda y) \Leftrightarrow (\lambda^2 + 1)x = 0 \Leftrightarrow \lambda = \pm i, x \neq 0$. Thus, T has no real eigenvalue.

5. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ given by $T(x, y) = (y, -x)$. Then $T(x, y) = \lambda(x, y) \Leftrightarrow (y, -x) = (\lambda x, \lambda y) \Leftrightarrow (\lambda^2 + 1)x = 0 \Leftrightarrow \lambda = \pm i, x \neq 0$. Thus, T has two complex eigenvalues $\pm i$ and $(1, i)$ is an eigenvector corresponding to i and $(1, -i)$ is an eigenvector corresponding to $-i$.

5. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (2x + 3y, 3x + 2y)$. To find $\lambda \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^2$ such that $(2x + 3y, 3x + 2y) = \lambda(x, y)$ or $(2 - \lambda)x + 3y = 0, 3x + (2 - \lambda)y = 0$. The system of linear equations has a non-zero solution if and only if the determinant of the coefficient matrix, $\det \begin{pmatrix} 2-\lambda & 3 \\ 3 & 2-\lambda \end{pmatrix} = 0$ or $\lambda = -1, 5$. When $\lambda = 1$, $3x + 3y = 0$ so that $(1, -1)$ is an eigenvector ($(-a, a)$ are eigenvectors of corresponding to eigenvalue -1 for every $a \neq 0$). For $\lambda = 5$, $3x - 3y = 0$ so that $(1, 1)$ is an eigenvector (in fact, (a, a) is an eigenvector corresponding to eigenvalue 5 for $a \neq 0$).

Theorem 3. *Let T be a linear operator on a finite-dimensional vector space $V(\mathbb{F})$ and $\lambda \in \mathbb{F}$. The following statements are equivalent.*

1. λ is an eigenvalue of T.

2. The operator $T - \lambda I$ is singular (not invertible).

3. $\det[(T - \lambda I)]_B = 0$, where B is an ordered basis of V.

Proof. A linear transformation T is singular if and only if $ker(T) \neq \{0\}$. Thus, (1) $\iff$ (2). if $V(\mathbb{F})$ is finite-dimensional, then the eigenvalues and eigenvectors of T can be determined by its matrix representation $[T]_B$ with respect to a basis B. A scalar λ is an eigenvalue of $T \Leftrightarrow Tv = \lambda v \Leftrightarrow [T]_B[v]_B = \lambda[v]_B \Leftrightarrow ([T]_B - \lambda I)[v]_B = 0$ for non zero v. Thus, (3) $\Leftrightarrow$ (1). $\square$

Definition 4. Let $A \in M_n(\mathbb{F})$. A scalar $\lambda \in \mathbb{F}$ is said to be an **eigenvalue of A** if there exists a non-zero vector $x \in \mathbb{F}^n$ such that $Ax = \lambda x$. Such a non-zero vector x is called an **eigenvector of A associated to the eigenvalue λ**.

Let $A \in M_n(\mathbb{F})$. Observe, $det(xI - A)$ is an n degree polynomial in x over $\mathbb{F}$. A scalar λ is an eigenvalue of $A \Leftrightarrow \det(A - \lambda I) = 0$ or $\det(\lambda I - A) = 0$.

Definition 5. Let $A \in M_n(\mathbb{F})$. Then the polynomial $f(x) = \det(xI - A)$ is called the **characteristic polynomial** of A. The equation $\det(xI - A) = 0$ is called the **characteristic equation** of A.

Theorem 6. A scalar $\lambda \in \mathbb{C}$ is an eigenvalue if and only if λ is a root of the characteristic polynomial of A.

Example 7. Let $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. The characteristic polynomial of A is $\det \begin{pmatrix} x-1 & -1 & 0 \\ 0 & x-1 & -1 \\ -1 & 0 & x-1 \end{pmatrix}$, that is, $x^3 - 3x^2 + 3x - 2 = (x-2)(x^2 - x + 1)$. Thus, the roots are $\lambda = 2, \frac{1 \pm \sqrt{3}i}{2}$. If $\mathbb{F} = \mathbb{R}$, the only eigenvalue of A is 2 and if $\mathbb{F} = \mathbb{C}$, the eigenvalues are $2, \frac{1 \pm \sqrt{3}i}{2}$. We leave it to the reader to find the corresponding eigenvectors over the field $\mathbb{C}$. In this example, we see that a real matrix over $\mathbb{C}$ may have complex eigenvalues.

Example 8. Consider a matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The characteristic polynomial is $x^2 + 1$ and the roots are $\pm i$. Thus, A has no eigenvalue over $\mathbb{R}$ and two eigenvalues over $\mathbb{C}$. Note that, the existence of eigenvalue depends on the field.

Properties of eigenvalue and eigenvector

1. Let $A \in M_n(\mathbb{C})$. Then the sum of eigenvalues is equal to the trace of the matrix and the product of eigenvalues is equal to the determinant of the matrix.

   Let
$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$
   Then the characteristic polynomial of A is $f(\lambda) = |\lambda I - A| = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n$ with roots $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then $\lambda_1 + \lambda_2 + \ldots + \lambda_n = -\frac{a_1}{a_0}$ and $\lambda_1\lambda_2 \ldots \lambda_n = (-1)^n \frac{a_n}{a_0}$.

Note that $a_0 = 1$, $f(0) = a_n = |-A| = (-1)^n|A|$ and $a_1 = -(a_{11} + a_{22} + \ldots + a_{nn})$. Therefore, $\lambda_1 + \lambda_2 + \ldots + \lambda_n = -\frac{a_1}{a_0} = (a_{11} + a_{22} + \ldots + a_{nn}) = trace(A)$ and $\lambda_1\lambda_2 \ldots \lambda_n = (-1)^n \frac{a_n}{a_0} = |A| = det(A)$.

2. If A is a non-singular matrix and λ is any eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}.

   Let λ be an eigenvalue of A, then there exists $0 \neq x \in \mathbb{F}^n$ such that $Ax = \lambda x \Leftrightarrow A^{-1}x = \frac{1}{\lambda}x$.

3. A and and A^T have the same eigenvalues.

   It is enough to show that A and A^T have the same characteristic polynomials. The characteristic polynomial of A is $|\lambda I - A| = |(\lambda I - A)^T| = |\lambda I - A^T|$=characteristic polynomial of A^T.

4. Similar matrices have the same eigenvalues (or characteristic equations).

   Let A and B are two matrices which are similar then there exists an invertible matrix P such that $A = P^{-1}BP$. Then characteristic polynomial of A is $|\lambda I - A| = |\lambda I - P^{-1}BP| = |P^{-1}(\lambda I - B)P| = |\lambda I - B|$.

5. If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k for a positive integer k.

6. Let $\mu \in \mathbb{F}$ and $A \in M_n(\mathbb{F})$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $\lambda \pm \mu$ is eigenvalue of $A \pm \mu I$.

Lecture 15 (Diagonalizability)

Definition 1. Let $A \in M_n(\mathbb{R})$ with the characteristic polynomial $f(x)$. Let λ be an eigenvalue of A then the largest power k such that $(x-\lambda)^k$ is a factor of $f(x)$ is called the algebraic multiplicity of λ (A.M.(λ)).

Theorem 2. Let λ be an eigenvalue of a matrix A. Then the set $E_\lambda = \{x \in \mathbb{F}^n \mid Ax = \lambda x\}$ forms a subspace of $\mathbb{F}^n$ and it is called eigenspace corresponding to the eigenvalue λ. Observe that E_λ is the set of all eigenvectors associated to λ including the zero vector.

Definition 3. The dimension of the eigenspace (E_λ) of eigenvalue λ is called the geometric multiplicity of λ (G.M.(λ)). Thus the geometric multiplicity of λ, G.M.(λ) = Nullity $(A-\lambda I) = n-$ Rank $(A-\lambda I)$.

Remark 4. *1. Thus the geometric multiplicity of λ, G.M.(λ) = Nullity $(A-\lambda I) = n-$ Rank $(A-\lambda I)$.*
2. G.M.(λ) ≥ 1.

Theorem 5. G.M. (λ) $\leq$ A.M. (λ), for an eigenvalue λ of A.

Proof: Let $dim(E_\lambda) = p$ and let $S = \{X_1, X_2, \dots, X_p\}$ be a basis of E_λ. Then S can be extended to a basis S' of $\mathbb{F}^n$. Let $S' = \{X_1, X_2, \dots, X_p, X_{p+1}, \dots, X_n\}$. Then

$$
\begin{aligned}
&AX_1 = \lambda X_1\\
&AX_2 = \lambda X_2\\
&\vdots\\
&AX_p = \lambda X_p\\
&AX_{p+1} = a_{(p+1)1}X_1 + a_{(p+1)2}X_2 + \ldots + a_{(p+n)n}X_n\\
&\vdots\\
&AX_n = a_{n1}X_1 + a_{n2}X_2 + \ldots + a_{nn}X_n.
\end{aligned}
$$

The matrix representation of the above system of equations is

$$A = \begin{bmatrix} \lambda I_p & B \\ 0 & C \end{bmatrix},$$

where I_p is the identity matrix of order p. Thus, the characteristic polynomial of A is $f(x) = det(xI - A) = (x-\lambda)^p g(x)$, where $g(x)$ is a polynomial. Hence, the algebraic multiplicity of λ is at least p.

Definition 6. Let $A \in M_n(\mathbb{F})$. Then A is called diagonalizable if it has n linearly independent eigenvectors.

linearly independent eigenvectors does not imply distinct eigenvalues.

Lemma 7. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be distinct eigenvalues of A and $v_1, v_2, \dots, v_k$ be the corresponding eigenvectors respectively. Then $v_1, v_2, \dots, v_k$ are linearly independent.

Proof. The proof is by induction. Let $k = 2$ and v_1, v_2 are linearly dependent. Then $v_1 = \alpha v_2 \Rightarrow$ for some $0 \neq \alpha \in \mathbb{F}$. Thus $Av_1 = \alpha A v_2 \Rightarrow \lambda_1 v_1 = \alpha \lambda_2 v_2 \Rightarrow \alpha(\lambda_1 - \lambda_2)v_2 \Rightarrow \lambda_1 = \lambda_2$, which is a contradiction. Suppose the result is true for $k-1$, that is, $v_1, v_2, \ldots, v_{k-1}$ are linearly independent. Let $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$. Then $A(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k) = 0 \Rightarrow \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_k \lambda_k v_k = 0 \Rightarrow$ $\alpha_1(\lambda_1 - \lambda_k)v_1 + \alpha_2(\lambda_2 - \lambda_k)v_2 + \cdots + \alpha_{k-1}(\lambda_{k-1} - \lambda_1)v_{k-1} = 0$ (since $\lambda_k(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k) = 0$). By induction hypothesis, $v_1, v_2, \ldots, v_{k-1}$ are linearly independent, hence $\alpha_i = 0$ for $1 \leq i \leq k-1$ as $\lambda_i \neq \lambda_k$. Thus, $\alpha_k v_k = 0$ so that $\alpha_k = 0$. □

Theorem 8. *Let $A \in M_n(\mathbb{F})$. The following statements are equivalent.*

1. A is diagonalizable.

2. There exists an invertible matrix P such that $P^{-1}AP = D$, where D is a diagonal matrix.

3. $A.M.(\lambda) = G.M.(\lambda)$ for each eigenvalue λ of A.

Proof. Let $X_1, X_2, \ldots, X_n$ be n independent eigenvectors of A. Construct a matrix P having X_i as its i-th column. Then $P^{-1}AP = D$, where D is a diagonal matrix and its i-th diagonal entry is the eigenvalue corresponding to X_i. Thus, $1 \Rightarrow 2$. For $2 \Rightarrow 1$, note that the columns of P are L.I. as P is invertible and each column of P is an eigenvector of A. By Lemma 7, $3 \Leftrightarrow 1$. □

Example 9. *Check diagonalizability of the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$. If diagonalizable, find a matrix P such that $P^{-1}AP$ is a diagonal matrix.*

Solution: The characteristic polynomial of A is $(x+1)(x-4)$. Hence $A.M.(\lambda) = 1 = G.M.(\lambda)$ for $\lambda = 4, -1$. Hence, A is diagonalizable. For finding P such that $P^{-1}AP$ is diagonal matrix, we find eigenvectors of A. Eigenvectors corresponding to $\lambda = -1$ and 4 are respectively $v_{-1} = (1, -1)$ and $v_4 = (2, 3)$. Since eigenvectors corresponding to distinct eigenvalues are LI, $\{(1, -1), (2, 3)\}$ is LI. Construct

$$P = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}.$$

One can see easily $D = P^{-1}AP$, where $D = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}$.

Definition 10. Let $T : V \to V$ be a linear transformation, where V is an n dimensional vector space over $\mathbb{F}$. Then T is called diagonalizable if V has a basis in which each vector is an eigenvector of T, that is, T has n independent eigenvectors.

Remark 11. *Let $V(\mathbb{F})$ is n-dimensional vector space and $T : V \to V$ be a linear operator. Then*

1. if T is diagonalizable and B is a basis of V consisting of eigenvectors, then $[T]_B = D$, where D is a diagonal matrix.

2. if T has n distinct eigenvalues, then T is diagonalizable.

3. if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T and E_{λ_i} are the associated eigenspaces, then T is diagonalizable if and only if $\dim V = \dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k}$.

Example: The operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (2x, x + 2y, 4x + 3z)$ is not diagonalizable. To see this, we consider the standard basis B of $\mathbb{R}^3$ and $[T]_B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 4 & 0 & 3 \end{pmatrix}$. The characteristic polynomial is $(x-2)^2(x-3)$. Thus $AM(2) = 2$ and $AM(3) = 1$. $E_2 = \{(0, x, 0) : x \in \mathbb{R}\}$ with $\dim E_2 = 1$ and $E_3 = \{(0, 0, x) : x \in \mathbb{R}\}$ with $\dim E_3 = 1$. Here, we get $\dim \mathbb{R}^3 \neq \dim E_2 + \dim E_3$. Hence, T is not diagonalizable.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (-x + 2y + 4z, -2x + 4y + 2z, -4x + 2y + 7z)$ is diagonalizable. To see this, we consider the standard basis B of $\mathbb{R}^3$ and $[T]_B = \begin{pmatrix} -1 & 2 & 4 \\ -2 & 4 & 2 \\ -4 & 2 & 7 \end{pmatrix}$. The characteristic polynomial is $\det(xI - [T]_B) = -x^3 + 10x^2 - 33x + 36 = (x-3)^2(x-4)$. Thus $AM(3) = 2$ and $AM(4) = 1$. Solving $([T]_B - 3I)X = 0$, we get $(1, 0, 1), (1, 2, 0)$ are independent solutions. Hence, $\dim E_3 = 2$ and $\dim E_4 = 1$. Here, we get $\dim \mathbb{R}^3 = \dim E_3 + \dim E_4$. Hence, T is diagonalizable.

Further, if we want to find a matrix P such that $P^{-1}[T]_B P = D$ for some diagonal matrix D. We need to compute a basis of eigen vectors. We have already found eigen vectors corresponding to $\lambda = 3$. Now let $\lambda = 4$, solving the system $([T]_B - 4I)X = 0$, we get $(2, 1, 2)$ is an eigen vector. The eigen vectors corresponding to distinct eigen values are linearly independent. Hence, $\{(1, 0, 1), (1, 2, 0), (2, 1, 2)\}$ is a basis consisting of eigen vectors. To find P, we will place basis vectors in the column, i.e., $P = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$. The diagonal matrix D is obtained by placing the eigen values on the diagonal in the same order as eigen vectors in P, that is, if the first column of P is corresponding to eigen vector of λ_1, the first diagonal entry is going to be λ_1 and so on. Here, $D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$. Verify yourself that $D = P^{-1}[T]_B P$.

Lecture 16

(Cayley Hamilton Theorem, minimal polynomial & Diagonalizability)

Theorem 1. Cayley-Hamilton Theorem: Every square matrix satisfies its characteristic equation, that is, if $f(x)$ is the characteristic polynomial of a square matrix A, then $f(A) = 0$.

Example 2. *Let* $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. *Find inverse of A using Cayley-Hamilton theorem.*

Solution: The characteristic polynomial of A is $f(x) = x^3 - 2x^2 + 1$. The constant term of $f(x) = 1 = \det(A)$, the matrix A is invertible. By Cayley-Hamilton Theorem $f(A) = 0$. Therefore $A^3 - 2A^2 + I = 0 \Rightarrow A^{-1} = -A^2 + 2A \Rightarrow -\begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 2 & 2 & 0 \end{pmatrix} \Rightarrow A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix}$.

Definition 3. A polynomial $m(x)$ is said to be the minimal polynomial of A if

(*i*) $m(A) = 0$;

(*ii*) $m(x)$ is a monic polynomial (the coefficient of the highest degree term is 1);

(*iii*) if a polynomial $g(x)$ is such that $g(A) = 0$, then $m(x)$ divides $g(x)$.

Remark 4. 1.The minimal polynomial of a matrix is unique.
2. The minimal polynomial divides its characteristic polynomial.

Theorem 5. The minimal polynomial and the characteristic polynomial have the same roots.

Proof: Let $f(x)$ and $m(x)$ be the characteristic and minimal polynomial of a matrix respectively. Then $f(x) = g(x)m(x)$. If α is a root of $m(x)$, then it is also a root of $f(x)$. Conversely, if α is a root of $f(x)$, then α is an eigenvalue of the matrix. Therefore, there is a non-zero eigenvector v such that $Av = \alpha v$, this implies $m(A)v = m(\alpha)v$, *i.e.*, $m(\alpha)v = 0$, and $v \neq 0$ so that $m(\alpha) = 0$. □

Theorem 6. Similar matrices have the same minimal polynomials.

Proof: Let A and B be two similar matrices. Then $A = P^{-1}BP$ for some invertible matrix P. Let $m_1(x) = a_0 + a_1x + \ldots + x^n$ and $m_2(x) = b_0 + b_1x + \ldots + x^l$ be the respective minimal polynomials of A and B. Then $m_2(A) = 0$, which implies $m_1(x)|m_2(x)$. Similarly $m_1(B) = 0$, which implies $m_2(x)|m_1(x)$. □

Theorem 7. Let $A \in M_n(\mathbb{F})$ and $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ be all eigenvalues of A, where $\lambda_i \neq \lambda_j$ for $i \neq j$. The matrix A is diagonalizable if and only if its minimal polynomial is a product of distinct linear polynomials, that is, $m(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k)$, where λ_i's are distinct elements of $\mathbb{F}$.

Example 8. *A matrices $A \in M_n(\mathbb{R})$ such that $A^2 - 3A + 2I = 0$ is diagonalizable.*

Solution: *Take $g(x) = x^2 - 3x + 2$, then $g(A) = 0$. Note that $g(x) = (x-1)(x-2)$ and the minimal polynomial $m(x)$ of A divides $g(x)$. Therefore, either $m(x) = (x-1)$ or $m(x) = (x-2)$ or $m(x) = (x-1)(x-2)$. In either of the case, the minimal polynomial is a product of distinct linear polynomials, hence diagonalizable.*

Lecture 17
Inner Product Space

Let $V = \mathbb{R}^2$ and $P = (x_1, x_2)$ and $Q = (y_1, y_2)$ be two vectors in V. The dot product of P and Q is defined as $(x_1, x_2) \cdot (y_1, y_2) = x_1y_1 + x_2y_2$. Then the length of P, $||P|| = \sqrt{(x_1, x_2) \cdot (x_1, x_2)}$, distance between P and Q is $d(p,q) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2} = \sqrt{(x_1-y_1, x_2-y_2).(x_1-y_1, x_2-y_2)}$ and the angle (θ) between P and Q is defined as $cos\theta = \frac{P.Q}{||P||||Q||}$.

Observe that the above dot product satisfies the following properties :

1. $(x \cdot x) \geq 0$ and $(x \cdot x) = 0$ if and only if $x = 0$;
2. $(x \cdot y) = (y \cdot x)$, $\forall x, y \in \mathbb{R}^n$;
3. $((\alpha x) \cdot y) = \alpha(x \cdot y)$, $\forall \alpha \in \mathbb{R}$;
4. $((x + y) \cdot z) = (x \cdot z) + (y \cdot z)$.

In an arbitrary vector space, we define a function which satisfies the above four conditions, we call this function **inner product**, with the help of this function we can define the geometric concepts such as length of a vector, distance between two vectors and angle between the vectors.

Definition 1. Let V be a vector space over $\mathbb{F}$. A function $\langle\, , \rangle : V \times V \longrightarrow \mathbb{F}$ is called an inner product on V if it satisfies the following properties.
1. $\langle x, x\rangle \geq 0\ \forall x \in V$ and $\langle x, x\rangle = 0$ if and only if $x = 0$;
2. $\langle x, y\rangle = \overline{\langle y, x\rangle}$, $\forall x, y \in V$;
3. $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$, $\forall \alpha \in \mathbb{F}$ and $\forall x, y, z \in V$.

A vector space $V(\mathbb{F})$ together with an inner product $\langle\, , \rangle$ is called an inner product space and denoted by $(V, \langle\, , \rangle)$.

Example 2. *1. Let $V = \mathbb{R}^n$ over $\mathbb{R}$ with $\langle x, y\rangle = x \cdot y$, that is, $\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.*
2. Let $V = \mathbb{C}^n$ over $\mathbb{C}$. Define $\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\rangle = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}$.
3. Let $V = \mathbb{R}^2$, $\mathbb{F} = \mathbb{R}$ and $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ such that $a, c > 0$ and $ac - b^2 > 0$. Define $\langle x, y\rangle = y^T Ax$.
4. Let $V = C[a, b]$, $\mathbb{F} = \mathbb{R}$. Define $\langle f(x), g(x)\rangle = \int_a^b f(x)\overline{g(x)}dx$.
5. Let $V = M_n(\mathbb{R})$, $\mathbb{F} = \mathbb{R}$. Then for $A, B \in V$, define $\langle A, B\rangle = trace(AB^T)$.

Proposition 3. *Every finite dimensional vector space is an inner product space.*

Proof. Let $B = \{v_1, \dots v_n\}$ be an ordered basis of $V(\mathbb{F})$. Then for $u, v \in V$, define $\langle u, v\rangle = \alpha_1\overline{\beta_1} + \dots + \alpha_n\overline{\beta_n}$, where $(\alpha_1, \dots, \alpha_n)^T = [u]_B$ and $(\beta_1, \dots, \beta_n)^T = [v]_B$. □

Note that $\langle v, v\rangle > 0$ for non-zero $v \in V$. This leads us to define the concept of length of a vector in an inner product space.

Definition 4. The length of a vector v (norm of a vector v) is defined as $||v|| = \sqrt{\langle v, v\rangle}$.

Theorem 5 (Cauchy-Schwartz Inequality). *Let V be an inner product space. Then $|\langle v, u\rangle| \leq ||v||\,||u||$, $\forall u, v \in V$. The equality holds if and only if the set $\{u, v\}$ is linearly dependent.*

Proof: Clearly, the result is true for $u = 0$. Suppose $u \neq 0$. Let $w = v - \frac{\langle v,u\rangle}{||u||^2}u$. Then $w \in V$. By the property $\langle w, w\rangle \geq 0$, we get $||v||^2 - \frac{|\langle v,u\rangle|^2}{||u||^2} \geq 0$. Therefore, $|\langle v, u\rangle| \leq ||v||\,||u||$.

For equality, if $u = 0$ then the set $\{0, v\}$ is L.D.. If $u \neq 0$ then from the above we have $v = \frac{\langle v,u\rangle}{||u||^2}u$. Conversely, let u, v are L.D. then $u = \alpha v$ for some $\alpha \in \mathbb{F}$. Then $|\langle u, v\rangle| = |\langle \alpha v, v\rangle| = |\alpha|||v||^2 = ||u||\,||v||$.

Proposition 6. *Let $(V(\mathbb{F}), \langle\,,\rangle)$ be an inner product space. Then*

1. $||u + v|| \leq ||v|| + ||u||$, $\forall u, v \in V$. (Triangle inequality)

2. $||u + v||^2 + ||u - v||^2 = 2(||v|| + ||u||)^2$ $\forall u, v \in V$. (Parallelogram law)

Proof: By definition, $||u+v||^2 = \langle u+v, u+v\rangle = ||v||^2+\langle u, v\rangle+\overline{\langle u, v\rangle}+||u||^2 = ||v||^2+2Re(\langle u, v\rangle)+||u||^2 \leq ||v||^2 + 2|\langle u, v\rangle| + ||u||^2 = (||u|| + ||v||)^2$. Prove the second statement yourself.

Definition 7. *Let u and v be vectors in an inner product space $(V, \langle\,,\rangle)$. Then u and v are* **orthogonal** *if $\langle u, v\rangle = 0$. A set S of an inner product space is called an* **orthogonal set** *of vectors if $\langle u, v\rangle = 0$ for all $u, v \in S$ and $u \neq v$. An* **orthonormal set** *is an orthogonal set S with the additional property that $||u|| = 1$ for every $u \in S$.*

Proposition 8. *An orthogonal set of non-zero vectors is linearly independent.*

Proof: Let S be an orthogonal set (finite or infinite) of non-zero vectors in a given inner product space. Suppose $v_I, v_2, \dots, v_m$ are distinct vectors in S and take $w = \alpha_1 v_1 + \cdots + \alpha_m v_m$. Then $\langle w, v_i\rangle = \langle \alpha_1 v_1 + \cdots + \alpha_m v_m, v_i\rangle = \alpha_1\langle v_1, v_i\rangle + \alpha_2\langle v_2, v_i\rangle + \cdots + \alpha_m\langle v_m, v_i\rangle = \alpha_i\langle v_i, v_i\rangle$. Note that $v_i \neq 0$ so that $\langle v_i, v_i\rangle \neq 0$. If $w = 0$, then $\alpha_i = 0$ for each i. Therefore, S is linearly independent.

Gram-Schmidt orthogonalization process

Theorem 9. *Let $(V, \langle\, ,\rangle)$ be an inner product space and $S = \{v_1, v_2, \ldots, v_n\}$ be a linearly independent set of vectors in V. Then we get an orthogonal set $\{w_1, w_2, \ldots, w_n\}$ in V such that*

$$L(\{v_1, v_2, \ldots, v_n\}) = L(\{w_1, w_2, \ldots, w_n\}).$$

Proof. $w_1 = v_1$, then $L(\{w_1\}) = L(\{v_1\})$;
$w_2 = v_2 - \frac{\angle v_2,w_1\rangle}{\langle w_1,w_1\rangle} w_1$, then $\langle w_2, w_1\rangle = 0$ with $L(\{w_1, w_2\}) = L(\{v_1, v_2\})$;
$w_3 = v_3 - \frac{\langle v_3,w_2\rangle}{\langle w_2,w_2\rangle} w_2 - \frac{\langle v_3,w_1\rangle}{\langle w_1,w_1\rangle} w_1$, then $\langle w_3, w_1\rangle = 0$, and $\langle w_3, w_2\rangle = 0$ with $L(\{w_1, w_2, w_3\}) = L(\{v_1, v_2, v_3\})$;
Inductively,
$\boxed{w_n = v_n - \frac{\langle v_n,w_{n-1}\rangle}{\langle w_{n-1},w_{n-1}\rangle} w_{n-1} - \frac{\langle v_n,w_{n-2}\rangle}{\langle w_{n-2},w_{n-2}\rangle} w_{n-2} - \cdots - \frac{\langle v_n,w_1\rangle}{\langle w_1,w_1\rangle} w_1,}$ then $\langle w_n, w_i\rangle = 0$ for $i \neq n$ with $L(\{v_1, v_2, \ldots, v_n\}) = L(\{w_1, w_2, \ldots, w_n\})$. □

Remark 10. *1. The method by means of which orthogonal vectors $w_1, \ldots, w_n$ are obtained is known as the* **Gram-Schmidt orthogonalization process***.*
2. Every finite-dimensional inner product space has an orthonormal basis.
3.Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for an inner product space V. Then for any $w \in V$, $w = \langle w, v_1\rangle v_1 + \cdots + \langle w, v_n\rangle v_n$.

Example 11. *Find an orthogonal basis of $\mathbb{R}^2$ with the inner product given by $\langle (x_1, y_1), (x_2, y_2)\rangle = x_1x_2 + 2x_1y_2 + 2x_2y_1 + 5y_1y_2$.*

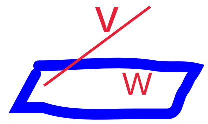
Solution: We know that $\{e_1, e_2\}$ is a basis of $\mathbb{R}^2$. Since $\langle e_1, e_2\rangle = 2 \neq 0$, the standard basis is not an orthogonal basis under the defined inner product. To get an orthogonal basis we use Gram-Schmidt process: $w_1 = e_1$ and $w_2 = e_2 - \langle e_2, e_1\rangle \frac{e_1}{||e_1||^2}$ and $||e_1||^2 = \langle e_1, e_1\rangle = 1$ so that $w_2 = e_2 - 2e_1$. Thus $\{e_1, e_2 - 2e_1\}$ is an orthogonal basis.

Lecture 18

Orthogonal Projection & Shortest Distance

Definition 1. *Let $(V, \langle\, , \rangle)$ be an inner product space and W be a subspace of V. Let $v \in V$. The orthogonal projection $P_W(v)$ of v onto W is a vector in W such that*

$$\langle (v - P_W(v)), w\rangle = 0 \ \ \forall w \in W.$$



Theorem 2. Let W be a finite-dimensional subspace of an inner product space V with an orthonormal basis $\{w_1, \ldots, w_n\}$. The orthogonal projection of $v \in V$ onto W is $P_W(v) = \langle v, w_1\rangle w_1 + \ldots + \langle v, w_n\rangle w_n$.

Proof. Let $P_W(v) = w_v$. Note that $w_v \in W$ and $\{w_1, \ldots, w_n\}$ is a basis of W. Hence, $w_v = \langle w_v, w_1\rangle w_1 + \langle w_v, w_2\rangle w_2 + \cdots + \langle w_v, w_n\rangle w_n$. Further, $\langle v - w_v, w_i\rangle = 0 \Rightarrow \langle v, w_i\rangle - \langle w_v, w_i\rangle = 0 \Rightarrow \langle w_v, w_i\rangle = \langle v, w_i\rangle \forall i$. $\square$

Remark 3. *1. $P_W(v) = w_v \Leftrightarrow ||v - w_v|| \leq ||v - w|| \,\forall\, w \in W$.*
2. Let v and u be two vectors in the inner product space V. Then orthogonal projection of u along v is $P_v(u) = \frac{<u,v>}{||v||^2} v$.
3. $P_W(v) \in W$ and $\langle v - P_W(v), w\rangle = 0$ for all $w \in W$, i.e., $v - P_W(v)$ is orthogonal to all the elements of W.

Definition 4. Let $(V, \langle\, , \rangle)$ be an inner product space and S be a non-empty subset of V. Then orthogonal complement of S, denoted by $S^\perp$, is defined as $S^\perp = \{v \in V \mid \langle v, s\rangle = 0 \,\forall s \in S\}$.

Definition 5. Let $(V, \langle\, , \rangle)$ be an inner product space and S_1 and S_2 be two subspaces of V. Then S_1 is perpendicular to S_2, $S_1 \perp S_2$, if $\langle s_1, s_2\rangle = 0$ for all $s_1 \in S_1$ and $s_2 \in S_2$.

Remark 6. *1. $V^\perp = \{0\}$.*
2. $\{0\}^\perp = V$.
3. Given any subset $W \subseteq V$, $W^\perp$ is a subspace of V.
4. $W \cap W^\perp = \{0\}$.

Theorem 7. Let $(V, \langle\, , \rangle)$ be an inner product space and and S_1 and S_2 are any two subsets of V. Then
1. $S \subseteq S^{\perp\perp}$.
2. if $S_1 \subseteq S_2$ then $S_2^\perp \subseteq S_1^\perp$.
3. if W is a finite-dimensional subspace of V, then $V = W \oplus W^\perp$.
4. if V is a finite-dimensional inner product space and W is a subspace of V, then $W = W^{\perp\perp}$.

Proof(i) Let $w \in S$ then $\langle w, v\rangle = 0$ for $v \in S^\perp$ which is equivalent to $w \in S^{\perp\perp}$. Thus $S \subseteq S^{\perp\perp}$

Proof(ii) Let $w \in S_2^\perp$ then $\langle w, v\rangle = 0$ for $v \in S_2$, but $S_1 \subseteq S_2$ hence $\langle w, v\rangle = 0$ for $v \in S_1$ this implies $w \in S_1^\perp$. Hence $S_2^\perp \subseteq S_1^\perp$.

Proof(iii) Let $\{v_1, v_2, \ldots, v_k\}$ be an orthogonal basis of W. Then for any $v \in V$, $P_W(v) = \sum_{i=1}^{k} \langle v, v_i \rangle \frac{v_i}{||v_i||^2}$. Thus, for $v \in V$, $v = P_W(v) + (v - P_W(v)) \in W + W^{\perp}$. Further, $W \cap W^{\perp} = \{0\}$. Therefore, $V = W \oplus W^{\perp}$.

Proof(iv) By the above result, we have $V = W \oplus W^{\perp}$. Since $W^{\perp}$ is a subspace of V, $V = W^{\perp} \oplus W^{\perp\perp}$. Moreover, $W \subseteq W^{\perp\perp}$. Let $v \in W^{\perp\perp}$. Since V is finite dimensional, $\dim W + \dim W^{\perp} = \dim W^{\perp} + \dim W^{\perp\perp}$ so that $W = W^{\perp\perp}$.

Example 8. *Let $V = M_n(\mathbb{R}), \mathbb{F} = \mathbb{R}$ with inner product given by $\langle A, B \rangle = tr(AB^T)$. Let W be the space of diagonal matrices. Find $W^{\perp}$.*

Solution: A basis of W is given by $B = \{e_{11}, e_{22}, \ldots, e_{nn}\}$. Note that B is orthonormal.
$W^{\perp} = \{A \in M_n(\mathbb{R}) \mid tr(AB^T) = 0\ \forall A \in W\} = \{A \in M_n(\mathbb{R}) \mid tr(Ae_{ii}^T) = 0 \text{ for } i = 1, 2, \ldots, n\}$ $= \{A \in M_n(\mathbb{R}) \mid tr(Ae_{ii}) = 0 \text{ for } i = 1, 2, \ldots, n\} = \{A \in M_n(\mathbb{R}) \mid a_{ii} = 0 \text{ for } i = 1, 2, \ldots, n\}$. Thus $W^{\perp}$ is collection of matrices having diagonal entries zero.

Shortest distance of a point from a subspace

Definition 9. *Let $(V, \langle\,,\rangle)$ be an inner product space and W be its finite dimensional subspace. Then the shortest distance of a vector $v \in V$ is given by $||v - P_W(v)||$.*

Example 10. *Find the shortest distance of $(1, 1)$ from the line $2y = x$.*

Solution: Here $W = L(\{(2, 1)\})$. Note that $||(2, 1)||^2 = 5$ so that orthonormal basis of W is $\{(2, 1)/\sqrt{5}\}$. Thus, $P_W((1, 1)) = \langle (1, 1), (2, 1) \rangle \frac{(2,1)}{5} = \frac{3}{5}(2, 1)$. The shortest distance of $(1, 1)$ from the line $y = 2x$ is $||(1, 1) - \frac{3}{5}(2, 1)|| = ||(-1, 2)/5|| = \frac{1}{\sqrt{5}}$.

Lecture 19

Fundamental Theorem of Linear Algebra & Least-Square Approximation

Fundamental Subspaces

Let $A \in M_{m\times n}(\mathbb{R})$. Suppose $N(A)$ is the null space of A, $C(A)$ is the column space of A, $C(A^T)$ is the column space of A^T and $N(A^T)$ is the null space of A^T. Then $N(A)$, $C(A^T)$ are subspaces of $\mathbb{R}^n$, and $C(A)$, $N(A^T)$ are subspaces of $\mathbb{R}^m$. These subspaces are called fundamental subspaces associated to A.

Lemma 1. *$N(A) \perp C(A^T)$ and $C(A) \perp N(A^T)$.*

?

Proof: Let $x \in N(A)$ and $y \in C(A^T)$. Then $A(x) = 0$ and $A^T z = y$ for some $z \in \mathbb{R}^m$. Then $y^T x = z^T A x = 0$, that is, $\langle x, y\rangle = 0$ so that $N(A) \perp C(A^T)$. Similarly, $C(A) \perp N(A^T)$. □

Theorem 2 (Fundamental Theorem of Linear Algebra)**.** *Let $A \in M_{m\times n}(\mathbb{R})$. Then*

1. $\mathbb{R}^n = N(A) \oplus C(A^T)$

2. $\mathbb{R}^m = C(A) \oplus N(A^T)$.

Proof: Since $C(A^T)$ is a subspace of $\mathbb{R}^n$, $\mathbb{R}^n = C(A^T) \oplus (C(A^T))^\perp$. We claim that $C(A^T)^\perp = N(A)$. By Lemma 1 , $N(A) \subseteq C(A^T)^\perp$. Note that $n = \dim(C(A^T)) + \dim((C(A^T))^\perp)$ and by rank-nullity theorem $n = rank(A) + nullity(A)$. This implies $\dim(N(A)) = dim((C(A^T))^\perp)$. Hence, $N(A) = (C(A^T))^\perp$. Similarly one can proof $\mathbb{R}^m = C(A) + N(A^T)$. □

Least-Square Approximation

R^m

Problem 3. *Let $A \in M_{m\times n}(\mathbb{R})$ and $b \in \mathbb{R}^n$ such that $b \notin C(A)$, where $C(A)$ is the column-space of A. In other words, the system $Ax = b$ is inconsistent. So the problem is to find a "pseudo solution" or "approximate solution" under certain condition in error term.*

Definition 4 (Least-Square Method)**.** *A method to approximate a solution of an inconsistent system of linear equations such that the solution minimizes the sum of square of errors made in every equation.*

Let $AX = b$ be an inconsistent system of linear equation, where $A \in M_{m\times n}(\mathbb{R}), X \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. Suppose $X_0 = (x_1, x_2, \dots, x_n)$ is an approximate solution of the system. Then $AX_0 = b'$ and $b' \neq b$. The error term for the i-th equation is $|b_i - b_i'| = |\sum_{j=1}^{n} a_{ij}x_j - b_i|$. For X_0 to be a least-square solution of the system, the sum of square of the errors made in each equation should be minimum, that is,

$$\sum_{i=1}^{m}\left|\sum_{j=1}^{n} a_{ij}x_j - b_i\right|^2 \text{ is minimum.}$$

Theorem 5. *Suppose X_0 is a least square solution. Then AX_0 is the orthogonal projection of b on the column-space of A.*

Proof. Let X_0 be the least-square approximation of $AX = b$. Then $\sum_{i=1}^{m}\left|\sum_{j=1}^{n} a_{ij}x_j - b_i\right|^2$ is minimum. For $Y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, $\sum_{i=1}^{m}\left|\sum_{j=1}^{n} a_{ij}y_j - b_i\right|^2 = ||AY - b||^2$. Thus $||AX_0 - b|| \leq ||AX - b||$ for all $X \in \mathbb{R}^n$

as $||AX_0 - b||$ is minimum. Recall that w_v is the orthogonal projection of v on to W if and only if $||v - w_v|| \leq ||v - w||$ for all $w \in W$. Take $V = \mathbb{R}^m$, $W = \{AX : X \in \mathbb{R}^n\}$ and $v = b$. Then AX_0 is the orthogonal projection of b on the column-space of A. $\square$

Theorem 6. *Let X_0 be a least-square approximation of $AX = b$ and $N(A)$ be the null space of A. Suppose S is the set of all least-square solutions of $AX = b$. Then $S = X_0 + N(A)$.*

Proof. Let $X \in X_0 + N(A)$. Then $X = X_0 + X_h$ so that $AX - b = AX_0 - b$. Thus $X \in S$. Now suppose $X \in S$. Then $||AX - b|| = ||AX_0 - b|| \Rightarrow ||(AX_0 - b) + A(X - X_0)||^2 = ||AX_0 - b||^2 \Rightarrow ||AX_0 - b||^2 + ||A(X - X_0)||^2 = ||AX_0 - b||$ since $A(X - X_0) \in C(A)$ and $(AX_0 - b) \perp Y$ for all $Y \in C(A)$. Therefore, $||A(X - X_0)|| = 0 \Rightarrow A(X - X_0) = 0 \Rightarrow X - X_0 \in N(A)$ so that $X = X_0 + (X - X_0) \in X_0 + N(A)$. $\square$

Application of Fundamental Theorem of Linear Algebra

Lemma 7. *Let $A \in M_{m\times n}(\mathbb{R})$. Then the $A^TAX = A^Tb$ is consistent for every $b \in \mathbb{R}^m$.*

Proof. It is enough to show that each A^Tb is in the column space of A^TA. By Fundamental Theorem of Linear Algebra, $\mathbb{R}^m = C(A) \oplus N(A^T)$. Thus, there exist $X \in \mathbb{R}^m$ and $Y \in N(A^T)$ such that $b = AX + Y$. Therefore, $A^Tb = A^T(AX) + A^TY = A^TAX + 0$. $\square$

Theorem 8. *Let $AX = b$ be an inconsistent system of linear equations and $X_0 \in \mathbb{R}^n$. Then X_0 is a least-square solution of $AX = b$ if and only if $A^TAX_0 = A^Tb$.*

Proof. Note that $N(A^T)^\perp = C(A)$. Then X_0 is a least-square solution if and only if $AX_0 - b \in C(A)^\perp$, that is, $(AX_0 - b) \in N(A^T) \Leftrightarrow A^T(AX_0 - b) = 0 \Leftrightarrow A^TAX_0 = A^Tb$.

$\square$

Remark 9. *For finding a least-square solution, one can solve the system $A^TAX = A^Tb$.*

Example 10. *Find a straight line $y = a + bx$ which fits best the given points $(1, 0), (2, 3), (3, 4), (4, 4)$ by least-square method.*

Solution: We get the following system of equations

$$\begin{aligned} a + b &= 0 \\ a + 2b &= 3 \\ a + 3b &= 4 \\ a + 4b &= 4 \end{aligned}$$

which is inconsistent. For finding a least-square solution, we will solve the system $A^TAX = A^Tb$, where $A \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$ and $b = \begin{pmatrix} 0 \\ 3 \\ 4 \\ 4 \end{pmatrix}$. Thus, $A^TA = \begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}$ and $A^Tb = \begin{pmatrix} 11 \\ 34 \end{pmatrix}$. Thus, $(A^TA|A^Tb =$ $\begin{pmatrix} 4 & 10 & | & 11 \\ 10 & 30 & | & 34 \end{pmatrix} \sim \begin{pmatrix} 1 & 3 & | & 34/10 \\ 4 & 10 & | & 11 \end{pmatrix} \sim \begin{pmatrix} 1 & 3 & | & 34/10 \\ 0 & -2 & | & -13/5 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & | & -1/2 \\ 0 & 1 & | & 13/10 \end{pmatrix}$. Thus, $y = -1/2 + 13/10x$ is a best fit.

For applying orthogonal projection method, $W = C(A) = \{(x+y, x+2y, x+3y, x+4y) \mid x, y \in \mathbb{R}\}$. Basis of W is $\{(1,1,1,1),(1,2,3,4)\}$. An orthogonal basis of W is $\{(1,1,1,1),(-3/2,-1/2,1/2,3/2)\}$, $||(1,1,1,1)||^2 = 4$ and $||(-3/2,-1/2,1/2,3/2)||^2 = 5$. Take $v = b = (0,3,4,4)$. Then $P_W(v) = 11/4(1,1,1,1) + 13/10(-3/2,-1/2,1/2,3/2) = 1/10(8,21,34,47)$. Then a least-square solution can be obtained by solving $AX = P_W(v)$ so that $\begin{pmatrix} 1 & 1 & | & 8/10 \\ 1 & 2 & | & 21/10 \\ 1 & 3 & | & 34/10 \\ 1 & 4 & | & 47/10 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & | & 8/10 \\ 0 & 1 & | & 13/10 \\ 0 & 2 & | & 26/10 \\ 0 & 3 & | & 39/10 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & | & 8/10 \\ 0 & 1 & | & 13/10 \\ 0 & 0 & | & 0 \\ 0 & 0 & | & 0 \end{pmatrix}$.
Thus $(-1/2, 13/10)$ is a least-square solution so that $y = -1/2x + 13/10$ is a best fit.

Lecture 20
Spectral Theorem

Definition 1 (Orthogonal Matrix). *A real square matrix is called orthogonal if $AA^T = I = A^TA$.*

Definition 2 (Unitary Matrix). *A complex square matrix is called unitary if $AA^* = I = A^*A$, where A^* is the conjugate transpose of A, that is, $A^* = \overline{A}^T$.*

Theorem 3. Let A be a unitary (real orthogonal) matrix. Then
(i) rows of A forms an orthonormal set;
(ii) columns of A forms an orthonormal set.

Remark 4. *1. P is orthogonal if and only if P^T is orthogonal.*
2. P is unitary if and only if P^* is unitary.
3. An orthogonal matrix (unitary) is invertible and its inverse is orthogonal (unitary).
4. Product of two orthogonal (unitary) matrices is orthogonal (unitary).

Theorem 5. The eigenvalues of a unitary matrix (an orthogonal matrix) has absolute value 1.

Proof: Let λ be an eigenvalue of a unitary matrix A. Then there exists a non-zero vector X such that $AX = \lambda X$. Thus, $(AX)^* = \bar{\lambda}X^* \Rightarrow (AX)^*(AX) = \bar{\lambda}X^*(\lambda X) \Rightarrow X^*A^*AX = \lambda\bar{\lambda}X^*X$. But $A^*A = I$, $(1 - |\lambda|^2)X^*X = 0$, *i.e.*, $|\lambda| = 1$.

Definition 6. A complex square matrix A is called a Hermitian matrix if $A = A^*$, where A^* is the conjugate transpose of A, that is, $A^* = \overline{A}^T$. A complex square matrix is called skew-Hermitian if $A = -A^*$.

Theorem 7. 1. The eigenvalues of a Hermitian matrix (real symmetric matrix) are real.
2. The eigenvalues of a skew-Hermitian matrix (real skew-symmetric matrix) are either purely imaginary or zero.

Proof: Let λ be an eigenvalue of a Hermitian matrix A. Then there exists a non-zero vector $X \in \mathbb{C}^n$ such that $AX = \lambda X$, multiplying both side by X^*, we get $X^*AX = \lambda X^*X$. Taking conjugate transpose both sides, we get $(X^*AX)^* = (\lambda X^*X)^* \Rightarrow X^*AX = \bar{\lambda}X^*X$. Thus we see that $\lambda X^*X = \bar{\lambda}X^*X$. Since $X \neq 0$, $X^*X = ||X||^2 \neq 0$ so that $\lambda = \bar{\lambda}$. For skew-Hermitian matrix, proceed in a similar way.

Theorem 8. Let A be a real symmetric matrix. Then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof: Let $\lambda_1 \neq \lambda_2$ be two eigenvalues of A and v_1 and v_2 be corresponding eigenvectors respectively. Then $Av_1 = \lambda_1 v_1 \Rightarrow v_1^TA^T = \lambda_1 v_1^T \Rightarrow v_1^TA^Tv_2 = \lambda_1 v_1^Tv_2$. Also $(Av_1)^Tv_2 = v_1^TA^Tv_2 = v_1^TAv_2 = \lambda_2 v_1^Tv_2$. Hence, $(\lambda_1 - \lambda_2)v_1^Tv_2 = 0$, and $\lambda_1 \neq \lambda_2$ so that $v_1^Tv_2 = 0 = \langle v_1, v_2\rangle \Rightarrow v_1 \perp v_2$.

Theorem 9. [**Spectral Theorem for a real symmetric matrix**] Let A be a real symmetric matrix. Then there exists an orthogonal matrix P such that $P^TAP = D$, where D is a diagonal matrix. In other words, a real symmetric matrix is orthogonally diagonalizable.

Proof: The proof is by induction on order of the matrix. The result holds for $n = 1$. Suppose the result holds for $(n-1) \times (n-1)$ symmetric matrix. Let A be a symmetric matrix of order $n \times n$. Note that A has real eigenvalues. Let $\lambda \in \mathbb{R}$ be one of the eigenvalue and $0 \neq X \in \mathbb{R}^n$ be a corresponding eigenvector with norm 1, then $AX = \lambda X$. Construct an orthonormal basis (by Gram-Schmidt process) $B = \{v_1, v_2, v_3, \ldots, v_n\}$, where $v_1 = X$ and $v_i \in \mathbb{R}^n$. Construct a matrix P such that the i-th column of P is v_i. Then P is an orthogonal matrix.

Note that the matrix $P^{-1}AP$ is symmetric and the first column of $P^{-1}AP$ is given by $P^{-1}AP(e_1)$, thus $P^{-1}A(Pe_1) = P^{-1}AX = P^{-1}\lambda X = \lambda e_1$. Therefore, the matrix can be represented as $P^{-1}AP = \begin{bmatrix} \lambda & 0 \\ 0 & C \end{bmatrix}$, where C is a symmetric matrix of order $(n-1) \times (n-1)$. Hence, by induction hypothesis, C is similar to a diagonal matrix, say D, *i.e.*, there is an orthogonal matrix Q such that $Q^{-1}CQ = Q^TCQ = D$. Let $R = P \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$. We claim that R is orthogonal and R^TAR is diagonal.

$$R^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix} P^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} P^T = R^T, \text{ and}$$

$$R^TAR = \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} P^TAP \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & Q^TCQ \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & D \end{bmatrix}.$$

Thus R is an orthogonal matrix such that R^TAR is diagonal. Therefore, A is orthogonally diagonalizable.
□

Theorem 10. Converse of the above theorem is also true, *i.e.*, if $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable, then A is symmetric.

Proof: Let A be a matrix which is orthogonally diagonalizable. Then there is an orthogonal matrix P s.t. $P^{-1}AP = P^TAP = D$, equivalentely, $A = PDP^{-1} = PDP^T$. This shows that $A^T = A$. Hence proved.

Example: Find an orthogonal matrix P and a diagonal matrix D such that $P^TAP = D$, where

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}.$$

The characteristic polynomial is $(x+1)^2(x-5)$. The eigenvalues are $5, -1, -1$. An eigenvector corresponding to $\lambda = 5$ is $v_1 = (1,1,1)$. The two independent eigenvectors corresponding to $\lambda = -1$ are $v_2 = (-1,0,1)$ and $v_3 = (-1,1,0)$. Thus, $B = \{v_1, v_2, v_3\}$ forms a basis of $\mathbb{R}^3$. To find an orthonormal basis, we apply Gram-Schmidt process on B. Thus

$w_1 = v_1$, $||w_1|| = \sqrt{3}$,
$w_2 = v_2$ (eigen vectors corresponding to distinct eigen values are orthogonal), $||w_2|| = \sqrt{2}$,
$w_3 = v_3 - \frac{\langle v_3, w_1 \rangle w_1}{||w_1||^2} - \frac{\langle v_3, w_2 \rangle w_2}{||w_2||^2} = (-1,1,0) - 0(1,1,1) - \frac{1(-1,0,1)}{2} = (-\frac{1}{2}, 1, -\frac{1}{2})$, $||w_3|| = \frac{\sqrt{6}}{2}$

Thus, $P = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Verify yourself that $P^TAP = D$.

Lecture 21
Decomposition of a Matrix in Terms of Projections

Here we discuss a special kind of linear maps (matrices), called projection and their properties. Further, we see that every diagonalizable matrix can be decomposed into projection matrices.

Definition 1. *Let V be a vector space over $\mathbb{F}$. A linear map $E : V \to V$ is called a projection if $E^2 = E$. A matrix M is called a projection matrix if $M^2 = M$, i.e., M is idempotent.*

Theorem 2. *Let $E : V \to V$ be a projection. Let R be the range of E and N be its null space. Then $V = R \oplus N$.*

Proof: It is easy to see that $R \cap N = \{0\}$. For $v \in V$, let $v = v - Ev + Ev \in N + R$.

Theorem 3. *Let R and N be subspaces of a vector space V such that $V = R \oplus N$. Then there is a projection map E on V such that the range of E is R and the null space of E is N.*

Proof: Define $E : V \to V$ as $E(r + n) = r$.

Definition 4. *A vector space V is said to be a direct sum of k subspaces $W_1, W_2, \ldots, W_k$ if $V = W_1 + W_2 + \cdots + W_k$ and $W_i \cap (W_1 + W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_k) = \{0\}$ for each i.*

Theorem 5. *If $V = W_1 \oplus W_2 \ldots \oplus W_k$, then there exist k linear maps $E_1, \ldots, E_k$ on V such that:*

?

1. *Each E_i is projection,*
2. *$E_iE_j = 0$ for all $i \neq j$,*
3. *$E_1 + \ldots + E_k = I$,*
4. *the range of E_i is W_i.*

Proof. Let $v \in V$. Then $v = w_1 + w_2 + \cdots + w_k$, where $w_i \in W_i$. Define $E_i : V \to V$ as $E_i(v) = E_i(w_1 + \ldots + w_k) = w_i$ for all i. Then E_i is linear with $E_i^2(v) = v$ for all $v \in V$. Also $E_iE_j = 0$ for all $i \neq j$ and $E_1 + \ldots + E_k = I$. By definition of E_i, range of E_i is W_i. $\square$

Lemma 6. *Let $A \in M_n(\mathbb{F})$. The matrix A is diagonalizable if and only if $\mathbb{F}^n = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$, where $\lambda_i \in \mathbb{F}$ and $\lambda_i \neq \lambda_j$ for $i \neq j$ and E_{λ_i} is the eigenspace of λ_i.*

Proof. Let A be diagonalizable. Recall that if B_i is a basis of the eigenspace E_{λ_i}, then $\cup_{i=1}^k B_i$ is a basis of $V = \mathbb{F}^n$. Thus $V = E_{\lambda_1} + \cdots + E_{\lambda_k}$. Let $v \in E_{\lambda_i} \cap (E_{\lambda_1} + E_{\lambda_2} + \cdots + E_{\lambda_{i-1}} + E_{\lambda_{i+1}} + \cdots + E_{\lambda_k})$. Then $Av = \lambda_i v$ and $v = v_1 + v_2 + \cdots + v_{i-1} + v_{i+1} + \cdots + v_k$, where $v_j \in E_{\lambda_j}$ and $j \neq i$. Then $Av = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_{i-1} v_{i-1} + \lambda_{i+1} v_{i+1} + \cdots + \lambda_k v_k$ so that $(\lambda_i - \lambda_1)v_1 + \cdots + (\lambda_i - \lambda_{i-1})v_{i-1} + (\lambda_i - \lambda_{i+1})v_{i+1} + \cdots + (\lambda_i - \lambda_k)v_k = 0$. If v is non-zero, not all v_i are zero. Note that if $v_j \neq 0$, it is an eigenvector corresponding to λ_j, but eigenvectors corresponding to distinct eigenvalues are independent, hence $\lambda_i = \lambda_j$ for some $j \neq i$, which is a contradiction. $\square$

Theorem 7. *Let A be a diagonalizable matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then A can be decomposed as a linear sum of idempotent (projection) matrices $E_1, \ldots, E_k$ given by $A = \lambda_1 E_1 + \ldots + \lambda_k E_k$.*

Proof: The matrix A is diagonalizable so that the minimal polynomial of A is $(x-\lambda_1)\dots(x-\lambda_k)$. Define

$$E_j = \frac{(A-\lambda_1 I)\dots(A-\lambda_{j-1}I)(A-\lambda_{j+1}I)\dots(A-\lambda_k I))}{(\lambda_j-\lambda_1)\dots(\lambda_j-\lambda_{j-1})(\lambda_j-\lambda_{j+1})\dots(\lambda_j-\lambda_k)}.$$

Let $v \in V$, then $v = v_1 + v_2 + \dots + v_k$, where $v_i \in E_{\lambda_i}$. Let $v_i \in E_{\lambda_i}$, then $E_j(v_i) = 0$ if $i \neq j$ and $E_j(v_j) = v_j$ so that $E_j(v) = E_j(v_1 + v_2 + \dots + v_k) = E_j(v_1) + E_j(v_2) + \dots + E_j(v_k) = v_j$. Thus E_j is a projection matrix. One can see that (i) $E_i^2 = E_i$, (ii) $E_iE_j = 0$ and $I = E_1 + \ldots + E_k$ (left to the reader to verify). Now $I = E_1 + E_2 + \dots + E_k$ so that $A = AE_1 + AE_2 + \dots + AE_k$. Then $Av = A(v_1 + v_2 + \dots + v_k) = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_k v_k = \lambda_1 E_1(v) + \lambda_2 E_2(v) + \dots + \lambda_k E_k(v)$ for all $v \in V$. Therefore, $A = \lambda_1 E_1 + \lambda_2 E_2 + \dots + \lambda_k E_k$.

Example: Check the diagonalizability of the given matrix $\begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}$. If diagonalizable, write the matrix as linear sum of projection matrices.

Solution: The characteristic polynomial $p(x) = (x-1)(x-2)^2$. Let $\lambda_1 = 1$ and $\lambda_2 = 2$. Then $GM(1) =$ and eigenvectors corresponding to 2 are $v_2 = (2,1,0)$ and $(2,0,1)$ so that $GM(2) = 2$. Hence, the matrix is diagonalizable. Then as per the above theory, $E_1 = (2I - A)$ and $E_2 = (A - I)$ and hence, $A = 1(2I - A) + 2(A - I)$. Verify yourself that $E_i^2 = E_i$ for $i = 1, 2$.

Lecture 22
Positive & Negative Definite Matrices & Singular Value Decomposition(SVD)

Definition 1. *Let A be a real symmetric matrix. Then A is said to be positive (negative) definite if all of its eigenvalues are positive (negative).*

Definition 2. *Let A be a real symmetric matrix. Then A is said to be positive (negative) semi-definite if all of its eigenvalues are non-negative (non-positive).*

Remark 3. *1. If A is positive definite, then $\det(A) > 0$ and $tr(A) > 0$.*
2. If A is negative definite matrix of order n, then $tr(A) < 0$. If n is even, $\det(A) > 0$ and if n is odd $\det(A) < 0$.
3. If A is positive semi-definite, then $\det(A) \geq 0$ and $tr(A) \geq 0$.
4. If A is negative semi-definite matrix of order n, then $tr(A) \leq 0$. If n is even, $\det(A) \geq 0$ and if n is odd $\det(A) \leq 0$.

Proposition 4. *Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then*
1. A is positive definite if and only if $X^TAX > 0$ for all $0 \neq X \in \mathbb{R}^n$.
2. A is negative definite if and only if $X^TAX < 0$ for all $0 \neq X \in \mathbb{R}^n$.

Proof. Let A be positive definite. Since A is a real symmetric matrix, A is orthogonally diagonalizable with positive eigenvalues. Therefore, $A = PDP^T$, where D is a diagonal matrix with entries as eigenvalues of A and P is an orthogonal matrix. Thus, $X^TAX = X^TPDP^TX = (P^TX)^TD(P^TX) = Y^TDY$, where $Y = P^TX \neq 0$. Let $Y = (y_1, y_2, \ldots, y_n)^T$. Then $X^TAX = Y^TDY = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 > 0$, where λ_i are eigenvalues of A.

Conversely, let $X^TAX > 0$ for all $X \in \mathbb{R}^n$. Let $\lambda \in \mathbb{R}$ be an eigenvalue of A and X_0 be an eigenvector corresponding to λ. Then $X_0^TAX_0 > 0 \Rightarrow \lambda X_0^TX_0 > 0$. Note that $X_0^TX_0 = \|X_0\|^2 > 0$ as $X_0 \neq 0$. Therefore, $\lambda > 0$. □

Proposition 5. *Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then*
1. A is positive definite if and only if $A = B^TB$ for some invertible matrix B.
2. A is positive semi-definite if and only if $A = B^TB$ for some matrix B.

Proof. Let A be a positive definite matrix. Then A is symmetric, by Spectral theorem, there exists an orthogonal matrix P such that $P^TAP = D$ with $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where λ_i's are eigenvalues of A. Here, $\lambda_i > 0$. Define $\sqrt{D} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$. Set $B = \sqrt{D}P^T$, then B is invertible and $B^TB = A$.

Conversely, $X^TAX = X^TB^TBX = (BX)^T(BX) = \|BX\|^2$. Therefore, for $X \neq 0$, $X^TAX > 0$. □

Let $A \in M_n(\mathbb{R})$. The leading principal minor D_k of A of order k, $1 \leq k \leq n$, is the determinant of the matrix obtained from A by deleting last $n-k$ rows and last $n-k$ columns of A.

Proposition 6. *Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then*
1. A is positive definite if and only if $D_k > 0$ for $1 \leq k \leq n$.
2. A is negative definite if and only if $(-1)^kD_k > 0$ for $1 \leq k \leq n$.
3. A is positive semi-definite, then $D_k \geq 0$ for $1 \leq k \leq n$. Show that the converse need not be true.
4. A is negative semi-definite, then $(-1)^kD_k \geq 0$ for $1 \leq k \leq n$. Show that the converse need not be true.

Proof. The prove for this result has been omitted. To see that converse is not true in case of (3), take $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1/2 \end{pmatrix}$. Then $D_1 = 1$, $D_2 = \det \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 0$ and $D_3 = \det(A) = 0$. The matrix is symmetric and $D_k \geq 0$ for $k = 1, 2, 3$. But $X^T AX = -2$ for $X = (1, 1, -2)^T$. Therefore, A is not positive semi-definite. □

Exercise 1. *Which of the following matrices is positive definite/negative definite/positive semi-definite/ negative semi-definite.*

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Singular-Value Decomposition

We know that every matrix is not diagonalizable and diagonalizability can be discussed only for square matrices. Here we discuss a decomposition of an $m \times n$ matrix which coincide with a known decomposition of a positive semi-definite matrix.

Let $A \in M_{m\times n}$. Then a decomposition of the form

$$A = U\Sigma V^T,$$

where $U \in M_m(\mathbb{R})$ and $V \in M_n(\mathbb{R})$ are orthogonal, and Σ is a rectangular diagonal matrix with non-negative real diagonal entries, is called Singular-Value Decomposition of A. The non-zero diagonal entries of Σ are called singular values of A.

When A is a positive semi-definite matrix, then SVD is nothing but $A = PDP^T$ for some orthogonal matrix P.

Theorem 7. *Let $A \in M_{m\times n}(\mathbb{R})$. Then A has a singular value decomposition.*

Proposition 8. *Let $A \in M_{m\times n}(\mathbb{R})$. Then*
1. *$A^T A$ is positive semi-definite.* nxn
2. *AA^T is positive semi-definite.* mxm
3. *If $m \geq n$, then $P^T(A^T A)P = D$ and $P'^T(AA^T)P' = D'$ for some orthogonal matrices $P \in M_n(\mathbb{R})$ and $P' \in M_m(\mathbb{R})$ with*

$$D' = \begin{pmatrix} D & 0_{m\times m-n} \\ 0_{m-n\times m} & 0_{m-n\times m-n} \end{pmatrix}.$$

X^T AA^T X

Proof. Note that $A^T A$ and AA^T are symmetric matrices. We claim that $X^T AX \geq 0$ for every $X \neq 0$. For $X \neq 0$, $X^T AA^T X = (A^T X)^T(A^T X) = \|A^T X\|^2 \geq 0$. Therefore, AA^T is positive semi-definite. Similarly for $A^T A$. Since the $A^T A$ and AA^T are symmetric, they are orthogonally diagonalizable. Therefore, $P^T(A^T A)P = D$ and $P'^T(AA^T)P = D'$ for some orthogonal matrices $P \in M_m(\mathbb{R})$ and $P' \in M_n(\mathbb{R})$. Recall that $p_{AA^T}x = x^{m-n}p_{A^T A}(x)$, where $p_{AA^T}(x)$ and $p_{A^T A}$ are the characteristic polynomial of AA^T and $A^T A$ respectively. Hence, $D' = \begin{pmatrix} D & 0_{m\times m-n} \\ 0_{m-n\times m} & 0_{m-n\times m-n} \end{pmatrix}$. □

Method to find SVD of A

Step 1: Find AA^T, which is positive semi-definite matrix. Therefore, we can find an orthogonal matrix $U \in M_m(\mathbb{R})$ such that

$$U^T(AA^T)U = D.$$

Note that columns of U are eigenvectors (orthonormal) of AA^T.

Step 2: Find A^TA, which is positive semi-definite matrix. We can find an orthogonal matrix $V \in M_n(\mathbb{R})$ such that

$$V^T(A^TA)V = D'.$$

Note that columns of V are eigenvectors (orthonormal) of A^TA.

Step 3: Define a rectangular diagonal matrix $\Sigma \in M_{m\times n}$ such that $\Sigma_{ii} = \sqrt{\lambda_i}$ for $i = 1, 2, \ldots, \min(m, n)$, where λ_i are the common eigenvalues of A^TA and AA^T. Note that non-zero diagonal entries σ_i are corresponding to non-zero eigenvalues of A^TA or AA^T.

Step 4: Verify that $U\Sigma V^T = A$.

Remark 9. *Let $A \in M_{m\times n}(\mathbb{R})$ and rank$(A) = r$. Let $U\Sigma V^T$ be a singular value decomposition of A. Let $U_1, U_2, \ldots, U_m$ are columns of U and $V_1, V_2, \ldots, V_n$ are columns of V. Then*
1. $\{U_1, U_2, \ldots, U_r\}$ is an orthonormal basis of column space(A).
2. $\{V_{r+1}, V_{r+2}, \ldots, V_n\}$ is an orthonormal basis of null space(A).
3. $\{V_1, V_2, \ldots, V_r\}$ is an orthonormal basis of Column space of (A^T) or row space of A.
4. $\{U_{r+1}, U_{r+2}, \ldots, U_n\}$ is an orthonormal basis of null space(A^T).

Proof. Note that $AV = U\Sigma \Rightarrow AV_j = \sigma_i U_j$ for $j = 1, 2, \ldots, r$ and $AV_j = 0$ for $j = r + 1, \ldots, n$. Since nullity of A is $n - r$ and $V_{r+1}, V_{r+2}, \ldots, V_n$ forms an orthonormal basis of $N(A)$. Since $\sigma_j > 0$ and $AV_j = \sigma_j U_j$, $U_j \in C(A)$ for $j = 1, 2, \ldots, r$. Thus $\{U_1, U_2, \ldots, U_r\}$ is an orthonormal basis of $C(A)$. Similarly, $A^TU = V\Sigma$ gives that first r columns of V forms a basis of the column space of A^T. □

Example 10. *Find SVD of $A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$.*

Solution: $AA^T = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ and $A^TA = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$. Then $U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Note that non-zero eigenvalue of A^TA is 2 (as non-zero eigenvalue of AA^T is 2) with eigenvectors $(0, 1, 0, 1)$ and $(1, 0, 1, 0)$ and the remaining eigenvalues of A^TA are all zero. The eigenvectors corresponding to 0 are $(1, 0, -1, 0)$ and $(0, 1, 0, -1)$. Thus $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \end{pmatrix}$. The rectangular diagonal matrix $\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{pmatrix}$.

Therefore, $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \end{pmatrix}$.

Remark: After finding U, one can find columns of V corresponding to non-zero eigenvalues by using the relation $V_i = \frac{1}{\sigma_i} A^T U_i$. The other columns of V can be found by finding vectors orthogonal to V_1, V_2 and to each other.

if some repeated eigenvalue, then we have to find the orthogonal eigenvector by Gram-Schmidt Process. Usually, eigenvalues will be distinct and since A^T A and AA^T are symmetric, distinct eigenvalues have orthogonal eigenvectors.

Lecture 23
Classification of Conics & Surfaces

Classification of Conics A conic is a curve in $\mathbb{R}^2$ which is represented by an equation of second degree in two variable, called quadratic curve. The general equation of such a conic (quadratic curve) is given by

$$\boxed{ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0} \tag{1}$$

where $a, b, h, g, f, c \in \mathbb{R}$ and $(a, b, h) \neq (0, 0, 0)$.

Then Equation (1) can be written as $\boxed{(x, y)\begin{pmatrix} a & h \\ h & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (2g, 2f)\begin{pmatrix} x \\ y \end{pmatrix} + c = 0}$. Here, $H(X) = (x, y)\begin{pmatrix} a & h \\ h & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = X^TAX$ is called the associated quadratic form of the conic (1), where $A = \begin{pmatrix} a & h \\ h & b \end{pmatrix}$ is a symmetric matrix. Suppose λ_1, λ_2 are eigenvalues of A and P is an orthogonal matrix such that $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = P^TAP$. Then Equation (1) can be written as $(x, y)P\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}P^T\begin{pmatrix} x \\ y \end{pmatrix} + +(2g, 2f)\begin{pmatrix} x \\ y \end{pmatrix} + c = 0$. Set $\begin{pmatrix} x' \\ y' \end{pmatrix} = P^T\begin{pmatrix} x \\ y \end{pmatrix}$. Then Equation (1) can be written as $\boxed{\lambda_1 x'^2 + \lambda_2 y'^2 + 2g'x' + 2f'y' + c' = 0.}$
If $\lambda_1, \lambda_2 \neq 0$, then equation can be reduced to the following form

$$\boxed{\lambda_1(x' + \alpha)^2 + \lambda_2(y' + \beta)^2 = \mu.}$$

If $\lambda_1 = 0$ and $\lambda_2 \neq 0$, the reduced equation is of the form $\lambda_2(y' + \beta)^2 = \gamma x + \mu$ (similarly when $\lambda_1 \neq 0, \lambda_2 = 0$). If $\lambda_1 = \lambda_2 = 0$, then $2g'x' + 2f'y' + c' = 0$.

Proposition 1. *Consider the quadratic $F(x, y) = ax^2 + 2hxy + by^2 + 2gx + 2fy + c$, for $a, b, c, g, f, h \in \mathbb{R}$. If $(a, b, h) \neq (0, 0, 0)$ then the conic $F(x, y) = 0$ can be classified as follows.*

λ_1	λ_2	μ	conic
+ve	+ve	+ve	ellipse
+ve, -ve	-ve, +ve	non-zero	hyperbola
+ve	+ve	-ve	no real curve exists
+ve	+ve	0	single point (-alpha, -beta)
-ve	-ve	0	single point (-alpha, -beta)
+ve, -ve	-ve, +ve	0	pair of straight lines
0	$\pm$ve		parabola ($\gamma \neq 0$) or single line ($\gamma = 0 = \mu$) or pair of parallel lines ($\mu\lambda_2 > 0$) or two imaginary lines ($\mu\lambda_2 < 0$)
$\pm$ve	0		similar as above
0	0		single straight line

Example 2. *Identify the conic $3x^2 - 2xy + 3y^2 - 8\sqrt{2}x + 10 = 0$*

Solution: The matrix form is $(x,y)\begin{pmatrix}3 & -1\\ -1 & 3\end{pmatrix}+(-8\sqrt{2},0)\begin{pmatrix}x\\ y\end{pmatrix}+10=0$. Eigenvalues of $A=\begin{pmatrix}3 & -1\\ -1 & 3\end{pmatrix}$ are $2,4$. The corresponding orthogonal matrix $P=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}$ such that $P^TAP=D$. Write $\begin{pmatrix}x\\ y\end{pmatrix}=P\begin{pmatrix}x'\\ y'\end{pmatrix}$, we get $(x',y')\begin{pmatrix}2 & 0\\ 0 & 4\end{pmatrix}\begin{pmatrix}x'\\ y'\end{pmatrix}+(-8\sqrt{2},0)\begin{pmatrix}\frac{x'-y'}{\sqrt{2}}\\ \frac{x'+y'}{\sqrt{2}}\end{pmatrix}+10=0$. By solving (expanding and making complete square), the reduced form is $2(x'-2)^2+4(y'+1)^2=2$, which represents an ellipse centered at $(2,-1)$.

Classification of Surfaces

A quadric surface is a surface in $\mathbb{R}^3$ described by a polynomial of degree 2 in three variables. A general equation of a surface is given by $F(x,y,z)=ax^2+by^2+cz^2+2hxy+2gxz+2fyz+2lx+2my+2nz+q$. The matrix form $F(x,y,z)=(x,y,z)\begin{pmatrix}a & h & g\\ g & b & f\\ g & f & c\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}+(2l,2m,2n)\begin{pmatrix}x\\ y\\ z\end{pmatrix}+q$. Let $A=\begin{pmatrix}a & h & g\\ g & b & f\\ g & f & c\end{pmatrix}$ and $\lambda_1,\lambda_2,\lambda_3$ be eigenvalues of A. Proceeding in a similar way as in the case of conics in $\mathbb{R}^2$, we get $F(x',y',z')=\lambda_1 x'^2+\lambda_2 y'^2+\lambda_3 z'2+l'x+m'y+n'z+q'$. If $\lambda_1,\lambda_2,\lambda_3\neq 0$, the equation can be reduced to the form $\lambda_1(x'+\alpha)^2+\lambda_2(y'+\beta^2)+\lambda_3(z'+\gamma)^2=\mu$. The classification of surfaces in $\mathbb{R}^3$ is as follows: If

λ_1	λ_2	λ_3	μ	conic
+ve	+ve	+ve	+ve	ellipsoid
+ve	+ve	-ve	+ve	hyperboloid of one sheet
+ve	-ve	-ve	+ve	hyperboloid of two sheet
+ve	+ve	+ve	0	single point
-ve	-ve	-ve	0	single point
+ve	+ve	-ve	0	cone
+ve	-ve	-ve	0	cone
+ve	+ve	0	+ve with coefficient of z is zero	elliptical cylinder
+ve	+ve	0	+ve with coefficient of z is non-zero	elliptical paraboloid
+ve	-ve	0	+ve with coefficient of z is zero	hyperbolic cylinder
+ve	-ve	0	+ve with coefficient of z is non-zero	hyperbolic paraboloid

two of the eigenvalues are zero, then the surface is either a parabolic cylinder or a pair planes or a singe plane.

Determine the following surface $F(x,y,z)=0$, where $F(x,y,z)=2x^2+2y^2+2z^2+2xy+2xz+2yz+4x+2y+4z+2$. Here $A=\begin{pmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{pmatrix}$, $b=\begin{pmatrix}2\\ 1\\ 2\end{pmatrix}$ and $q=2$. The eigenvalues of A are 4,1,1 and $P=\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}}\end{pmatrix}$ such that $P^TAP=D$, where $D=diag(4,1,1)$. Hence, $F(x,y,z)=0$ reduces to

$4\left(\frac{x+y+z}{\sqrt{3}}\right)^2+\left(\frac{x-y}{\sqrt{2}}\right)^2+\left(\frac{x+y-2z}{\sqrt{6}}\right)^2=-(4x+2y+4z+2)$. Further, we get $4\left(\frac{4(x+y+z)+5}{4\sqrt{3}}\right)^2+\left(\frac{x-y+1}{\sqrt{2}}\right)^2+\left(\frac{x+y-2z-1}{\sqrt{6}}\right)^2=9/12$. Equivalently, the surface can be written as $4(x'+5/4)^2+1(y'+1)^2+1(z'-1)^2=9/12$, where $x'=\frac{x+y+z}{\sqrt{3}}, y'=\frac{x-y}{\sqrt{2}}, z'=\frac{x+y-2z}{\sqrt{6}}$. Thus, the given equation describes an ellipsoid and the principal axes are $4(x+y+z)=-5$; $x-y=1$ and $x+y-2z=1$.

Lecture 24
Jordan-Canonical Form

We know that not every matrix is similar to a diagonal matrix. Here, we discuss the simplest matrix to which a square matrix is similar. This simplest matrix coincides with a diagonal matrix if the matrix is diagonalizable.

Definition 1. *A square matrix A is called* **block diagonal** *if A has the form*

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & A_k \end{pmatrix},$$

where A_i is a square matrix and the diagonal entries of A_i lie on the diagonal of A.

Definition 2. *Let $\lambda \in \mathbb{C}$. A* **Jordan block** $J(\lambda)$ *is an upper triangular matrix whose all diagonal entries are λ, all entries of the superdiagonal (entries just above the diagonal) are 1 and other entries are zero. Therefore,*

$$J(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix}.$$

Definition 3. *A* **Jordan form** *or* **Jordan-Canonical form** *is a block diagonal matrix whose each block is a Jordan block, that is, Jordan form is a matrix of the following form*

$$\begin{pmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & J_k \end{pmatrix}.$$

Definition 4. *Let $T : V \to V$ be a linear transformation and $\lambda \in \mathbb{C}$. A non-zero vector $v \in V$ is called a generalized eigenvector of T corresponding to Λ if $(T - \lambda I)^p(v) = 0$ for some positive integer p.*

The generalized eigenspace of T corresponding to λ, denoted by K_λ, is the subset of V defined by

$$K_\lambda = \{v \in V \mid (T - \lambda I)^p(v) = 0 \textit{ for some natural number } p\}.$$

Remark 5. *1. If $v \in V$ is a generalized eigenvector of a linear transformation T corresponding to $\lambda \in \mathbb{C}$, then λ is an eigenvalue of T.*
2. The generalized eigenspace K_λ is a subspace of V and $Tx \in K_\lambda$ for all $x \in K_\lambda$.
3. Let E_λ be the eigenspace corresponding to λ. Them $E_\lambda \subset K_\lambda$.

Theorem 6. *Let J be an $m \times m$ Jordan block with eigenvalue λ. Then characteristic polynomial of J is equal to its minimal polynomial, that is $p_J(x) = (x - \lambda)^m = m_J(x)$.*

Proof. Note that J is an upper-triangular matrix, hence the characteristic polynomial is $(x-\lambda)^m$ and the minimal polynomial is $(x-\lambda)^k$ for some $1 \le k \le m$. Here, we claim that $(J-\lambda I)^k \neq 0$ for $k < m$.

Observe that, $J - \lambda I = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$, $(J-\lambda I)^2 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$ and $(J-\lambda I)^{m-1} =$

multiplication wrong but understandable.

$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$ so that the minimal polynomial of J is $(x-\lambda)^m$. □

Remark 7. *1. If a matrix A is similar to a Jordan block of order m with eigenvalue λ, then there exist an invertible matrix P such that $P^{-1}AP = J$. Let X_i be the i-th column of P. Then $\{X_1, X_2, \dots, X_m\}$ is a basis of $\mathbb{R}^m$, which is called* **Jordan basis**.
2. The vector X_1 is an eigenvector corresponding to λ and $X_{j-1} = (A-\lambda)X_j$ for $j = 2, \dots, m$.

Theorem 8. *An $m \times m$ matrix A is similar to an $m \times m$ Jordan block J with eigenvalue λ if and only if there exist m independent vectors $X_1, X_2, \dots, X_m$ such that $(A-\lambda I)X_1 = 0$, $(A-\lambda I)X_2 = X_1, \dots, (A-\lambda I)X_m = X_{m-1}$.*

Example 9. *Consider $A = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$. Then the characteristic polynomial and minimal polynomial are the same which is $(x-2)^2$. Hence the matrix in not diagonalizable. Here, $(1, -1)$ is an eigenvector corresponding to 2. If J is the Jordan form of A, then we have a basis $\{X_1, X_2\}$ with respect to which the matrix representation of A is J. By previous theorem X_1 is an eigenvector and X_2 can be found by solving $(A-2I)X_2 = X_1$. Set $X_1 = (1,-1)$, then $X_2 = (1, 0)$. Now construct $P = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ and verify that $P^{-1}AP = J$, where $J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$.*

Theorem 10. *Let A be an $n \times n$ matrix with the characteristic polynomial $(x-\lambda_1)^{r_1} \cdots (x-\lambda_k)^{r_k}$, where λ_i's are distinct. Then A is similar to a matrix of the following form*

$$\begin{pmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & J_k \end{pmatrix},$$

where $J_1, J_2, \dots, J_k$ are Jordan blocks. The matrix J is unique except for the order of the blocks $J_1, J_2, \dots, J_k$.

Remark 11. *1. The sum of orders of the blocks corresponding to λ_i is r_i (the A.M.(λ_i)).*
2. The order of the largest block associated to λ_i is s_i, the exponent of $x - \lambda_i$ in the minimal polynomial of A.
3. The number of blocks associated with the eigenvalue λ_i is equal to the $GM(\lambda_i)$.
4. Knowing the characteristic polynomial and the minimal polynomial and the geometric multiplicity of each eigenvalue λ_i need not be sufficient to determine Jordan form of a matrix.

Example 12. *Let A be a matrix with characteristic polynomial $(x-1)^3(x-2)^2$ and minimal polynomial $(x-1)^2(x-2)$. Then we can find the Jordan form J of A by using above remarks,*
(i) The eigenvalue 1 appears on the diagonal 3 times, and 2 appears 2 times.
(ii) The largest Jordan block corresponding to $\lambda = 1$ is of order 2 (exponent of $(x-1)$ in the minimal polynomial), and the largest Jordan block corresponding to $\lambda = 2$ is of order 1.
(iii) The number number of Jordan blocks corresponding to $\lambda = 1$ is 2 where one block is of order 2 and other is of order 1. (iv) The number number of Jordan blocks corresponding to $\lambda = 2$ is 2 where both the blocks are of order 1. Therefore, the Jordan form of A is

$$\begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & & & \\ & (1) & & \\ & & (2) & \\ & & & (2) \end{pmatrix}.$$

Example 13. *Let A be a matrix with characteristic polynomial $(x-1)^3(x-2)^2$ and minimal polynomial $(x-1)^3(x-2)^2$. Then*
(i) The eigenvalue 1 appears on the diagonal 3 times, and 2 appears 2 times.
(ii) The largest Jordan block corresponding to $\lambda = 1$ is of order 3 (exponent of $(x-1)$ in the minimal polynomial), and the largest Jordan block corresponding to $\lambda = 2$ is of order 2.
(iii) The number number of Jordan blocks corresponding to $\lambda = 1$ is 1. (iv) The number number of Jordan blocks corresponding to $\lambda = 2$ is 1. Therefore, the Jordan form of A is

$$\begin{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} & \\ & \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \end{pmatrix}.$$

Example 14. minimal and characteristic are not always sufficient *Let A be a matrix with characteristic polynomial $(x-1)^4$ and minimal polynomial $(x-1)^2$. Then*
(i) The eigenvalue 1 appears on the diagonal 4 times.
(ii) The largest Jordan block corresponding to $\lambda = 1$ is of order 2 (exponent of $(x-1)$ in the minimal polynomial).
(iii) The number number of Jordan blocks corresponding to $\lambda = 1$ is GM(1) which is not known. Note that $GM(1) \leq 4$ as minimal polynomial confirms that A is not diagonalizable. Also, $GM(1) \neq 1$, if $GM(1) = 1$, the the Jordan matrix has only one block corresponding to $\lambda = 1$ which must be of order 4, which is not true.
(iv) Thus $GM(1) = 2$ or 3.

(v) If $GM(1) = 2$ the Jordan form of A is

$$\begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \\ & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{pmatrix}.$$

(vi) If $GM(1) = 3$, the Jordan form of A is

$$\begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & & \\ & (1) & \\ & & (1) \end{pmatrix}.$$

Example 15. Possible Jordan forms for a given characteristic polynomial *Let A be a matrix with characteristic polynomial $(x-1)^3(x-2)^2$. Then choices of minimal polynomials are*
(i)$(x-1)(x-2)$, then Jordan form is the diagonal matrix.
(ii) $(x-1)^2(x-2)$, Example 12.

(iii) $(x-1)^3(x-2)$, the Jordan form is $\begin{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} & & \\ & (2) & \\ & & (2) \end{pmatrix}$.

(iv) $(x-1)(x-2)^2$, $\begin{pmatrix} (1) & & & \\ & (1) & & \\ & & (1) & \\ & & & \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \end{pmatrix}$.

(v) $(x-1)^2(x-2)^2$, $\begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & & \\ & (1) & \\ & & \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \end{pmatrix}$.

(vi) $(x-1)^3(x-2)^2$, Example 13.

Example 16. (**minimal, characteristic and $GM(\lambda)$ are not always sufficient)** *Let A be a matrix with characteristic polynomial $(x-1)^7$ and minimal polynomial $(x-1)^3$ and $GM(1) = 3$. Then there are two possible Jordan forms (write the corresponding Jordan forms yourself!):*
(i) One Jordan block of order 3 and other two blocks of order 2.
(ii) Two Jordan blocks of order 3 and one of order 1.

Example 17. Find a Jordan basis *Let $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix}$. The characteristic polynomial of A is $(x-1)^3$ and $A - I = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$. Thus nullity$(A-I)$ is 1=GM(1). Therefore, the Jordan form of A is $J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. The problem is to find a Jordan basis or a matrix P such that $P^{-1}AP = J$. $P = [X_1\, X_2\, X_3]$, where $(A-I)X_1 = X_1$, $(A-I)X_2 = X_1$, $(A-I)X_3 = X_2$. On solving, we get $X_1 = (1, 0, -1)$, $X_2 = (1, 1, -1)$ or $(-1, 1, 1)$ and $X_3 = (1, 1, 0)$ or $(0, 1, 1)$. Hence, $P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 0 \end{pmatrix}$.*

Example 18. *(***Finding a Jordan basis is not always straight forward***)*

Let $A = \begin{pmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$. *The characteristic polynomial of* A *is* $(x-2)^3$ *and* $A - 2I = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$.

Thus nullity$(A - I2)$ *is 2=GM(2). Therefore, the Jordan form of* A *is* $J = \begin{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} & \\ & (2) \end{pmatrix}$. *The problem is to find a Jordan basis or a matrix* P *such that* $P^{-1}AP = J$. $P = [X_1\, X_2\, X_3]$. *Here, we get an eigenvector* (x, y, z) *satisfies* $x + y = 0$, *two independent eigenvectors are* $(0, 0, 1)$ *and* $(-1, 1, 0)$. *Note that each eigenvector corresponds to a Jordan block. Thus, set* $X_1 = (0, 0, 1)$, $(A - 2I)X_2 = X_1$, $X_3 = (-1, 1, 0)$ *or . But,* $(A - 2I)X_2 = X_1 \Rightarrow \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ *which is an inconsistent system. Similarly,* $(A - 2I)X_2 = X_1$, *where* $X_1 = (-1, 1, 0)$ *is inconsistent. For finding a Jordan basis, we will change the eigenvector, let* $X_1 = (-1, 1, -1)$, *then* $X_2 = (0, -1, 0)$, *and* $X_3 = (-1, 1, 0)$ *Hence,* $P = \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 1 \\ -1 & 0 & 0 \end{pmatrix}$.

each independent eigenvector obtained on solving (A-lambdaI)X = 0 is associated to each Jordan Block of that eigenvalue. But if in a jordan block, other eigenvectors have to be found, use the relation (A-lambdaI) X_n = Xn-1