MTH 100: Lecture 21

Example of another infinite dimensional vector space:

Ex: Let $C[a,b]$ be the vector space of all real valued continuous functions defined on $[a,b]$.

Question: Is $C[a,b]$ finite dimensional?

Answer: No. $C[a,b]$ is infinite dimensional.

Assume BWOC that $C[a,b]$ is finite dimensional.

Let $P[a,b]$ be the set of all (real valued) polynomials with domain $[a,b]$.

Now $P[a,b] \subset C[a,b]$

Furthermore $P[a,b]$ is a subspace of $C[a,b]$.
(check!)

Now the space $P[a,b]$ is infinite dimensional.
(The proof essentially uses the same argument we used to prove that $\mathbb{R}[t]$ is infinite dimensional)

Now $P[a,b]$ is a subspace of $C[a,b]$.

Thus if $C[a,b]$ is finite dimensional, then $P[a,b]$ will also be finite dimensional. — a contradiction.

Hence $\dim(C[a,b]) = \infty$.

<u>Proof of the fact that $P[a,b]$ is infinite dimensional:</u>

Suppose BWOC that $P[a,b]$ is finite dimensional. Then it has a finite Basis, say $\{p_1(x), p_2(x), \ldots, p_k(x)\}$

Let $N = \max\{\deg p_1, \deg p_2, \ldots, \deg p_k\}$

and let $p(x) = x^{N+1}$

Then $p(x)$ can't be written as a linear combination of $p_1, p_2, \ldots, p_k$ because any linear combination of $p_1, p_2, \ldots, p_k$ will be a polynomial of degree $\leq N$

and $\deg p(x) = N+1$,

a contradiction

Hence $P[a,b]$ is infinite dimensional.

Note: For the space $\mathbb{R}[t]$, $1, t, t^2, \ldots, t^n, \ldots$ is a basis because these are linearly independent and any polynomial can be written as a finite linear combination of these polynomials.

If $p(t) \in \mathbb{R}[t]$ and $\deg p(t) = N \ (< \infty)$ then there exist scalars $c_0, c_1, \ldots, c_N$ such that $p(t) = \sum_{i=0}^{N} c_i t^i$

Sum of Subspaces:

Definition: Let U and W be subspaces of the vector space V.

Then the sum of U and W is defined by

$$U + W = \{u + w : u \in U, w \in W\}$$

Furthermore,

- $U + W$ is a subspace of V.
- In fact $U + W$ is the smallest subspace of V containing U and W.

Proposition: If U and W are finite-dimensional subspaces of the vector space V, then

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$$

Proof: • If either U or $W = \{0\}$, the result is obvious.

• Now let $B = \{k_1, k_2, \ldots, k_m\}$ be a basis of $U \cap W$

(If $U \cap W = \{0\}$, this step is not needed)

Since $U \cap W \subseteq U$, we can expand B to a basis B_1 of U, by adjoining vectors $u_1, u_2, \ldots, u_n$

i.e. $B_1 = \{k_1, k_2, \ldots, k_m, u_1, \ldots, u_n\}$, $m \geq 0$, $n \geq 0$

Similarly since $U \cap W \subseteq W$, we can expand B to a Basis B_2 of W, by adjoining the vectors $w_1, \ldots, w_p$

i.e. $B_2 = \{k_1, k_2, \ldots, k_m, w_1, \ldots, w_p\}$ $m \geq 0$, $p \geq 0$

Let $C = B \cup B_1 \cup B_2 = \{k_1, k_2, \ldots, k_m, u_1, \ldots, u_n, w_1, \ldots, w_p\}$

We claim that C is a basis for $U+W$

So, we need to prove that

(1) $\operatorname{Span} C = U+W$ (2) C is linearly independent.

① Let $v \in U+W$. Then $v = u + w$ where $u \in U$ $w \in W$

Then there exist scalars $c_1, \dots, c_m, d_1, \dots, d_n, f_1, \dots f_m, g_1, \dots, g_p \in F$ such that $u = c_1 k_1 + \dots + c_m k_m + d_1 u_1 + \dots + d_n u_n$

$w = f_1 k_1 + \dots + f_m k_m + g_1 w_1 + \dots + g_p w_p$

So, $v = u + w = (c_1 + f_1) k_1 + \dots + (c_m + f_m) k_m$
$+ d_1 u_1 + \dots + d_n u_n + g_1 w_1 + \dots + g_p w_p$

Thus v is a linear combination of the elements of C.

Hence $U+W = \operatorname{Span}(C)$

② Now suppose

$c_1 k_1 + \dots + c_m k_m + d_1 u_1 + \dots + d_n u_n + g_1 w_1 + \dots + g_p w_p = \bar{0}$ .... ①

Then $c_1 k_1 + \dots + c_m k_m + d_1 u_1 + \dots + d_n u_n$

$= -g_1 w_1 - \dots - g_p w_p$ .... ②

Now the L.H.S. of ② is a vector in U and the R.H.S. of ② is a vector in W and so it is in $U \cap W$.

Hence we can write

$c_1 k_1 + \dots + c_m k_m + d_1 u_1 + \dots + d_n u_n$

$= f_1 k_1 + \dots + f_m k_m$ where $f_1, \dots, f_m \in F$

$\Rightarrow (c_1 - f_1) k_1 + \dots + (c_m - f_m) k_m + d_1 u_1 + \dots + d_n u_n = \bar{0}$

Since $\{k_1, \dots, k_m, u_1, \dots, u_n\}$ is a basis for U,

it is linearly independent.

So, $d_1 = d_2 = \dots = d_n = 0$

Then ① becomes

$$c_1 k_1 + \cdots + c_m k_m + g_1 w_1 + \cdots + g_p w_p = \bar{0}$$

Since $\{k_1, \ldots, k_m, w_1, \ldots, w_p\}$ is a basis for W, it is linearly independent and

therefore $c_1 = \cdots = c_m = g_1 = \cdots = g_p = 0$

Hence C is linearly independent and so, C is a basis for $U+W$.

Now $\dim U + \dim W - \dim(U \cap W)$

$$= (m+n) + (m+p) - m$$

$$= \cancel{m} + n + m + p - \cancel{m}$$

$$= m + n + p$$

$$= \dim(U+W)$$

Therefore $\boxed{\dim(U+W) = \dim U + \dim W - \dim(U \cap W)}$

MTH 100 : Lecture 22

From Last time:

Sum of two Subspaces:

If U and W are two subspaces of a vector space V, then their sum $U+W$ is defined by:

$$U+W = \{u+w : u \in U, w \in W\}$$

$U+W$ is a subspace of V.

- If U and W are finite dimensional subspaces of a vector space V, then

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$$

Direct Sums

Definition: V is said to be the direct sum of the subspaces U and W if every vector $v \in V$ is <u>uniquely</u> expressible in the form $v = u + w$ where $u \in U$, $w \in W$.

Notation: We will use the notation $V = U \oplus W$ to indicate that V is the direct sum of U and W.

Proposition: If U and W are subspaces of a vector space V, then $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$

Proof: Exercise (Try it!!)

Remark: The subspace W in the above is often called a Complement or Complementary subspace of U.

Corollary: If V is the direct sum of the finite dimensional subspaces U and W, then $\dim V = \dim(U \oplus W) = \dim U + \dim W$

Fundamental Subspaces

Definition: The null space of an $m \times n$ matrix A is the set of all solutions to the homogeneous system $Ax = 0$.

It is denoted by Nul A.

Thus $\text{Nul } A = \{ x \in \mathbb{R}^n : Ax = 0 \}$

So, Nul A is a subset of $\mathbb{R}^n$.

Proposition: The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$.

Proof: Let $\bar{0} \in \mathbb{R}^n$ be the zero vector of $\mathbb{R}^n$.

- Now $A \underbrace{\bar{0}}_{\text{(zero vector in } \mathbb{R}^n)} = \underbrace{\bar{0}}_{\text{(zero vector in } \mathbb{R}^m)}$ and so $\bar{0} \in \text{Nul } A$

- If $\bar{u}, \bar{v} \in \text{Nul } A$, then $A\bar{u} = \bar{0}$, $A\bar{v} = \bar{0}$

  Then $A(\bar{u} + \bar{v}) = A\bar{u} + A\bar{v} = \bar{0} + \bar{0} = \bar{0}$

  $\Rightarrow \bar{u} + \bar{v} \in \text{Nul } A$

- Finally if $\bar{u} \in \text{Nul } A$ and $c \in \mathbb{R}$ (any scalar)

  then $A(c\bar{u}) = c(A\bar{u}) = c(\bar{0}) = \bar{0} \Rightarrow c\bar{u} \in \text{Nul } A$

Therefore Nul A is a subspace of $\mathbb{R}^n$.

Remark:

- We need to take a homogeneous system of equations to get a subspace.
- The solution set of a non-homogeneous system is not a subspace.
- Nul A is defined implicitly.
  To describe Nul A explicitly, we need to solve the system of linear equation
  $$A\bar{x} = \bar{0}.$$

How to find a Basis for Nul A:

- Reduce A to an RREF matrix.
- Express the solution vector of the simplified system as a linear combination where the coefficients are the free variables.
- The spanning set produced by this method is a basis for Nul A.

Remark: Either Nul A is the zero subspace

or $\dim(\text{Nul } A) =$ Number of free variables in the solution.

Column Space:

Let $A_{m\times n} = [a_1, a_2, \dots, a_n]$, $a_i \in \mathbb{R}^m$ for $i = 1, 2, \dots, n$

Then column space of A (denoted by $\text{Col } A$) is the set of all linear combinations of the columns of A.

i.e. $\text{Col } A = \text{Span}\{a_1, a_2, \dots, a_n\}$

Proposition: $\text{Col } A$ is a subspace of $\mathbb{R}^m$.

Proof: Since A is an $m\times n$ matrix, its columns are vectors in $\mathbb{R}^m$.
Since $\text{Col } A$ is the span of the columns of A, it is a subspace of $\mathbb{R}^m$.

Remark: $\boxed{\text{Col } A = \{b \in \mathbb{R}^m : b = Ax \text{ for some } x \in \mathbb{R}^n\}}$

If $b \in \text{Col } A$, b can be written as
$b = x_1 a_1 + x_2 a_2 + \dots + x_n a_n$ where x_i's are scalars for $i = 1, 2, \dots, n$

Hence
$$b = x_1 a_1 + x_2 a_2 + \dots + x_n a_n$$
$$= [a_1, a_2, \dots, a_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = Ax$$
where,
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

Basis for col A:

Proposition: The pivot columns of a matrix A form a basis for col A.

Proof: Any linear dependence relationship between the columns of A can be written in the form $Ax = 0$

(Note that $x_1 a_1 + \cdots + x_n a_n = 0 \Rightarrow [a_1, \cdots, a_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 0$

$$\Rightarrow \quad Ax = 0 \text{)}$$

When the matrix A is row reduced to R, the columns of A change but the equation $Rx = 0$ has the same set of solutions as $Ax = 0$.

Thus row reduction does not change the dependence relations between the columns.

The pivot columns of A must be linearly independent since the pivot columns R are linearly independent.

Also, non-pivot columns are linear combinations of the preceding (i.e. left) pivot columns.

Hence Pivot columns of A form a basis for Col A.

(QED)

Note: We must take the columns of A for Basis (Not of its RREF matrix R)

Comparison between Nul A and Col A:

Nul A	Col A
• Nul A is a subspace of $\mathbb{R}^n$	• Col A is a subspace of $\mathbb{R}^m$
• Nul A is defined implicitly	• Col A is defined explicitly.
• To find vectors in Nul A, we have to solve an equation	• Vectors in Col A can be found directly.
• There is no obvious relation between Nul A and entries of A.	• There is a definite relation between Col A and entries of A.
• If $v \in$ Nul A, $Av = 0$	• If $v \in$ Col A, the system $Ax = v$ is consistent.
• Given a specific vector v, we can easily test whether it is in Nul A.	• Given a specific vector v, to test whether it is in Col A, we have to solve an equation.
• Nul $A = \{0\}$ if and only if $Ax = 0$ has only the trivial solution.	• Col $A = \mathbb{R}^m$ if and only if $Ax = b$ has a solution for every $b \in \mathbb{R}^m$.

Row space: Let A be an $m \times n$ matrix. The Row space of A (denoted by Row A) is the set of all linear combinations of the rows of A.

If $A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$, then Row $A = \text{Span}\{r_1, r_2, \ldots, r_m\}$

Proposition: Row A is a subspace of $\mathbb{R}^n$.

Proposition: Row equivalent matrices have the same row space.

Proof: Elementary row operations replace rows of the original matrix by rows which are the same or linearly dependent on them.

Hence row space does not get enlarged by row operations.

Thus if B is obtained from A by an elementary row operation, then Row $B \subseteq$ Row A

But since elementary row operations are reversible, we also have Row $A \subseteq$ Row B

Therefore $\boxed{\text{Row } B = \text{Row } A}$

(QED)

How to find a Basis for Row A:

- Given a matrix A, reduce it to an RREF matrix R.
- The non-zero rows of R are linearly independent and they form a Basis for the row space of R and also for the row space of A.

Alternate Method:

- Since the rows of A are the columns of A^T (Transpose of A), we can find a Basis for Row A by using the method to find a Basis for Col A^T.
- This method can be used to find a Basis for Row A consisting of actual rows of A.

Ex: Let $A = \begin{bmatrix} 1 & 3 & 2 & -3 \\ 2 & 6 & 4 & -6 \\ 3 & 9 & 7 & -11 \\ 8 & 24 & 9 & -10 \end{bmatrix} = [v_1, v_2, v_3, v_4]$ (say)

Find a Basis for Nul A, a Basis for Col A and a Basis for Row A.

$$A = \begin{bmatrix} 1 & 3 & 2 & -3 \\ 2 & 6 & 4 & -6 \\ 3 & 9 & 7 & -11 \\ 8 & 24 & 9 & -10 \end{bmatrix} \xrightarrow[\substack{R_3 \to R_3 - 3R_1 \\ R_4 \to R_4 - 8R_1}]{R_2 \to R_2 - 2R_1} \begin{bmatrix} 1 & 3 & 2 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 7 & -14 \end{bmatrix}$$

$$\xrightarrow{R_4 \to R_4 + 7R_3} \begin{bmatrix} 1 & 3 & 2 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \longleftrightarrow R_3} \begin{bmatrix} 1 & 3 & 2 & -3 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{R_1 \to R_1 - 2R_2} \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R = \begin{bmatrix} r_1 \\ r_2 \\ 0 \\ 0 \end{bmatrix} \quad \text{(say)}$$

$$\begin{array}{cccc} \downarrow & & \downarrow & \\ \text{Basic} & & \text{Basic} & \end{array}$$

The corresponding system $R\bar{x} = \bar{0}$ is

$$\left.\begin{aligned} x_1 + 3x_2 + x_4 &= 0 \\ x_2 &= x_2 \\ x_3 - 2x_4 &= 0 \\ x_4 &= x_4 \end{aligned}\right\} \Rightarrow \left.\begin{aligned} x_1 &= -3x_2 - x_4 \\ x_2 &= x_2 \\ x_3 &= 2x_4 \\ x_4 &= x_4 \end{aligned}\right\}$$

Thus $\bar{x} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} = x_2 u_1 + x_4 u_2$ (say)

- Then a Basis for Nul $A = \{u_1, u_2\}$

$$= \left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}$$

- A Basis for Col $A = \{v_1, v_3\}$

$$= \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 7 \\ 9 \end{bmatrix} \right\}$$

- A Basis for Row $A = \{r_1, r_2\}$

$$= \left\{ \begin{bmatrix} 1 & 3 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -2 \end{bmatrix} \right\}$$

MTH 100 : Lecture 23

Last time: For any $m \times n$ matrix A, we defined Nul A, Col A and Row A.

- We have seen how to find Bases for all these three spaces.
- Note that there is no containment relationship between Nul A, Col A and Row A.

  In general Nul A and Col A are not even subspaces of the same space because Col $A \subseteq \mathbb{R}^m$ and Nul $A \subseteq \mathbb{R}^n$.

The Rank Theorem:

Definition: If A is an $m \times n$ matrix, the column rank of A is defined to be $\boxed{\dim(\text{Col } A)}$.

Similarly, the row rank of A is defined to be $\boxed{\dim(\text{Row } A)}$.

- The nullity of A is defined to be $\boxed{\dim(\text{Nul } A)}$.

Ex: In the last example of Lecture 24,

A is a 4×4 matrix.

row rank $= 2$, Column rank $= 2$, nullity $= 2$

Theorem: (The Rank Theorem for Matrices):

(a) The row rank and column rank of a matrix A are equal. This number is called the rank of A.

(b) The rank of A is equal to the number of pivot positions in the RREF matrix obtained from A.

(c) $\text{rank}(A) + \text{nullity}(A) = n$ = number of columns of A.

Sketch of a Proof:

- (a) and (b) follow from our discussion of finding the Basis of Col A and Row A. In each case, the number of basis vectors corresponded to the number of pivot elements in the RREF matrix R of a given matrix A.

- For (c),
  Pivot columns of R will correspond to a basis of Col A (leading variables of the homogeneous system).
- The remaining columns correspond to a basis of Nul A (free variables of the homogeneous system).

Since, the total number of columns $= n$
$=$ number of variables,

we get

$n =$ number of basis vectors in Col A
$\quad +$ number of basis vectors in Nul A

$\Rightarrow \boxed{n = \text{rank}(A) + \text{nullity}(A)}$ (QED)

<u>Note</u>

(1) Col $A = \mathbb{R}^m$ if and only if the system $Ax = b$ has a solution for each $b \in \mathbb{R}^m$.
(This follows from the description of Col A)

(2) An $m \times m$ matrix A is invertible if and only if its columns form a basis of $\mathbb{R}^m$.
(This follows from Note (1) above and part (d) of the first Theorem of the course.)

<u>Corollary to Rank Theorem:</u>
A square $m \times m$ matrix A is invertible if and only if $\text{rank}(A) = m$ (or equivalently nullity $= 0$)

- In view of today's discussion, an <u>extended version</u> of the <u>first theorem of our course</u> can be given in the following way.

Theorem:

The following are equivalent for an $m \times m$ square matrix A.

(a) A is invertible.

(b) A is row equivalent to the identity matrix.

(c) The homogeneous system $Ax = 0$ has only the trivial solution.

(d) The system of equations $Ax = b$ has at least one solution for every $b \in \mathbb{R}^m$.

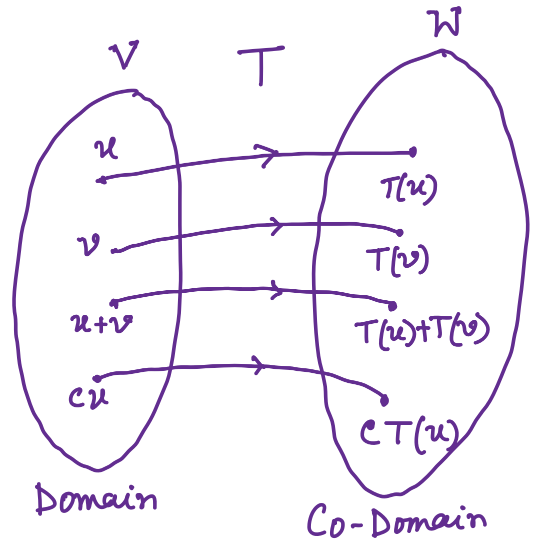
(e) $\text{Nullity}(A) = 0$

(f) $\text{Rank}(A) = m$

(g) The columns of A form a basis for $\mathbb{R}^m$.

(h) $\text{Det}\, A \neq 0$

Linear Transformations:



Definition: A map or function $T : V \to W$ from a vector space V to a vector space W is called a Linear Transformation (or briefly linear)

if (1) $T(u+v) = T(u) + T(v) \quad \forall\, u, v \in V$

(2) $T(cu) = cT(u) \quad \forall\, u \in V$ and $\forall\, c \in F$

(F is the scalar field)

Note: (1) The space W (the Co-Domain) may be the space V or a subspace of V or may be an entirely different space (but over the same field F).

(2) We may write either $T(v)$ or Tv to indicate the image of the vector v under the transformation T.

(3) Some books use the term homomorphism for a linear transformation (map or function) from a vector space V to a vector space W.

Examples:

(1) The zero transformation $0 : V \longrightarrow W$ defined by $0(u) = \bar{0}$ (zero vector in W) $\forall\, u \in V$

$$\begin{aligned} 0(u+v) &= \bar{0} = \bar{0}+\bar{0} = 0(u)+0(v) \\ 0(cu) &= \bar{0} = c.\bar{0} = c\,0(u) \quad \forall\, u,v \in V,\ \forall\, c \in \mathbb{R} \end{aligned}$$

(2) The identity transformation $I : V \longrightarrow V$ defined by $I(u) = u \ \forall\, u \in V$

$$I(u+v) = u+v = I(u)+I(v)$$
$$I(cu) = cu = cI(u) \quad \forall\, u,v \in V \ \forall\, c \in \mathbb{R}$$

(3) Projection: Define the function $P_i : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by

$$P_i(x_1, x_2, \ldots, x_i, \ldots, x_n) = (0, 0, \ldots, 0, x_i, 0, \ldots, 0)$$

(all coordinates other than the i-th coordinate are replaced by 0.)

- Then P_i is a linear transformation.
  We can extend this idea by projecting onto any selection of coordinates.

To show that P_i is linear: If $(x_1, x_2, \dots, x_i, \dots, x_n), (y_1, y_2, \dots, y_i, \dots y_n) \in \mathbb{R}^n$

then $P_i\left[(x_1, x_2, \dots, x_i, \dots, x_n) + (y_1, y_2, \dots, y_i, \dots, y_n)\right]$

$$= P_i\,(x_1+y_1, x_2+y_2, \dots, x_i+y_i, \dots, x_n+y_n)$$

$$= (0, 0, \dots, 0, x_i+y_i, 0, \dots, 0)$$

$$= (0, 0, \dots, 0, x_i, 0, \dots, 0) + (0, 0, \dots, 0, y_i, 0, \dots, 0)$$

$$= P_i\,(x_1, x_2, \dots, x_i, \dots, x_n) + P_i\,(y_1, y_2, \dots, y_i, \dots, y_n)$$

Now if $c \in \mathbb{R}$ and $(x_1, x_2, \dots, x_i, \dots x_n) \in \mathbb{R}^n$

then $P_i\left[c\,(x_1, x_2, \dots, x_i, \dots, x_n)\right]$

$$= P_i\,(cx_1, cx_2, \dots, cx_i, \dots, cx_n)$$

$$= (0, 0, \dots, 0, cx_i, 0, \dots, 0)$$

$$= c\,(0, 0, \dots, 0, x_i, 0, \dots, 0)$$

$$= c\,P_i\,(x_1, x_2, \dots, x_i, \dots, x_n)$$

Therefore P_i is a linear transformation.

MTH 100 : Lecture 24

Ex: Fix $1 \leq i \leq n$

Let $P_i : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be defined by

$$P_i(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots x_n) = (0, 0, \ldots, 0, x_i, 0, \ldots 0)$$

Then P_i is a Linear transformation:

- Let $u = (x_1, x_2, \ldots, x_n)$ and $v = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$

Now $P_i(u+v) = P_i((x_1, \ldots, x_n) + (y_1, \ldots, y_n))$

$$= P_i((x_1 + x_2, \ldots, x_n + y_n))$$

$$= (0, \ldots, 0, x_i + y_i, 0, \ldots, 0)$$

$$= (0, \ldots, 0, x_i, 0, \ldots, 0) + (0, \ldots, 0, y_i, 0, \ldots, 0)$$

$$= P_i(u) + P_i(v)$$

- Next let $u = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and let $c \in \mathbb{R}$

Then $P_i(cu) = P_i(c(x_1, \ldots, x_n)) = P_i((cx_1, \ldots, cx_n))$

$$= (0, \ldots, 0, cx_i, 0, \ldots, 0) = c(0, \ldots, 0, x_i, 0, \ldots, 0)$$

$$= cP_i(u)$$

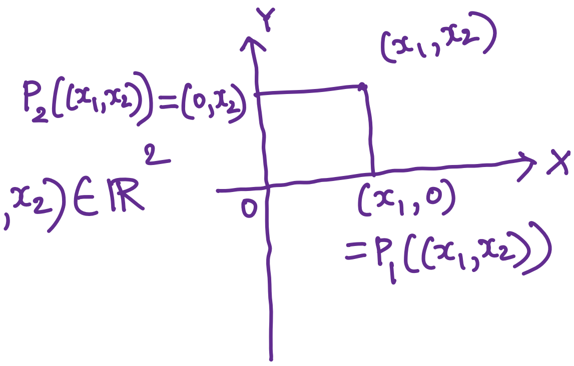
Therefore P_i is a linear transformation.

Ex: Let us take $n=2$.

Then we will get two linear transformations

$$P_1, P_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$\left.\begin{aligned} P_1((x_1,x_2)) &= (x_1,0) \\ \text{and } P_2((x_1,x_2)) &= (0,x_2) \end{aligned}\right\} \forall (x_1,x_2)\in\mathbb{R}^2$$



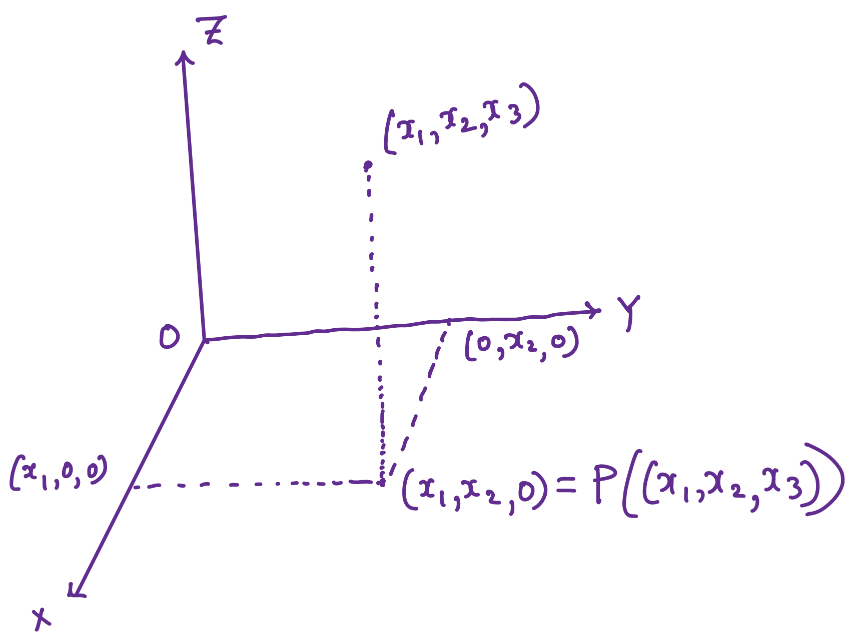
Ex: Look at the transformation for $\mathbb{R}^3$

Ex: Define $P_{12} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by

$$P_{12}((x_1,x_2,x_3)) = (x_1,x_2,0)$$

Show that P_{12} is a linear transformation

Similarly define transformations P_{23} and P_{31}.



<u>Ex:</u> Fix i, j such that $1 \leq i < j \leq n$

Define $P_{ij} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by

$$P_{ij}((x_1, \dots, x_i, \dots, x_j, \dots, x_n))$$
$$= (0, \dots, 0, x_i, 0, \dots, 0, x_j, 0, \dots, 0)$$

Show that P_{ij} is a linear transformation.

<u>Remarks:</u>

(1) If $T : V \to W$ is linear, then

(a) $T(0) = 0$ (b) $T(-v) = -T(v)$

<u>Proof:</u>

(a) $T(v) = T(v+0) = T(v) + T(0)$ (Since T is Linear)

$\Rightarrow (-T(v)) + T(v) = (-T(v)) + T(v) + T(0)$

$\Rightarrow 0 = 0 + T(0) \Rightarrow \boxed{T(0) = 0}$

(b) $T((-v) + v) = T(0) = 0$ (By (a))

$\Rightarrow T(-v) + T(v) = 0$ (Since T is Linear)

$\Rightarrow T(-v) + T(v) + (-T(v)) = 0 + (-T(v))$

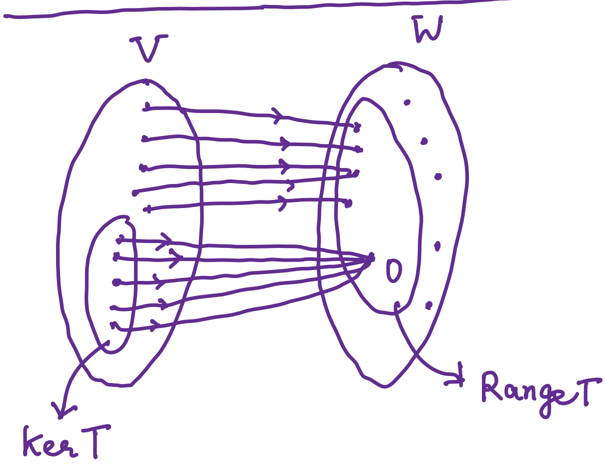
$\Rightarrow T(-v) + 0 = -T(v) \Rightarrow \boxed{T(-v) = -T(v)}$

(2) If T is linear, T preserves linear combinations:

i.e. $T(c_1 v_1 + c_2 v_2 + \dots + c_k v_k)$
$$= c_1 T(v_1) + c_2 T(v_2) + \dots + c_k T(v_k)$$

<u>Proof:</u> Exercise

Two important Subspaces associated with a Linear Transformation:



(1) Let $T: V \longrightarrow W$ be a linear transformation.

- Then the kernel of T, $\ker T = \{v \in V : T(v) = 0 \in W\}$ is a subspace of V
  $\ker T$ is also called the null space of T, denoted by $\text{Nul}\, T$.
- The range of T,
  $\text{Range}\, T = \{ w \in W : w = T(v) \text{ for some } v \in V\}$
  is a subspace of W.

Proof:

- If $u, v \in \ker T$, then $T(u) = 0$, $T(v) = 0$
  Now, $T(u+v) = T(u) + T(v) = 0 + 0 = 0$
  $\Rightarrow u + v \in \ker T$
  If $u \in \ker T$ and $c \in F$ then $T(u) = 0$
  Now $T(cu) = c\,T(u) = c.0 = 0 \Rightarrow cu \in \ker T$
  Hence $\ker T$ is a subspace of V.

- If $w_1, w_2 \in \text{Range}\,T$, then there exist $u, v \in V$ such that $T(u) = w_1$ and $T(v) = w_2$

Now, $w_1 + w_2 = T(u) + T(v) = T(u+v)$ (since T is linear)

and $u+v \in V$ (since V is a vector space)

So, $w_1 + w_2 \in \text{Range}\,T$

If $w \in \text{Range}\,T$ and $c \in F$, then there exists $v \in V$ such that $T(v) = w$

Now, $cw = cT(v) = T(cv)$ (since T is linear)

and $cv \in V$ (since V is a vector space)

So, $cw \in \text{Range}\,T$

Hence $\text{Range}\,T$ is a subspace of W.

Note: • For any linear transformation T, $\ker T \neq \phi$ since $0 \in \ker T$ (since $T(0) = 0$)

Definition: A linear transformation $T : V \to W$ is called injective (1-1) if $Tu = Tv \Rightarrow u = v \ \forall u, v \in W$

(or equivalently $u \neq v \Rightarrow Tu \neq Tv \ \forall u, v \in V$)

Important Remark:

If $T : V \to W$ is a linear transformation, then T is injective (1-1) if and only if $\ker T = \{0\}$.

Proof:

$\Rightarrow$: Assume that T is 1-1. We have seen that $0 \in \ker(T)$ (since $T(0) = 0$)

Now let $v \in V$, $v \neq 0 \Rightarrow T(v) \neq T(0)$ (since T is 1-1)

$$\Rightarrow T(v) \neq 0$$

$$\Rightarrow v \notin \ker T$$

So, $\ker T = \{0\}$

$\Leftarrow$: Assume that $\ker T = \{0\}$

Now $T(u) = T(v)$

$$\Rightarrow T(u) - T(v) = 0$$

$$\Rightarrow T(u - v) = 0 \quad \text{(since } T \text{ is linear)}$$

$$\Rightarrow u - v \in \ker T$$

$$\Rightarrow u - v = 0 \quad (\text{since } \ker T = \{0\})$$

$$\Rightarrow u = v \Rightarrow T \text{ is 1-1.} \quad \text{(QED)}$$

Proposition:

(a) A linear transformation $T: V \longrightarrow W$ is completely determined by its action on a basis of V.

(b) Conversely, given a basis $B = \{v_1, \ldots, v_n\}$ of V, and a list of n vectors $w_1, \ldots, w_n$ (not necessarily distinct) in the co-domain space W, there is a unique linear transformation T such that $T(v_1) = w_1$, $T(v_2) = w_2, \ldots, T(v_n) = w_n$.

Proof: Exercise

Outline of a proof:

(a) If $\{v_1, \ldots, v_n\}$ is a Basis of V, then any arbitrary vector $v \in V$ can be written as a linear combination of $v_1, \ldots, v_n$.

Thus there exist $c_1, \ldots, c_n \in V$ s.t. $v = c_1 v_1 + \cdots + c_n v_n$

$\Rightarrow T(v) = T(c_1 v_1 + \cdots + c_n v_n) = c_1 T(v_1) + \cdots + c_n T(v_n)$

Thus the image of any arbitrary vector under T is a linear combination of $Tv_1, \ldots, Tv_n$

i.e. T is completely determined by $Tv_1, \ldots, Tv_n$

(b) First note that the transformation T defined by $Tv_i = w_i$ for $i = 1, 2, \ldots, n$ is a linear transformation (check!!)

Now if there are two linear transformations T_1 & T_2 with the same properties i.e. $T_1(v_i) = w_i$, $T_2(v_i) = w_i$ for $i = 1, 2, \ldots, n$

then for any arbitrary vector $v = c_1 v_1 + \cdots + c_n v_n$, $c_1, \ldots, c_n \in F$

$T_1(v) = T_1(c_1 v_1 + \cdots + c_n v_n) = c_1 T_1 v_1 + \cdots + c_n T_1 v_n = c_1 T_2 v_1 + \cdots + c_n T_2 v_n = T_2(c_1 v_1 + \cdots + c_n v_n)$

Thus $T_1 = T_2$ and so such T is unique.

Rank of a Linear Transformation:

- For the time being, we will assume V to be finite-dimensional.

Definition: Let $T: V \longrightarrow W$ be a linear transformation. Then Rank of T is defined to be the $\boxed{\text{dimension of the Range of } T}$

Remark:
Range$[T]$ is finite dimensional and

$$\dim(\text{Range}(T)) \leq \dim V$$

Thus definition of Rank (T) is valid.

Proof: Use the previous proposition: given that V is finite dimensional.

Let $B = \{v_1, \dots, v_n\}$ is a Basis for V.

Let $Tv_1 = w_1, \dots, Tv_n = w_n$

Then $w_1, w_2, \dots, w_n$ Span Range T:

Let $w \in$ Range T, Then there exists an $v \in V$ s.t. $Tv = w$, Since B is a basis of V, there exist scalars $c_1, \dots, c_n$ s.t. $v = c_1 v_1 + \dots + c_n v_n$

Now $w = Tv = T(c_1 v_1 + \dots + c_n v_n) = c_1 T(v_1) + \dots + c_n T(v_n)$
$= c_1 w_1 + \dots + c_n w_n$

Therefore $\boxed{\dim(\text{Range}(T)) \leq \dim(V) = n}$

Recall from Last time:

- For a linear transformation $T: V \to W$, we defined $\ker T = \text{Null}\, T = \{ v \in V : T(v) = 0 \}$ and showed that it is a subspace of V.
- If $\ker T$ is finite-dimensional, then $\dim(\ker T)$ is called the $\boxed{\text{nullity of } T}$

Theorem (Rank Theorem for Linear Transformations):

Suppose that $T: V \to W$ is a linear transformation and V is finite dimensional. Then $\boxed{\text{Rank}(T) + \text{nullity}(T) = \dim V}$.

Note: We have already seen that if $T: V \to W$ is a linear transformation and V is finite-dimensional, then range T is also finite dimensional and

$\dim(\text{range}\, T) \leq \dim V$.

i.e. $\text{Rank}(T) \leq \dim V$.

Proof of the Rank Theorem:

- Assume that $\dim V = n$ and $\text{nullity}(T) = k$. Let $v_1, v_2, \dots, v_k$ be a basis of $\ker T$. Expand this to a basis B of V by inserting the additional vectors $v_{k+1}, \dots, v_n$.
- We will show that $T(v_{k+1}), \dots, T(v_n)$ form a basis for Range(T).

Firstly all the vectors $T(v_1), \dots, T(v_n)$ span Range(T)

Any element of Range(T) is of the form $T(v)$ for some $v \in V$.

Since $v_1, \dots\dots, v_n$ form a basis of V, there exist scalars $c_1, \dots, c_n \in F$ such that $v = c_1 v_1 + \dots + c_n v_n$

$$\Rightarrow \quad T(v) = T(c_1 v_1 + \dots + c_n v_n)$$
$$= c_1 T(v_1) + \dots + c_n T(v_n)$$

Since, $T(v_1) = T(v_2) = \dots = T(v_k) = 0$, actually $T(v_{k+1}), \dots, T(v_k)$ span Range T.

Now suppose that

$$c_{k+1}T(v_{k+1}) + c_{k+2}T(v_{k+2}) + \cdots + c_nT(v_n) = 0$$

$$\Rightarrow T(c_{k+1}v_{k+1} + c_{k+2}v_{k+2} + \cdots + c_nv_n) = 0$$

$$\Rightarrow c_{k+1}v_{k+1} + c_{k+2}v_{k+2} + \cdots + c_nv_n \in \operatorname{Ker} T$$

So, there exist scalars $b_1, \ldots, b_k$ such that $c_{k+1}v_{k+1} + \cdots + c_nv_n = b_1v_1 + \cdots + b_kv_k$

$$\Rightarrow b_1v_1 + \cdots + b_kv_k - c_{k+1}v_{k+1} - \cdots - c_nv_n = 0$$

Since $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ form a basis of V, they are linearly independent and hence $c_{k+1} = \cdots = c_n = 0$

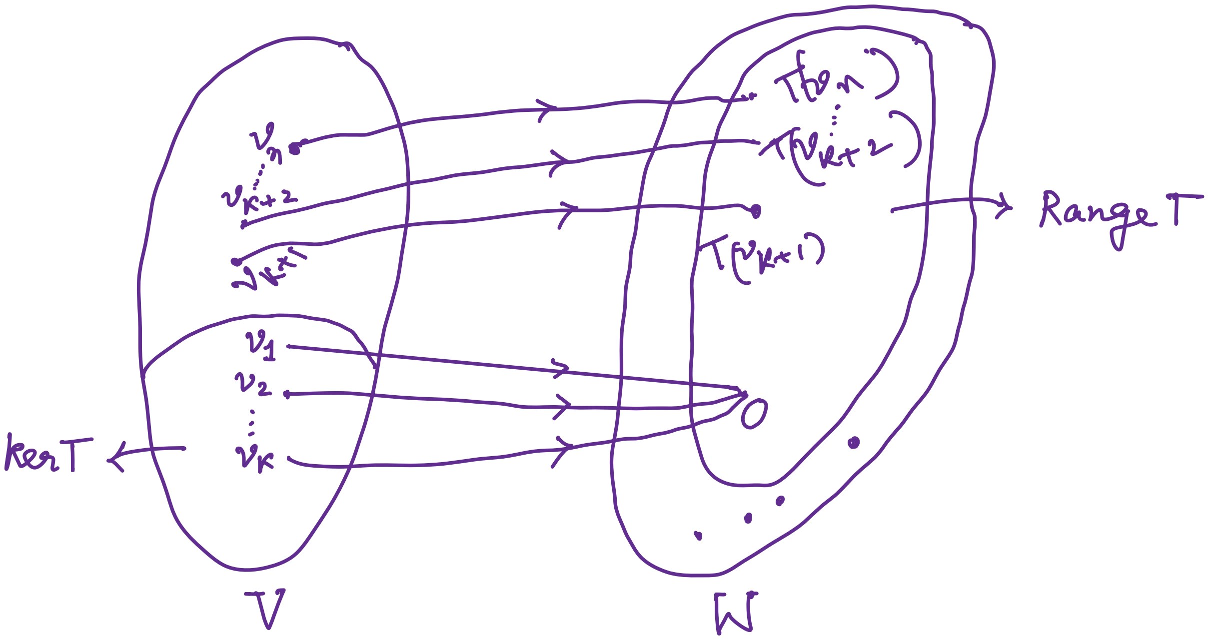
Thus $T(v_{k+1}), \ldots, T(v_n)$ form a basis of Range T.

Now $\operatorname{Rank} T = \dim(\operatorname{Range} T) = n-k$

Hence $\operatorname{Rank} T + \operatorname{nullity} T = n-k+k$

$$= n = \dim V$$

(QED)


v_n
v_{k+2}
v_{k+1}
v_1
v_2
v_k
ker T
$T(v_n)$
$T(v_{k+2})$
$T(v_{k+1})$
0
Range T
V
W

MTH 100: Lecture 26

Definition: A linear transformation $T: V \to W$ is called an isomorphism if it is injective and surjective (i.e. if $\text{Range}\, T = W$)

Proposition: Let V and W be finite dimensional spaces.

(a) An isomorphism $T: V \to W$ takes any arbitrary basis of V to a basis of W.

(b) Conversely, if a linear transformation $T: V \to W$ takes some basis of V to a basis of W, then it is an isomorphism.

Sketch of a Proof:

(a): Given: $T: V \to W$ is an isomorphism

Now, T is onto $\Rightarrow \text{Range}\, T = W \Rightarrow \text{Rank}\, T = \dim W$

T is 1-1 $\Rightarrow \ker T = \{0\} \Rightarrow \text{nullity}\, T = 0$

By the Rank Theorem,

$$\text{Rank } T + \text{nullity } T = \dim V = n \text{ (say)}$$
$$\Rightarrow \dim W + 0 = n$$
$$= \dim W = \dim V = n$$

Let $\{v_1, \ldots, v_n\}$ be a Basis of V.

Then $\{Tv_1, \ldots, Tv_n\}$ is a spanning set of $\text{Range } T = W$. Since $\dim W = n$, $\{Tv_1, \ldots, Tv_n\}$ forms a basis of W.

(b) $\Leftarrow$:

Assume that $\dim V = n$ and T takes some basis $\{v_1, \ldots, v_n\}$ of V to a basis $\{Tv_1, \ldots, Tv_n\}$ of W

Therefore $\dim W = n$ and $\text{Rank } T = n$

Hence $\text{Range } T = W$ and T is onto.

Now using nullity Theorem:

$$\text{Rank } T + \text{nullity } T = \dim V = n$$
$$\Rightarrow n + \text{nullity } T = n$$
$$\Rightarrow \text{nullity } T = 0$$
$$\Rightarrow \text{Ker } T = \{0\}$$
$$\Rightarrow T \text{ is } 1\text{-}1.$$

Therefore T is an isomorphism.

Proposition: Two finite dimensional vector spaces V and W (over the same field F) are isomorphic if and only if $\dim V = \dim W$.

Proof: $\Rightarrow$: Assume $T: V \longrightarrow W$ is an isomorphism.
Want to show $\dim V = \dim W$.

Suppose $\dim V = n$ and let $\{v_1, \dots, v_n\}$ be a basis of V.
Then by the previous proposition $\{Tv_1, \dots, Tv_n\}$ is a basis of W. Hence $\dim W = n$
So, $\boxed{\dim V = \dim W}$

$\Leftarrow$: Assume that $\dim V = \dim W$.
Want to show that V and W are isomorphic.

Let $\{v_1, \dots, v_n\}$ be a basis of V
and $\{w_1, \dots, w_n\}$ be a basis of W.
Consider the unique linear transformation $T: V \longrightarrow W$ such that $Tv_i = w_i$ for $i = 1, 2, \dots, n$.

Since T takes a basis of V to a basis of W, by the previous proposition, T is an isomorphism and hence V and W are isomorphic.

Remark: Every vector space of dimension n over a field F is isomorphic to F^n. In particular, every vector space of dimension n over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$.

An Important Linear Transformation:

Left multiplication by a Matrix:

Let A be a $m\times n$ matrix.

Define $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

by $T_A(x) = Ax$ [Note that $\underbrace{A_{m\times n}\, x_{n\times 1}}_{m\times 1}$ is a $m\times 1$ matrix]

- T_A is a linear transformation.

- For $x, y \in \mathbb{R}^n$,

$$T_A(x+y) = A(x+y) = Ax + Ay = T_A(x) + T_A(y)$$

- For $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

$$T_A(cx) = A(cx) = cAx = cT_A(x)$$

Hence T_A is a linear transformation.

Consider the Reverse Problem:

Suppose V and W are finite dimensional vector spaces over the field F.

Suppose $T: V \to W$ is a linear transformation
We will associate a matrix with this linear transformation.

Coordinate Systems:

Suppose V is a finite dimensional vector space.
An ordered basis for a finite dimensional vector space V is a finite sequence of vectors which is linearly independent and spans V.

In other words,

an ordered basis is a basis with the vectors taken in a specified fixed order.

Thus, given an order basis of V
$$B = \{u_1, \dots, u_n\},$$
we can express any vector $u \in V$ uniquely in the form $u = x_1 u_1 + x_2 u_2 + \dots + x_n u_n$
The scalars x_i are called the coordinates of u relative to the (ordered) basis B

Remark: given a fixed ordered basis B for a finite dimensional vector space V, we can find an n-tuple in F^n

(usually F is $\mathbb{R}$ or $\mathbb{C}$) corresponding to any vector u in V as follows : $u \longrightarrow (x_1, x_2, \dots, x_n)$ where x_i are the coordinates of u relative to B

Rather than the n-tuple $(x_1, x_2, \dots, x_n)$ we express it as a column vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

- This vector is called the coordinate vector of u (relative to B) and is written $[u]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Ex: Let $V = R_n(t)$ be the vector space of polynomials of degree $\leq n$

Then $B = \{1, t, t^2, \dots, t^n\}$ is an ordered basis of $R_n(t)$

Let $v = 2t^3 \in R_n(t)$

Then v can be written as :

$$v = 0.1 + 0.t + 0.t^2 + 2.t^3 + 0.t^4 + \dots + 0.t^n$$

Hence the coordinate vector of v relative to B is $[v]_B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

The Determinant

- **Introduction:** The following material about determinants has been collated from Chapter 3 of the textbook by Lay. It contains all the information about determinants that you are expected to be familiar with. From now onwards, you would be free to use any of the definitions and results presented here. Probably, you are already familiar with most of these results. They will not be presented in class. However, you are welcome to review Chapter 3 of Lay and try out some practice exercises if desired.

The Determinant

- **Remark:** Propositions about determinants will be numbered independently as Prop D1, Prop D2, etc.
- **Definition of the Determinant:** If $A \in F^{2\times2}$ where $A = [a_{ij}]$, then det A is defined to be the scalar $a_{11}\, a_{22} - a_{12}\, a_{21}$. Thus det is a function from $F^{2\times2}$ to F.
- We extend this definition recursively to $F^{n\times n}$. as follows:
- **Notation:** If $A \in F^{n\times n}$, let $A_{i,j}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by omission of the i-th row and j-th column.
- **Column expansion formula:** A formula for the determinant is given by:

  $\det A = \Sigma\, (-1)^{i+j}\, a_{ij} \det A_{i,j}$, where the summation is taken for i = 1 to n.
- **Row expansion formula:** Another formula for the determinant is given by:

  $\det A = \Sigma\, (-1)^{i+j}\, a_{ij} \det A_{i,j}$, where the summation is taken for j = 1 to n.

The Determinant - 1

- **Proposition D1:** The following hold for the determinant of a square matrix A:
  i. If the matrix A’ is obtained from A by interchanging two rows, then det A’ = – det A
  ii. If the matrix A’ is obtained from A by multiplying some row by $\lambda \in F$, then det A’ = λ det A
  iii. If the matrix A’ is obtained from A by adding a multiple of one row to another row, then det A’ = det A
- **Remark 1**: The above indicates what happens to the determinant when an elementary row operation – interchange, scaling, or replacement – is applied.
- **Remark 2:** The above holds if row is replaced by column.
- **Remark 3**: It follows directly from the above that if the rows (or columns) of A are linearly dependent, then det A = 0.

Procedure for Computing the Determinant

- **Proposition D2:** If an $n \times n$ matrix A is upper triangular, then $\det A = a_{11}a_{22}\ldots\ldots a_{nn}$
- **Corollary D2.1:** In order to determine the determinant of an $n \times n$ matrix, use elementary row operations of interchange and replacement type only to reduce A to an upper triangular matrix A'. If r is the number of row interchanges carried out, then $\det A = (-1)^r \det A'$.
- **Remark 1**: This follows directly from Proposition D1 and the definition (using the column expansion).
- **Remark 2**: The above method is far less computationally intensive than using either row or column expansion. **NB**: Most advanced textbooks use a different definition (formula) for the determinant; however, it is equally inefficient computationally.

Further Properties of the Determinant - 1

- **Proposition D3:** An $n \times n$ matrix A is invertible if and only if $\det A \neq 0$.
- **Remark:** The above gives another useful property equivalent to invertibility for square matrices. Consequently, we need to extend our theorem on invertibility of matrices (see next slide).

Very Important Theorem – Ver 1.1

- **Theorem 1**: The following are equivalent for an $m \times m$ square matrix A:

a. A is invertible

b. A is row equivalent to the identity matrix

c. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution

d. The system of equations $A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b}$ in $\mathbb{R}^m$.

e. Det $A \neq 0$

Further Properties of the Determinant - 2

- **Proposition D4:** for all A,B $\in F^{n\times n}$, det (AB) = (det A) (det B)
- **Corollary D4.1:** If A is invertible, then det A^{-1} = (det A)$^{-1}$
- **Remark: W**hile det (AB) = (detA)(detB), in general det(A + B) $\neq$ det A + det B. *As we shall later, the determinant is not a linear function or linear transformation.*
- **Proposition D5:** For all A $\in F^{n\times n}$, det A^T = det A.

Cramer's Rule

- **Remark:** If you have not studied this topic before, it is nicely presented in the book by Lay: Section 3.3
- **Definition:** For any n×n matrix A and any vector **b** in R^n, define $A_i(\mathbf{b})$ to be the matrix obtained by replacing the i-th column of A by **b**.
- **Proposition D6 (Cramer's Rule):** Let A be any invertible n×n matrix. For any vector **b** in R^n, the unique solution **x** of A**x** = **b** has entries given by:

  $x_i = (\det A_i(\mathbf{b}))/(\det A)$ for i = 1,2,..,n

- Cramer's Rule is (usually) not a practical method for solving systems of linear equations since it requires computation of (n + 1) determinants.

Application of Cramer’s Rule

- **Terminology and Notation:** For any n×n matrix A, we define the cofactor $C_{ij} = (-1)^{i+j} \det(A_{ij})$
- Definition: the classical adjoint of A (written adj A) is the matrix whose entries are the cofactors of A transposed. In other words, adj A is the matrix:

$$\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & & C_{nn} \end{bmatrix}$$

- **Proposition D7: Inverse Formula:** Let A be any invertible n×n matrix. Then:

$A^{-1} = (1/\det A)(\text{ adj } A)$

Application of Determinants to Areas and Volumes - 1

- **Proposition D8:** (a) If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$.

  (b) If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$.

Application of Determinants to Areas and Volumes - 2

- **Remark:** *We will shortly introduce the topic of linear transformations. You should revisit the following propositions after that.*
- **Proposition D9:** (a) Let $T:R^2 \to R^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in R^2, then {area of T(S)} = |det A| × {area of S}.

  (b) Let $T:R^3 \to R^3$ be the linear transformation determined by a 3×3 matrix A. If S is a parallelepiped in R^3, then {volume of T(S)} = |det A| × {volume of S}.

Application of Determinants to Areas and Volumes - 3

- **Proposition D10:** The conclusions of Proposition D9 hold whenever S is a region in R^2 with finite area or a region in R^3 with finite volume. In other words:

  {area or volume of T(S)} = |det A| × {area or volume of S}.

MTH 100 : Lecture 27

Coordinate Systems:

Let V be a finite dimensional vector space and $B = \{v_1, \ldots, v_n\}$ be an ordered basis of V.

Then any vector $u \in V$ can be uniquely written as $u = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n$

The vector $[u]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is called the coordinate vector of u (relative to B) and is denoted by $[u]_B$

- Now the mapping or correspondence between V and F^n given by : $u \longmapsto [u]_B$ is called the coordinate mapping determined by B

- It is an one-to-one correspondence.
  i.e. each vector has a unique corresponding n-tuple and each n-tuple has a unique corresponding vector.

- The sum of two vectors corresponds to the sum of the two n-tuples.

- The scalar multiple of a vector corresponds to a scalar multiple of the n-tuple.

Therefore the coordinate mapping is actually an isomorphism from an n-dimensional vector space V over the field F to F^n.

- Note that we get a <u>different isomorphism for each choice of an ordered basis for V.</u> (Recall proposition of last class).

<u>Ex:</u> Let $V = \mathbb{R}^3$

Now, $E = \{e_1, e_2, e_3\}$ is an ordered Basis of $\mathbb{R}^3$ where $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Let $v = \begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix} \in V \in \mathbb{R}^3$

$$[v]_E = \begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix}_E$$

(because $\begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix} = 14e_1 + 11e_2 + 7e_3$, $\underbrace{14,\ 11,\ 7}_{\text{coordinates}}$)

Now $B = \{v_1, v_2, v_3\}$ be another ordered Basis of V

where $v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Now

$$v = \begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

So, $$[v]_B = \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}_B$$

Similarly

since $$w = \begin{bmatrix} 12 \\ 15 \\ 9 \end{bmatrix} = -3\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 6\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 9\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

we have $$[w]_B = \begin{bmatrix} -3 \\ 6 \\ 9 \end{bmatrix}_B$$

---x---x---x---x---x---x---x

In general if $z \in V = \mathbb{R}^3$, then to find $[z]_B$, we need to find coefficients x_1, x_2, x_3 such that $x_1 v_1 + x_2 v_2 + x_3 v_3 = z$

$$\Rightarrow \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = z$$

$$\Rightarrow \quad Ax = z \quad \text{where } A = [v_1 \; v_2 \; v_3]$$
and $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$\Rightarrow \quad x = A^{-1} z$$

(Note that A is invertible as its columns are linearly independent)

Here $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

Then $A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ (Please calculate and check !!)

Now for
$z = v = \begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix}$

$$[z]_B = A^{-1} z = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 14 \\ 11 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}_B \quad \text{(as expected)}$$

Ex:
If $z_1 = \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix}$, check that $[z_1]_B = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix}$

$$= \begin{bmatrix} -2 \\ 5 \\ 9 \end{bmatrix}_B$$

Matrix of a Linear Transformation

Suppose V and W are finite dimensional vector spaces over the field F and $T: V \to W$ is a linear transformation.

Suppose $\dim V = n$ and $\dim W = m$

Let $B = \{v_1, \dots, v_n\}$ be an ordered basis of V and $C = \{w_1, \dots, w_m\}$ be an ordered basis of W.

Since $Tv_1, Tv_2, \dots, Tv_n \in W$, we can express them uniquely as linear combinations of $w_1, \dots, w_m$.

Thus we can write
$$Tv_1 = A_{11} w_1 + A_{21} w_2 + \cdots + A_{m1} w_m$$
$$Tv_2 = A_{12} w_1 + A_{22} w_2 + \cdots + A_{m2} w_m$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$Tv_n = A_{1n} w_1 + A_{2n} w_2 + \cdots + A_{mn} w_m$$

We now form the $m \times n$ matrix A with these coefficients as columns.

i.e.
$$A_{m \times n} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

- The matrix A is called the matrix of T with respect to the bases B and C and is denoted by $[T]_{B \to C}$
- For any vector $v \in V$, we can find the coordinates of Tv in W by left multiplying the coordinate vector of v by the matrix $A = [T]_{B \to C}$

- In terms of coordinate vectors, we can write : $\boxed{[T(v)]_C = [T]_{B\to C}[v]_B}$

- In the special case of a linear operator, i.e. a linear transformation from V into itself, the bases B and C are usually taken as the same, and the matrix A is called the B-matrix for T, written $[T]_B$.

  Then the above equation becomes:

$$\boxed{[T(v)]_B = [T]_B [v]_B}$$

<u>Ex:</u> Let $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be given by

$$T(x,y,z) = (x+y+z,\ x+2y+3z)$$

Let

$$B = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

and $C = \left\{ e_1' = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ e_2' = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

Then B is an ordered basis of $\mathbb{R}^3$ and C is an ordered basis of $\mathbb{R}^2$

$$Te_1 = T\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1\,e_1' + 1\,e_2'$$

$$Te_2 = T\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\,e_1' + 2\,e_2'$$

$$Te_3 = T\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 1\,e_1' + 3\,e_2'$$

<u>So, the matrix of T:</u>

$$[T]_{B \to C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}_{2\times 3}$$

If $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$,

then $$[T(v)]_C = [T]_{B \to C} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$= \begin{bmatrix} x+y+z \\ x+2y+3z \end{bmatrix} = [Tv]_C$$

<u>Ex:</u> Let $D : \mathbb{R}_3[t] \longrightarrow \mathbb{R}_2[t]$ be defined by $D[p(t)] = p'(t)$ for any $p(t) \in \mathbb{R}_3(t)$

Note that $\dim(\mathbb{R}_3[t]) = 4$ and $\dim(\mathbb{R}_2[t]) = 3$ and

$B = \{1, t, t^2, t^3\}$ is an ordered basis of $\mathbb{R}_3[t]$

and $C = \{1, t, t^2\}$ is an ordered basis of $\mathbb{R}_2[t]$

Now

$$D(1) = 0 = 0.1 + 0.t + 0.t^2$$
$$D(t) = 1 = 1.1 + 0.t + 0.t^2$$
$$D(t^2) = 2t = 0.1 + 2.t + 0.t^2$$
$$D(t^3) = 3t^2 = 0.1 + 0.t + 3.t^2$$

Therefore $[D]_{B\to C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

Now let us take a particular polynomial

$$p_1(t) = 4 + 5t + 2t^2 + 3t^3$$

Then $[p_1(t)]_B = \begin{bmatrix} 4 \\ 5 \\ 2 \\ 3 \end{bmatrix}$

Now $[T]_{B\to C} \; [p_1(t)]_B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 2 \\ 3 \end{bmatrix}$

$$= \begin{bmatrix} 5 \\ 4 \\ 9 \end{bmatrix}_C$$

Note that $D[p_1(t)] = 5 + 4t + 9t^2$

whose coordinate vector with respect to C is

$$\begin{bmatrix} 5 \\ 4 \\ 9 \end{bmatrix}_C$$

MTH 100 : Lecture 28

Change of Basis

- We would like to know what happens to the matrix of a linear transformation if the basis gets change.
- We will only consider the case when T is a linear operator from a finite dimensional vector space V to V

Preliminary Result:

Proposition: Let $B = \{u_1, \ldots, u_n\}$ and $C = \{v_1, \ldots, v_n\}$ be two ordered bases of a vector space V. Then there is an invertible $n \times n$ matrix P such that

$$[x]_C = P\,[x]_B \quad \text{for any } x \in V.$$

Proof: Will be given as a note. (rather technical)

Note: The columns of P are the C-coordinate vectors of the basis B.

The matrix P is called the change of coordinate matrix from B to C and is denoted by $P_{B \to C}$

Remark: To change coordinates between two bases, we need the coordinate vectors of the old basis B relative to the new basis C. These become the columns of the change of matrix P.

- In practice $P = Q^{-1}$ where Q has its columns the coordinate vectors of the new basis C relative to the old basis B.

In most of the applications, the old basis is the standard basis for $\mathbb{R}^n$ and so Q can be found directly.

- Recall the first example of last lecture.

Ex (First part):

$$\text{Let } V = \mathbb{R}^2$$

Let old (ordered) basis $\alpha = \{e_1, e_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

(standard basis in $\mathbb{R}^2$)

and Let

new (ordered) basis $\beta = \{u_1, u_2\} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right\}$

(Should be clear that this is a basis)

Step 1: Construct the matrix Q

$$= \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

New basis in terms of old basis

(Since $u_1 = 2e_1 + 1e_2$
and $u_2 = 5e_1 + 3e_2$
we have $[u_1]_\alpha = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_\alpha$
and $[u_2]_\alpha = \begin{bmatrix} 5 \\ 3 \end{bmatrix}_\alpha$)

Step 2 Change of Basis Matrix $= P = \boxed{P_{\alpha\to\beta} = Q^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}}$

check: Let us determine $[v]_\beta$ for a specific vector v, say $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_\alpha$

Then $[v]_\beta = P[v]_\alpha = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -7 \\ 3 \end{bmatrix}_\beta$

Note that $\begin{bmatrix} -7 \\ 3 \end{bmatrix}_\beta = -7u_1 + 3u_2 = -7\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_\alpha = v$

Verification that columns of P are the coordinate vectors of the old basis in terms of the new basis.

First column of $P = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_\beta = 3u_1 + (-1)u_2 = 3\begin{bmatrix} 2 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = e_1$

i.e. $[e_1]_\beta = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_\beta$

Similarly,

Second column of $P = \begin{bmatrix} 5 \\ -2 \end{bmatrix}_\beta = 5u_1 + (-2)u_2 = 5\begin{bmatrix} 2 \\ 1 \end{bmatrix} + (-2)\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = e_2$

i.e. $[e_2]_\beta = \begin{bmatrix} 5 \\ -2 \end{bmatrix}_\beta$

Similarity of Matrices

- A $n\times n$ matrix B is called similar to an $n\times n$ matrix A if there exists an invertible matrix P such that $\boxed{B = PAP^{-1}.}$

Proposition: Similarity of matrices is an equivalence relation on $F^{n\times n}$

($F^{n\times n}$ is the set of $n\times n$ matrices with entries taken from a field F)

Remarks (1): If A is similar to B, then B is similar to A.

(So, we will say A and B are similar matrices)

(2) If A and B are similar matrices then $\boxed{\det(A) = \det(B)}$

Effect of change of Basis

Proposition: Suppose A and B are the matrices of the linear operator T relative to the ordered basis α and β respectively.

Then A and B are similar matrices.

Infact, $\boxed{B = PAP^{-1}}$, where $P = P_{\alpha\to\beta}$ is the change of Basis matrix.

$$\boxed{\text{i.e. } \underset{\substack{\downarrow \\ B}}{[T]_\beta} = P \underset{\substack{\downarrow \\ A}}{[T]_\alpha} P^{-1}}$$

Proof: If P is the change of basis matrix from α to β, then P^{-1} is the change of basis matrix from β to α.

Let $[T]_\alpha = A$ and $[T]_\beta = B$

Then for any $v \in V$, $(PAP^{-1})[v]_\beta = (PA)(P^{-1}[v]_\beta)$

$$= (PA)[v]_\alpha = P(A[v]_\alpha)$$

$$= P([T]_\alpha [v]_\alpha) = P[Tv]_\alpha$$

$$= [Tv]_\beta = [T]_\beta [v]_\beta = B[v]_\beta$$

Since the above holds for all vectors $v \in V$, it follows that

$$\boxed{PAP^{-1} = B = [T]_\beta}$$

Ex: (2nd part): Let $V = \mathbb{R}^2$

Let $\alpha = \{e_1, e_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be the standard basis of $\mathbb{R}^2$

Let $\beta = \{x_1, x_2\} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right\}$ be another basis of $\mathbb{R}^2$

We have seen that $Q = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ and the change of basis matrix

$$P = P_{\alpha \to \beta} = Q^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$$

Define a linear operator $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + 2y \\ 3x + 4y \end{bmatrix};$$

Since $Te_1 = T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 1e_1 + 3e_2$

and $Te_2 = T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 2e_1 + 4e_2$,

We have $[T]_\alpha = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = A$ (say)

By the proposition,

$$B = [T]_\beta = PAP^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -38 & -102 \\ 16 & 43 \end{bmatrix}$$

Verification with a specific vector:

$$\text{Let } v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_\alpha = \begin{bmatrix} -7 \\ 3 \end{bmatrix}_\beta$$

Now, $$[Tv]_\alpha = [T]_\alpha [v]_\alpha = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}_\alpha$$

and $$[Tv]_\beta = [T]_\beta [v]_\beta = \begin{bmatrix} -38 & -102 \\ 16 & 43 \end{bmatrix} \begin{bmatrix} -7 \\ 3 \end{bmatrix} = \begin{bmatrix} -40 \\ 17 \end{bmatrix}_\beta$$

Note that

$$\begin{bmatrix} -40 \\ 17 \end{bmatrix}_\beta = -40 u_1 + 17 u_2 = -40 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 17 \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}_\alpha$$

So, we get the same vector but expressed in two different coordinate system.

MTH 100 : Lecture 29

Algebra of Linear Transformations

Let V and W be vector spaces over a field F.

Notation: $W^V \equiv$ The set of all functions from V to W

$L(V,W) \equiv$ The set of all linear transformations from V to W

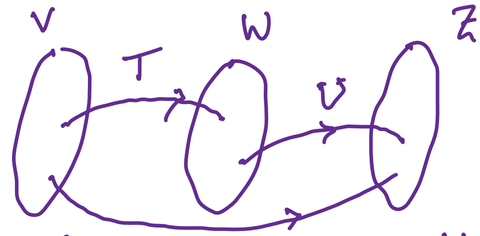
Proposition:
(a) The set W^V of all functions from V to W is a vector space over F.

(b) The set $L(V,W)$ of all linear transformations from V to W is a subspace of W^V

Proof: Exercise (For (b) use test for a subspace)

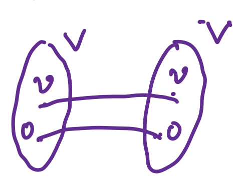
Proposition: Let V, W and Z be vector spaces over a field F. Let T be a linear transformation from V to W and U be a linear transformation from W to Z. Then the composed function UT from V to Z defined by $(UT)(v) = U(T(v))$ for all $v \in V$ is a linear transformation from V into Z.

Proof: Exercise



Linear operators:

- A linear operator on a vector space V is a linear transformation from V to V
- $L(V,V) \equiv$ The space of all linear operators on V
  (i.e. the space of all linear transformations from V into V)
- $L(V,V)$ is of primary importance because we can define a 'multiplication', i.e. composition of linear operators.
  (Note that we can't do this on $L(V,W)$ when W is different from V.)
- By the previous proposition, the composition of two linear operators on V is a linear operator on V.
- Note Define the identity operator $I: V \to V$ by $I(v) = v \;\; \forall\, v \in V$



- Composition of Linear operators satisfies the following properties:

(a) $IU = UI = U$ for all linear operator U where I is the identity operator on V

(b) $(T_1 T_2)T_3 = T_1(T_2 T_3)$ (Associative Law)

(c) $U(T_1+T_2) = UT_1 + UT_2$

(d) $(T_1+T_2)U = T_1 U + T_2 U$

(e) $c(UT_1) = (cU)T_1 = U(cT_1)$.

*(f) However, this multiplication is not commutative.

- Verification of the above properties is left as an exercise.

- A vector space with a multiplication which satisfies properties (a) through (e) above is called an $\boxed{\text{ALGEBRA}}$.

Thus $L(V\ V)$ is an Algebra.

A Fundamental Isomorphism

Proposition: Let V be an n-dimensional vector space over the field F and let W be an m-dimensional vector space over F.
Then there is an isomorphism between
$L(V, W)$ and $F^{m \times n}$

Outline of a Proof:
We take a fixed ordered basis $\alpha = \{v_1, v_2, \ldots, v_n\}$ for V and a fixed ordered basis $\beta = \{w_1, w_2, \ldots, w_m\}$ for W.
For any linear transformation $T \in L(V, W)$, we can find the matrix of T with respect to the bases α and β. Let us denote it by $[T]_{\alpha \to \beta}$.
Clearly $[T]_{\alpha \to \beta} \in F^{m \times n}$

Define the mapping $\phi : L(V, W) \longrightarrow F^{m \times n}$
by $\phi(T) = [T]_{\alpha \to \beta}$

- Now show that ϕ is linear, 1-1 and onto. (need to show!!)
  Hence ϕ is an isomorphism.

Note: The isomorphism ϕ above is defined in terms of the bases α and β and is therefore dependent on the choice of α and β.

Proposition: If $\dim V = n$, $\dim W = m$
then $\dim L(V, W) = mn$

Proof: can be proved in two different ways.

First way: Using the previous proposition, we can say that $L(V,W)$ is isomorphic to $F^{m\times n}$.

Since $\dim(F^{m\times n}) = mn$, we conclude that $\boxed{\dim L(V,W) = mn}$

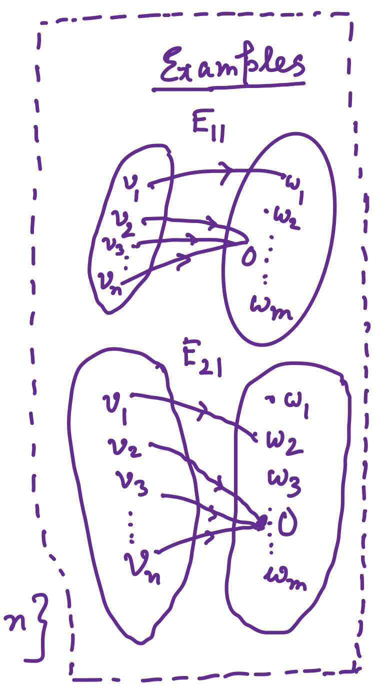
Second way: Let us take a fixed ordered basis $\alpha = \{v_1, \dots, v_n\}$ for V and a fixed ordered basis $\beta = \{w_1, \dots, w_m\}$ for W.

Define a linear transformation

$$E_{ij} : V \longrightarrow W$$

by $E_{ij}(v_j) = w_i$ and

$E_{ij}(v_k) = 0$ for $k \neq j$



- It can be shown that the family $S = \{E_{ij} : i = 1, \dots m;\ j = 1, \dots, n\}$ forms a basis for $L(V,W)$.

Since the number of elements in S is mn, we conclude that $\dim L(V,W) = \boxed{mn}$

Note: The matrix of the linear transformation E_{ij} with regard to the basis α and β is the matrix $[E_{ij}]$ defined earlier.

(Recall the $m \times n$ matrix $[E_{ij}]_{m\times n}$ is defined as the matrix whose (i,j)th entry is 1 and rest of the entries are zero.)

i.e. $[E_{ij}]_{\alpha\to\beta} = [E_{ij}]$

Recall that the matrices $[E_{ij}]$ form a basis for $F^{m\times n}$.

Proposition ⊛:

Suppose T and U are linear operators on a finite dimensional vector space V and β is a fixed ordered basis for V.

Then $[UT]_\beta = [U]_\beta [T]_\beta$

(So, the matrix of the product of two linear operators is the product of their matrices. Thus we have a seamless transition from operators to matrices and vice-versa.)

Note: If $\left.\begin{array}{l} T : V \to V \\ U : V \to V \end{array}\right\}$ then $UT : V \to V$ and all the three operators have corresponding matrices with respect to a fixed ordered basis of V.

Alternative Statement of previous proposition ⊛ :

The mapping $\phi : L(V,V) \longrightarrow F^{n\times n}$ given by $\phi(T) = [T]_\beta$ is a vector space isomorphism which also preserves products,

i.e. $\phi(UT) = \phi(U)\phi(T)$

Generalization of Proposition ⊛ :

The Proposition ⊛ can be extended to the composition of linear transformations $T : V \to W$ and $U : W \longrightarrow Z$ in the following way :

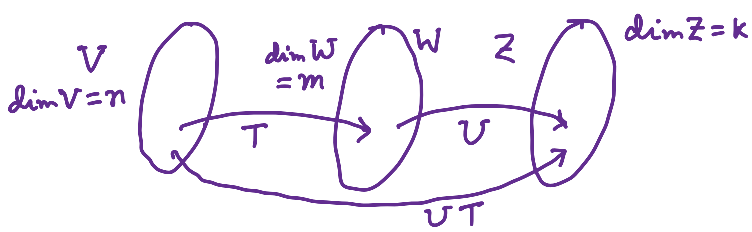
Suppose that $\dim V = n$, $\dim W = m$, $\dim Z = k$

Then $UT : V \longrightarrow Z$ will be a linear transformation from a space of dimension n to a space of dimension k

i.e. its matrix would be an $k \times n$ matrix.

Let α, β, γ be bases of V, W, Z respectively,

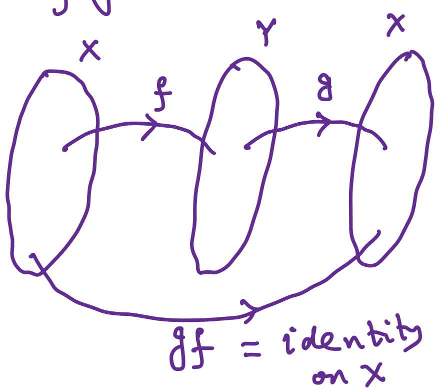
then $$\underbrace{[UT]_{\alpha\to\gamma}}_{k\times n} = \underbrace{[U]_{\beta\to\gamma}}_{k\times m} \underbrace{[T]_{\alpha\to\beta}}_{m\times n}$$

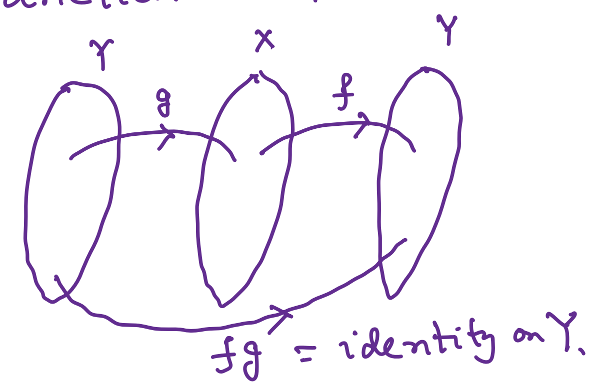


MTH 100: Lecture 30

Invertible functions

Definition: A function $f : X \to Y$ is called invertible if there exists a function $g : Y \to X$ such that gf is the identity function on X and fg is the identity function on Y

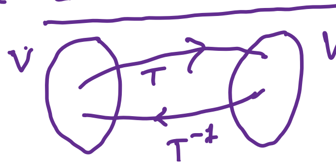




- If f is invertible, then the function g is unique and is called the inverse of f, denoted by f^{-1}.
- A function f is invertible if and only if f is 1-1 and onto (i.e. bijective)

Invertibility of Linear Transformations

Proposition: If $T : V \to W$ is an invertible linear transformation, its inverse function $T^{-1} : W \to V$ is also a linear transformation.



Proof: Exercise

Note: We had earlier referred to invertible linear transformations as isomorphisms. We may use either term.

Corollary: Isomorphism is an equivalence relation on the set of all vector spaces over a given field F.

Outline of Proof: Reflexive property is obvious. Prove Symmetric and transitive property using earlier propositions.

Eigen Vectors and Eigen Values:

- A scalar λ is called an eigen value of a $n \times n$ matrix A if there is a non trivial solution of $Ax = \lambda x$. Such a vector x is called an eigen vector corresponding to the eigen value λ.

- Thus an eigen vector of an $n \times n$ matrix A is a non-zero vector x such that $Ax = \lambda x$ for some scalar λ.

- If v is an eigen vector corresponding to an eigen value λ_1, then it cannot be an eigen vector corresponding to some different eigen value λ_2.

For if $Av = \lambda_1 v$ and $Av = \lambda_2 v \Rightarrow \lambda_1 v = \lambda_2 v \Rightarrow (\lambda_1 - \lambda_2) v = 0$

If $\lambda_1 \neq \lambda_2$, $\lambda_1 - \lambda_2 \neq 0$ and so $v = 0$, not possible for an eigen vector.

- Eigen values are sometimes called characteristic values or latent roots.

  Eigen vectors are sometimes called characteristic vectors.

Note:
- The "zero vector" is not considered as an eigen vector since $A0 = \lambda 0$ for all matrices A and all scalars λ.
- However 0 is allowed to be an eigen value for a matrix A. (Note that $Ax = 0.x \Rightarrow Ax = 0$)

Proposition: An $n \times n$ matrix A is invertible if and only if 0 is not an eigen value for A.

Proof: Let 0 be an eigen value for A.

Then the equation $Ax = 0.x$ has a nontrivial solution.

But $Ax = 0$ has a nontrivial solution if and only if A is not invertible.

Therefore an $n \times n$ matrix A is invertible if and only if 0 is not an eigen value of A.

- Thus we have another condition to add to our first theorem (of the course).

Note: An eigen vector is <u>not unique</u> since all scalar multiples of an eigen vectors are also eigen vectors (corresponding to the same eigen value) $Ax = \lambda x \Rightarrow A(cx) = c(Ax)$
$= c(\lambda x) = \lambda(cx)$

Proposition: Let A be an $n \times n$ matrix and $V = F^n$

Then the set $X = \{ v \in V : v$ is an eigen vector of A corresponding to $\lambda \} \cup \{0\}$

$$= \{ v \in V : Av = \lambda v \}$$

is a subspace of V.

Proof: • Let $v_1, v_2 \in X$. Then $Av_1 = \lambda v_1$ and $Av_2 = \lambda v_2$

$\Rightarrow A(v_1 + v_2) = \lambda v_1 + \lambda v_2 = \lambda(v_1 + v_2)$

$\Rightarrow v_1 + v_2 \in X$

Similarly if $v \in X$ and $c \in F$, then $Av = \lambda v \Rightarrow A(cv) = cAv = c(\lambda v)$

$\Rightarrow A(cv) = \lambda(cv)$

$\Rightarrow cv \in X$

So, $\boxed{X \text{ is a subspace of } V}$

- <u>Another Proof:</u>

Note that $v \in X \Leftrightarrow Av = \lambda v \Leftrightarrow Av - \lambda v = 0 \Leftrightarrow (A - \lambda I)v = 0$
$\Leftrightarrow v \in \mathrm{Nul}(A - \lambda I)$

Hence $X = \mathrm{Nul}(A - \lambda I)$ and therefore X is a subspace of V

Note: The subspace X defined above is called the <u>eigen space</u> of A <u>corresponding to λ</u>.

Fundamental Result about Eigenvectors and Eigen Values

Proposition: If $v_1, v_2, \ldots, v_p$ are eigen vectors corresponding to distinct eigen values $\lambda_1, \lambda_2, \ldots, \lambda_p$ of the matrix A, then the set $\{v_1, v_2, \ldots, v_p\}$ is linearly independent.

Corollary: An $n \times n$ matrix A can have atmost n distinct eigen values.

Proof of the Proposition:

Proof will be by contradiction:

Assume that $v_1, v_2, \ldots, v_p$ are linearly dependent.

Let m be the smallest number such that $v_1, v_2, \ldots, v_m$ are linearly independent and v_{m+1} is a linear combination of the preceeding vectors.

Then there exist scalars $c_1, c_2, \ldots c_m \in F$ such that $c_1 v_1 + c_2 v_2 + \ldots + c_m v_m = v_{m+1}$ .... (1)

Then $A(c_1 v_1 + c_2 v_2 + \ldots + c_m v_m) = A v_{m+1}$

$\Rightarrow \quad c_1 A v_1 + c_2 A v_2 + \cdots + c_m A v_m = A v_{m+1}$

$\Rightarrow \quad c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_m \lambda_m v_m = \lambda_{m+1} v_{m+1}$ ..... (2)

Now multiplying ① by λ_{m+1}, we get

$$c_1 \lambda_{m+1} v_1 + c_2 \lambda_{m+1} v_2 + \cdots + c_m \lambda_{m+1} v_m = \lambda_{m+1} v_{m+1} \quad \cdots\cdots ③$$

Now ② − ③ ⇒

$$c_1(\lambda_1 - \lambda_{m+1}) v_1 + c_2(\lambda_2 - \lambda_{m+1}) v_2 + \cdots + c_m(\lambda_m - \lambda_{m+1}) v_m = 0$$

Since $v_1, v_2, \ldots, v_m$ are <u>linearly independent</u>,

$$c_1(\lambda_1 - \lambda_{m+1}) = c_2(\lambda_2 - \lambda_{m+1}) = \cdots = c_m(\lambda_m - \lambda_{m+1}) = 0$$

But $\lambda_1 - \lambda_{m+1} \neq 0$, $\lambda_2 - \lambda_{m+1} \neq 0, \ldots, \lambda_m - \lambda_{m+1} \neq 0$

Hence $c_1 = c_2 = \cdots = c_m = 0$

and so from ① we conclude $v_{m+1} = 0$

But v_{m+1} is an eigen vector of A, corresponding to the eigen value λ_{m+1} and so $v_{m+1} \neq 0$, a contradiction.

Therefore $\boxed{\{v_1, v_2, \ldots, v_p\} \text{ is linearly independent}}$

MTH 100 : Lecture 31

How to Determine Eigen Values and Eigen Vectors :

Note: It is easy to verify whether a particular vector is an eigen vector of a given matrix A or not.

Similarly, given some scalar, we can verify whether it is an eigen value or not

- However we need to find a systematic method to find eigen values.

Ex: Let $A = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix}$

Let $v = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, $v_1 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$, $v_2 = \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}$

$v_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

To check if v is an eigen vector of A :

Calculate:

$$Av = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

$$= \begin{bmatrix} 20 \\ -11 \\ 38 \end{bmatrix}$$

But $\begin{bmatrix} 20 \\ -11 \\ 38 \end{bmatrix} \neq \lambda \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ for any $\lambda \in F$.

So, $Av \neq \lambda v$ for any λ.

So, v is not an eigen vector of A.

Now $Av_1 = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = 1 . \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$

So, v_1 is an eigen vector corresponding to the eigen value 1.

Again, $Av_2 = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix} = 1 . \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}$

So, v_2 is an eigen vector corresponding to the eigen value 1.

Now $Av_3 = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

So, v_3 is an eigen vector corresponding to the eigen value 0.

Now $v_4 = v_1 + v_2 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 11 \end{bmatrix}$

and $Av_4 = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 11 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 11 \end{bmatrix} = 1 . \begin{bmatrix} 3 \\ 1 \\ 11 \end{bmatrix}$

So, v_4 is (as expected) an eigen vector corresponding to the eigen value 1.

<u>Ex:</u> $A = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix}$

Let $\lambda = 3$: can we find out if 3 is an eigen value of A or not.

If λ is an eigen value of A, then there exists a vector $v \in \mathbb{R}^3$, $v \neq 0$ such that

$$Av = \lambda v \Rightarrow Av - \lambda v = 0$$
$$\Rightarrow Av - \lambda I v = 0$$
$$\Rightarrow (A - \lambda I)v = 0$$

i.e. the homogeneous system $(A-\lambda I)X = 0$ has a nontrivial solution.

Now $A - \lambda I$

$$= A - 3I = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -1 \\ -3 & -4 & 1 \\ 6 & 4 & -4 \end{bmatrix}$$

$$\xrightarrow[R_3 \to R_3 - 6R_1]{R_2 \to R_2 + 3R_1} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 2 & -2 \\ 0 & -8 & 2 \end{bmatrix} \xrightarrow{R_2 \to \frac{1}{2}R_2} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -1 \\ 0 & -8 & 2 \end{bmatrix} \xrightarrow{R_3 \to R_3 + 8R_2} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & -6 \end{bmatrix}$$

$$\xrightarrow[R_3 \to -\frac{1}{6}R_3]{R_1 \to R_1 - 2R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[R_2 \to R_2 + R_3]{R_1 \to R_1 - R_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

= RREF matrix

So, $(A - 3I)x = 0$ has only trivial solution.

Therefore $\lambda = 3$ is not an eigen value of A.

Proposition: A scalar λ is an eigen value of an $n\times n$ matrix A if and only if λ satisfies the characteristic equation $\det(A-\lambda I)=0$

(**Note:** Characteristic Equation of A : $\det(A-\lambda I)=0$
Characteristic Polynomial of A : $\det(A-\lambda I)$)

Proof: λ is an eigen vector of A

$\Leftrightarrow$ There is a non zero vector v such that
$$Av=\lambda v$$

$\Leftrightarrow$ The system $(A-\lambda I)x=0$ has a non-trivial solution

$\Leftrightarrow$ The matrix $(A-\lambda I)$ is not invertible (By the first theorem of the course)

$\Leftrightarrow$ $\det(A-\lambda I)=0$ (By the extended version of the first theorem)

$\Leftrightarrow$ λ is a root of the characteristic equation.

Ex: Given $A=\begin{bmatrix}4 & 2 & -1\\ -3 & -1 & 1\\ 6 & 4 & -1\end{bmatrix}$

Find the characteristic polynomial of A and eigen values of A.

Characteristic Polynomial of A

$$=\det(A-\lambda I)=\begin{vmatrix}4-\lambda & 2 & -1\\ -3 & -1-\lambda & 1\\ 6 & 4 & -1-\lambda\end{vmatrix}$$

$$=(4-\lambda)\left[(-1-\lambda)(-1-\lambda)-4\right]+2\left[1\times 6-(-3)(-1-\lambda)\right]+(-1)\left[(-3)(4)-6(-1-\lambda)\right]$$

$= (4-\lambda)(1+2\lambda+\lambda^2-4) + 2(3-3\lambda) - (-6+6\lambda)$

$= (4-\lambda)(\lambda^2+2\lambda-3) + 2(3-3\lambda) - (-6+6\lambda)$

$= 4\lambda^2 + 8\lambda - \cancel{12} - \lambda^3 - 2\lambda^2 + 3\lambda + \cancel{6} - 6\lambda + \cancel{6} - 6\lambda$

$= \boxed{-\lambda^3 + 2\lambda^2 - \lambda}$

Now the characteristic polynomial of A

$= -\lambda^3 + 2\lambda^2 - \lambda \;=\; -\lambda(\lambda^2 - 2\lambda + 1)$

$= -\lambda(\lambda-1)^2$

Hence the eigen values of A are 0 and 1.

<u>Note:</u> • 0 is an eigen value of A with <u>multiplicity 1</u>

and 1 is an eigen value of A with <u>multiplicity 2</u>.

<u>Note:</u> • det $(A-\lambda I)$ is a polynomial of degree n and it is called the characteristic polynomial of A.

• It has atmost n roots, counting multiplicities. Hence an $n \times n$ matrix can have atmost n eigen values (counting multiplicities)

• It is possible for a matrix with real entries to have no real eigen values.

<u>Ex:</u> Given $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

Then $\det(A-\lambda I) = \begin{vmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - (-1)(1)$

$$= \lambda^2 - 2\lambda + 1 + 1 = \lambda^2 - 2\lambda + 2$$

So, the characteristic equation is $\lambda^2 - 2\lambda + 2 = 0$

$$\Rightarrow \quad \lambda = \frac{2 \pm \sqrt{(-2)^2 - 4(2)(1)}}{2(1)} = \frac{2 \pm \sqrt{-4}}{2} = \frac{2 \pm 2i}{2}$$

$$\Rightarrow \quad \lambda = 1 \pm i$$

Note: If complex roots are allowed, an $n \times n$ matrix has exactly n eigen values (counting multiplicities): This follows from the so-called Fundamental Theorem of Algebra. Therefore, we must clearly specify which field is being considered when we talk about the eigen values of a matrix.

Eigenvalues of Similar Matrices

- Recall that an $n \times n$ matrix B is similar to an $n \times n$ matrix A if there exists an invertible matrix P such that $B = PAP^{-1}$ (or $A = P^{-1}BP$).

- Note that if A and B are similar matrices then $B = PAP^{-1}$ for some invertible matrix P and so $\det B = \det(PAP^{-1})$

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

$$= \det A$$

• Proposition: If the $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues with the same multiplicities.

Proof:
$$\begin{aligned}
\det(B-\lambda I) &= \det(PAP^{-1}-\lambda I) \\
&= \det(PAP^{-1}-P(\lambda I)P^{-1}) \\
&= \det(P(A-\lambda I)P^{-1}) \\
&= (\det P)\det(A-\lambda I)\det(P^{-1}) \\
&= \det(A-\lambda I)(\det P)\det(P^{-1}) \\
&= \det(A-\lambda I)\cdot 1 \\
&= \det(A-\lambda I)
\end{aligned}$$

(QED)

Note: The eigenvectors of A and B are not necessarily the same.

$\boxed{\text{MTH 100 : Lecture 32}}$

Polynomials Applied to Matrices

- **Matrix Powers:** If A is an $n \times n$ matrix (we may take the entries either real or complex), then the product matrix $A.A$ is well-defined and can be written as A^2, $A.A.A = A^3$, ...

In general, $A^m = A.A \dots A$ (m times) for any positive integer m.

For convenience, we define $A^0 = I_n$, the identity matrix.

- If A is invertible and A^{-1} is its inverse, then for any positive integer m, $\boxed{(A^m)^{-1} = (A^{-1})^m}$

- **Remark:**

$$A^i . A^j = A^{i+j} \text{ and } (A^i)^j = A^{ij}$$

where i, j can be arbitrary integers if A is invertible and non-negative integers if A is not invertible.

Definition: If p is a polynomial given by

$$p(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m$$

and A is an $n \times n$ matrix, then $p(A)$ is the matrix given by

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \dots + a_m A^m$$

- Note that this is a new use of the symbol p because we are applying it to matrices and not just scalars.
- If p and q are two polynomials, then

$$pq(A) = p(A)q(A) = q(A)p(A) = (qp)(A)$$

(where pq is the polynomial defined by $pq(t) = p(t)q(t)$ (usual multiplication of polynomials))

The Minimal Polynomial of a Matrix

Definition: Given an $n \times n$ matrix A, the minimal polynomial of A is the (non zero) monic polynomial of minimal degree such that $p(A) = 0$ (i.e. the zero matrix)

(Monic polynomial means the nonzero coefficient of highest power of the variable is equal to 1. The monic condition is inserted so as to make the minimal polynomial unique.)

Note: Every square matrix must have a minimal polynomial:

Suppose A is an $n \times n$ matrix with entries from a field F, then the set $\{I, A, A^2, \ldots, A^{n^2}\}$ cannot be linearly independant because this set has $\boxed{n^2+1}$ matrices and $\dim(F^{n\times n}) = \boxed{n^2}$

Let m be the smallest positive integer such that $\{I, A, A^2, \ldots, A^m\}$ is linearly dependent. Then A^m is a linear combination of the preceding matrices.

Thus there exist scalars $a_0, a_1, \ldots, a_{m-1}$ such that

$$\boxed{a_0 I + a_1 A + a_2 A^2 + \cdots + a_{m-1} A^{m-1} + A^m = 0}$$

Note: Reference: Problem 5 of Worksheet 6.

A Famous Result

Theorem (Caley - Hamilton Theorem):

Let q denote the characteristic polynomial of an $n \times n$ matrix A.

Then $\boxed{q(A) = 0}$

Corollary: The degree of minimal polynomial of any $n \times n$ matrix is atmost n.

(Recall that the degree of the characteristic polynomial of A is n.)

Note: We will omit the proof of Caley-Hamilton Theorem. You may refer to advanced textbooks.

Remark:

Using Remainder Theorem for Polynomial Division, we can see that if $p(x)$ is the minimal polynomial of A and if $q(x)$ is any other polynomial satisfied by A, then $p(x)$ divides $q(x)$.

Review of Polynomials:

- We use the notation $F[t]$ to indicate the vector space of polynomials with coefficients from the field F (F could be either $\mathbb{R}$ or $\mathbb{C}$)
- $\lambda \in F$ is called a root of a polynomial $p(t)$ if $p(\lambda) = 0$

- **Lemma:** Suppose $p \in F[t]$ is a polynomial of degree $m \geq 1$.
  Then λ is a root of p if and only if there exists a polynomial $q \in F[t]$ with degree $(m-1)$ such that

$$\boxed{p(t) = (t-\lambda)\, q(t)}$$

- **Lemma:** Suppose $p \in F[t]$ is a polynomial of degree $m \geq 0$, then p has atmost m distinct roots in F.

- **Lemma (Division Algorithm or Remainder Theorem):**
  Suppose $p, q \in F[t]$ with $p \neq 0$.
  Then there exist polynomials $r, s \in F[t]$ with $q(t) = p(t)s(t) + r(t)$ and either $r = 0$ or $\deg r < \deg p$.

- **Fundamental Theorem of Algebra:**
  Suppose $p \in \mathbb{C}[t]$ is a polynomial of degree $m \geq 1$. Then p has a root.
  Furthermore, p has a factorization of the form $\boxed{p(t) = c(t-\lambda_1)(t-\lambda_2)\cdots\cdots(t-\lambda_m)}$

Lemma: Suppose $q \in F[t]$.

Then $q(A)=0$ if and only if the minimal polynomial of A divides q.

Proof: **$\Rightarrow$:** Suppose $q(A)=0$

Let $p(t)$ be the minimal polynomial of A.

Then $p \neq 0$.

Using Remainder Theorem, we can write

$$q(t) = p(t)s(t) + r(t) \quad \ldots\ldots (1)$$

where either $r=0$ or $\deg r < \deg p$

Evaluating (1) at A,

$$q(A) = p(A)s(A) + r(A)$$

Since $\underline{q(A)=0}$, $\quad p(A)s(A) + r(A) = 0 \quad \ldots\ldots (2)$

Now, since p is the <u>minimal polynomial</u> of A, $p(A)=0$ and hence from (2), $r(A)=0$

But this is not possible unless $r=0$ because $\deg r < \deg p$ and p is the <u>minimal polynomial</u> of A

Therefore from (1), $q(t) = p(t)s(t)$ and so p divides q.

$\Leftarrow$ If $p(t)$ is the minimal polynomial of A, then $p(A) = 0$.

If $p(t)$ divides $q(t)$, there exists a polynomial $s(t)$ such that

$$q(t) = p(t)s(t)$$

Evaluating at A,

$$q(A) = p(A)s(A) = 0 \cdot s(A) = 0$$

$$\Rightarrow \boxed{q(A) = 0}$$

(QED).

Diagonalization of Matrices:

- If A is a diagonal matrix, then its diagonal elements are its eigen values and the standard basis vectors are its eigen vectors.
- i.e. if $A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ then $Ae_i = \lambda_i e_i$ where $e_i = \begin{array}{c} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \end{array} \rightarrow$ ith entry

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

Definition:

An $n \times n$ matrix A is said to be diagonalizable if A is similar to a diagonal matrix D. i.e. if there exists an invertible matrix P and a diagonal matrix D such that

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

Note: If A is diagonalizable then its powers are easy to compute.

Note: If A is diagonalizable, then its eigen values can be found by inspection of D. However, in practice, we have to do things the other way round. First, we find the eigenvalues from the characteristic equation, then we find P and the diagonal matrix D.

MTH 100 : Lecture 33

Theorem (Diagonalization Theorem):

(a) An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

(b) In this case, $A = PDP^{-1}$ where the columns of P are n linearly independent eigen vectors of A, and the diagonal entries of D are eigenvalues corresponding to these eigenvectors.

Another way to express the above Theorem

An $n \times n$ matrix A is diagonalizable if and only if it has enough (linearly independent) eigenvectors to form a basis of $\mathbb{R}^n$.

Such a basis is called an eigenvector basis.

Note: Check if the matrix considered in the last lecture is diagonalizable.

Proof of the Diagonalization Theorem:

(a) $\Rightarrow$: Suppose A is diagonalizable
Want to show that A has n linearly independent eigen vectors.

Now $A = PDP^{-1}$ for some diagonal matrix D and some invertible matrix P.

Therefore $AP = PD$ ...... (1)

Let $P = [v_1 \ v_2 \ \ldots \ v_n]$ in column form

and Let $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ where the λ's need not be distinct.

Then (1) becomes

$$A[v_1 \ v_2 \ \ldots \ v_n] = [v_1 \ v_2 \ \ldots \ v_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\Rightarrow [Av_1 \ Av_2 \ \ldots \ Av_n] = [\lambda_1 v_1 \ \lambda v_2 \ \ldots \ \lambda v_n]$$

$$\Rightarrow \quad Av_i = \lambda_i v_i \quad \text{for } i = 1, 2, \ldots, n$$

Therefore $v_1, v_2, \ldots, v_n$ are eigen vectors corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$

Also the vectors $v_1, v_2, \ldots, v_n$ being columns of an invertible matrix P are linearly independent.

$\underline{\Leftarrow}$: Suppose A has n linearly independent eigen vectors.

Want to show that A is diagonalizable.

Let $v_1, v_2, \ldots, v_n$ be the n linearly independent eigen vectors of A corresponding to the eigen values $\lambda_1, \lambda_2, \ldots, \lambda_n$ (not necessarily distinct).

Then $Av_i = \lambda_i v_i$ for $i = 1, 2, \ldots, n$

We form the matrix P with the v_i's as columns.

i.e. $P = [v_1 \; v_2 \; \cdots \; v_n]$

Then
$$\begin{aligned}
AP &= A[v_1 \; v_2 \; \cdots \; v_n] \\
&= [Av_1 \; Av_2 \; \cdots \; Av_n] \\
&= [\lambda_1 v_1 \; \lambda_2 v_2 \cdots, \lambda_n v_n] \\
&= [v_1 \; v_2 \; \cdots \; v_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\
&= PD \quad \text{where } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}
\end{aligned}$$

Now since $v_1, v_2, \ldots, v_n$ are linearly independent, the matrix $P = [v_1 \; v_2 \cdots v_n]$ is invertible and $\boxed{A = PDP^{-1}}$. Hence A is diagonalizable.

(b) Part (b) of the theorem has been proved enroute to proving part (a) of the theorem. (QED)

In practice, we can distinguish three cases:

- Case 1: An $n \times n$ matrix A has n distinct (real) eigen values.

Then we have the following result.

Proposition: An $n \times n$ matrix A with n distinct eigen values is diagonalizable.

Proof: By an earlier proposition, eigen vectors corresponding to distinct eigen-values are linearly independent. Therefore in this case A has n linearly independent eigen vectors. Hence by Diagonalization Theorem, A is diagonalizable.

Two Preliminary Definitions:

Given an eigenvalue λ_1 for a matrix A we define

- The algebraic multiplicity of λ_1 is the power of the factor $(\lambda-\lambda_1)$ in the characteristic polynomial of A.

- The <u>geometric multiplicity of λ_1</u> is the dimension of the eigen space corresponding to λ_1.

<u>Note:</u> Algebraic multiplicity applies to polynomials in general (not only characteristic polynomial). The geometric multiplicity applies specifically to the characteristic polynomial (since its roots are eigenvalues which have corresponding eigenspaces).

<u>Case 2:</u> An $n \times n$ matrix A has $p < n$ distinct eigenvalues, but counting the (algebraic) multiplicities, there are <u>n real</u> eigenvalues (<u>not distinct</u>).

We then have a weaker result for this case.

<u>Proposition:</u> Let A be an $n \times n$ matrix with n (real) eigenvalues (counting algebraic multiplicities) of which only $\lambda_1, \lambda_2, \dots, \lambda_p$ are distinct $(p < n)$.

Then the following hold:

(a) For $1 \le k \le p$, the <u>geometric multiplicity</u>

of λ_k is <u>less than or equal</u> to the <u>algebraic multiplicity</u> of λ_k.

(b) A is diagonalizable if and only if <u>the sum of the dimensions</u> of the <u>distinct eigenspaces is n</u> and this happens if and only if the <u>geometric multiplicity</u> for each λ_k equals its <u>algebraic multiplicity</u>.

(c) If A is diagonalizable and B_k is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in $B_1, B_2, \ldots, B_p$ forms an eigenvector basis for $\mathbb{R}^n$.

<u>Note:</u> This result is weaker because unlike case 1, A is not automatically diagonalizable. A has to satisfy the additional condition (b) and this may not happen for all matrices.

Ex (1) Let $A = \begin{bmatrix} 42 & -33 \\ 22 & -13 \end{bmatrix}$

Then $\det(A-\lambda I) = \begin{vmatrix} 42-\lambda & -33 \\ 22 & -13-\lambda \end{vmatrix}$

$$= (42-\lambda)(-13-\lambda) + 22(33)$$
$$= -546 - 42\lambda + 13\lambda + \lambda^2 + 726$$
$$= 180 - 29\lambda + \lambda^2 = (\lambda-20)(\lambda-9)$$

Thus here there are 2 distinct eigen values

$\lambda_1 = 20$, $\lambda_2 = 9$

For $\lambda_1 = 20$ $\quad A-\lambda_1 I = \begin{bmatrix} 42-20 & -33 \\ 22 & -13-20 \end{bmatrix}$

$$= \begin{bmatrix} 22 & -33 \\ 22 & -33 \end{bmatrix} \xrightarrow[R_2 \to R_2 - R_1]{} \begin{bmatrix} 22 & -33 \\ 0 & 0 \end{bmatrix} \xrightarrow[R_1 \to \frac{1}{22}R_1]{} \begin{bmatrix} 1 & -\frac{3}{2} \\ 0 & 0 \end{bmatrix} = \text{RREF matrix}$$

So, $(A-\lambda_1 I)x = 0 \Rightarrow \left.\begin{array}{l} x_1 = \frac{3}{2}x_2 \\ x_2 = x_2 \end{array}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix}$

Let us take $v_1 = 2\begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as an eigen vector corresponding to $\lambda_1 = 20$
(by taking $x_2 = 2$)

(Check $A v_1 = \begin{bmatrix} 42 & -33 \\ 22 & -13 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 60 \\ 40 \end{bmatrix} = 20\begin{bmatrix} 3 \\ 2 \end{bmatrix} = 20 v_1$)

For $\lambda_2 = 9$ $\quad A - \lambda_2 I = \begin{bmatrix} 42-9 & -33 \\ 22-9 & -13-9 \end{bmatrix} = \begin{bmatrix} 33 & -33 \\ 22 & -22 \end{bmatrix}$

$\xrightarrow{R_1 \to \frac{1}{33} R_1} \begin{bmatrix} 1 & -1 \\ 22 & -22 \end{bmatrix} \xrightarrow{R_2 \to R_2 - 22R_1}$

RREF matrix $= \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$

So, $(A - \lambda_2 I)x = 0 \Rightarrow \left.\begin{array}{l} x_1 = x_2 \\ x_2 = x_2 \end{array}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Let us take $v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as an eigen-vector corresponding to eigenvalue $\lambda_2 = 9$

(By taking $x_2 = 1$)

(Check $Av_2 = \begin{bmatrix} 42 & -33 \\ 22 & -13 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 9 \end{bmatrix} = 9v_2$)

Note that we should get $A = PDP^{-1}$

where $D = \begin{bmatrix} 20 & 0 \\ 0 & 9 \end{bmatrix}$ and $P = [v_1, v_2]$

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

Easier to check $AP = PD$

Now $AP = \begin{bmatrix} 42 & -33 \\ 22 & -13 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 60 & 9 \\ 40 & 9 \end{bmatrix}$

and $PD = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 20 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 60 & 9 \\ 40 & 9 \end{bmatrix}$

So, $AP = PD$ as desired.

MTH 100 : Lecture 34

Ex: $A = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix}$

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 2 & -1 \\ -3 & -1-\lambda & 1 \\ 6 & 4 & -1-\lambda \end{vmatrix} = -\lambda^3 + 2\lambda^2 - \lambda$$

$$= -\lambda(\lambda^2 - 2\lambda + 1)$$

(as in a previous problem)

$$= (-\lambda)(1-\lambda)^2$$

So, the eigen values are

$\lambda_1 = 1$ with algebraic multiplicity 2

and $\lambda_2 = 0$ with algebraic multiplicity 1

For $\lambda_1 = 1$

$$A - \lambda_1 I = \begin{bmatrix} 4-1 & 2 & -1 \\ -3 & -1-1 & 1 \\ 6 & 4 & -1-1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \\ -3 & -2 & 1 \\ 6 & 4 & -2 \end{bmatrix}$$

$$\xrightarrow{\substack{R_1 \to R_1 + R_2 \\ R_3 \to R_3 + 2R_2}} \begin{bmatrix} 0 & 0 & 0 \\ -3 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} -3 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{R_1 \to -\frac{1}{3} R_1} \begin{bmatrix} 1 & \frac{2}{3} & -\frac{1}{3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{RREF matrix}$$

So, $(A - \lambda_1 I)X = 0 \Rightarrow x_1 + \frac{2}{3}x_2 - \frac{1}{3}x_3 = 0$

$$x_2 = x_2$$

$$x_3 = x_3$$

So, $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{2}{3} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{1}{3} \\ 0 \\ 1 \end{bmatrix}$

Taking, $x_2 = 0$, $x_3 = 3$ we have $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$

Taking, $x_2 = 1$, $x_3 = 8$ we have $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}$

<u>Now for $\lambda_2 = 0$</u>

$$A - \lambda_2 I = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \xrightarrow{R_1 \to \frac{1}{4}R_1} \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{4} \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix}$$

$$\xrightarrow[R_3 \to R_3 - 6R_1]{R_2 \to R_2 + 3R_1} \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & \frac{1}{2} \end{bmatrix} \xrightarrow{R_2 \to 2R_2} \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{bmatrix}$$

$$\xrightarrow[R_3 \to R_3 - R_2]{R_1 \to R_1 - \frac{1}{2}R_2} \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} = \text{RREF matrix}$$

So, $(A - \lambda_2 I)x = 0 \Rightarrow \left.\begin{aligned} x_1 - \tfrac{1}{2}x_3 &= 0 \\ x_2 + \tfrac{1}{2}x_3 &= 0 \\ x_3 &= x_3 \end{aligned}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$

Taking $x_3 = 2$ we get $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

Now in this example,
geometric multiplicity of λ_1 is equal to its algebraic multiplicity $= 2$

and geometric multiplicity of λ_2 is equal to its algebraic multiplicity $= 1$

Therefore the matrix A is diagonalizable.

Check that $AP = DP$ where $P = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 3 & 8 & 2 \end{bmatrix}$

and $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$$AP = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -1 & 1 \\ 6 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 3 & 8 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 8 & 0 \end{bmatrix}$$

and
$$PD = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 3 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 8 & 0 \end{bmatrix}$$

Therefore $AP = PD$ as desired.

Ex: Look at the matrix of Worksheet 11 (Problem 10)
(H.W.) and find if it is diagonalizable.

Ex: Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 0 \end{bmatrix}$

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & 0 \\ 2 & 1-\lambda & 0 \\ 3 & 2 & -\lambda \end{vmatrix}$$

$$= -\lambda(1-\lambda)^2$$

So, the eigen values are
$\lambda_1 = 1$ with algebraic multiplicity 2
and $\lambda_2 = 0$ with algebraic multiplicity 1.

Now for $\lambda_1 = 1$:

$$A - \lambda_1 I = \begin{bmatrix} 1-1 & 0 & 0 \\ 2 & 1-1 & 0 \\ 3 & 2 & 0-1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 3 & 2 & -1 \end{bmatrix}$$

$$\xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 2 & -1 \end{bmatrix} \xrightarrow{R_1 \to \frac{1}{2}R_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 2 & -1 \end{bmatrix} \xrightarrow{R_3 \leftrightarrow R_2} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{R_2 \to R_2 - 3R_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \to \frac{1}{2}R_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

= RREF matrix

So, $(A-\lambda_1 I)x=0 \Rightarrow \left.\begin{aligned} x_1 &= 0 \\ x_2-\tfrac{1}{2}x_3 &= 0 \\ x_3 &= x_3 \end{aligned}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \end{bmatrix}$

Taking $x_3 = 2$ we get $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = v_1$ (say)

(Check that $Av_1 = 1.v_1$)

Now for $\lambda_2 = 0$

$$A-\lambda_2 I = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 0 \end{bmatrix} \xrightarrow[R_3 \to R_3 - 3R_1]{R_2 \to R_2 - 2R_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 0 \end{bmatrix}$$

$$\xrightarrow{R_3 \to R_3 - 2R_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{RREF matrix}$$

So, $(A-\lambda_2 I)x=0 \Rightarrow \left.\begin{aligned} x_1 &= 0 \\ x_2 &= 0 \\ x_3 &= x_3 \end{aligned}\right\}$

Hence $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Taking $x_3 = 1$ we get $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ as an eigen vector.

Now here

- geometric multiplicity of λ_1 is 1 but its algebraic multiplicity is 2 $\Big\}$ ..... (*)
- geometric multiplicity of λ_2 is 1 and its algebraic multiplicity is also 1

So, the matrix A is not diagonalizable.
(because of (*))

Last case for Diagonalization: Case 3:

An $n \times n$ matrix A has $p < n$ distinct eigenvalues, but even after counting the algebraic multiplicities, there are $< n$ real eigenvalues (p could even be 0).

Then A is not diagonalizable over the real field. If we want to diagonalize, we have to admit complex eigenvalues and eigenvectors.

Remark: Even if we admit complex eigenvalues and eigenvectors, a real matrix does not have to be diagonalizable. The case is quite complicated and we will not go into the details. However we will consider the case of 2×2 real matrix with a complex eigenvalue and describe the nature of such a matrix and its corresponding transformation (i.e. a linear operator on $\mathbb{R}^2$).

Basic Result for Complex Eigenvalues:

Suppose A is a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$, $b \neq 0$ and associated eigenvector v in $\mathbb{C}^2$.

Then $A = PBP^{-1}$ where $P = \begin{bmatrix} \mathrm{Re}\, v & \mathrm{Im}\, v \end{bmatrix}$

and $B = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$

- Furthermore, the transformation (left multiplication by B) corresponds to a rotation followed by a scaling.
- The rotation is through the angle ϕ between the positive x-axis and the ray from the origin to (a, b).
  The angle ϕ is called the argument of λ.
- The scaling is by the factor $r = |\lambda| = \sqrt{a^2 + b^2}$.
  The quantity $r = |\lambda|$ is known as the modulus of λ.

Ex: Let $A = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix}$

The characteristic polynomial

$$= \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -8 & 4-\lambda \end{vmatrix} = -\lambda(4-\lambda) - (-8)$$

$$= \lambda^2 - 4\lambda + 8$$

Then the eigen values are roots of $\lambda^2 - 4\lambda + 8 = 0$

$$\Rightarrow \quad \lambda = \frac{-(-4) \pm \sqrt{(-4)^2 - 4 \times 8}}{2 \times 1}$$

$$= \frac{4 \pm \sqrt{-16}}{2} = \frac{4 \pm 4i}{2} = 2 \pm 2i$$

We take $\lambda = 2 + 2i = 2 - (-2)i$ and so $\boxed{a=2}$ and $\boxed{b=-2}$

Then the matrix $A - \lambda I = \begin{bmatrix} -2-2i & 1 \\ -8 & 4-(2+2i) \end{bmatrix}$

$$= \begin{bmatrix} -2-2i & 1 \\ -8 & 2-2i \end{bmatrix}$$

Need to find the Corresponding Eigen Vector $v = \begin{bmatrix} x \\ y \end{bmatrix}$

$$(A - \lambda I)v = 0 \Rightarrow \begin{bmatrix} -2-2i & 1 \\ -8 & 2-2i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \left.\begin{aligned} (-2-2i)x + y &= 0 \\ -8x + (2-2i)y &= 0 \end{aligned}\right\}$$

- Since the system has a nontrivial solution, its two rows are linearly dependent i.e. the two equations represent the same relationship between x and y.
- Taking the first equation,

$$(-2-2i)x + y = 0$$

$$\Rightarrow \quad y = (2+2i)x$$

Taking $x = 1$, $y = 2+2i$.

Thus $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2+2i \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + i \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

Thus $P = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}$

$\begin{pmatrix} a = 2 \\ b = -2 \end{pmatrix}$

Now we can verify that $A = PBP^{-1}$

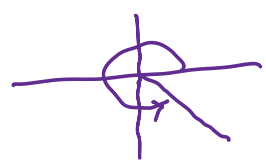
- $$PBP^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1/2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} = A$$

- Now $$B = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix} = \sqrt{8} \begin{bmatrix} \frac{2}{\sqrt{8}} & \frac{2}{\sqrt{8}} \\ -\frac{2}{\sqrt{8}} & \frac{2}{\sqrt{8}} \end{bmatrix}$$

$$= \sqrt{8} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

The matrix represents a rotation through $\frac{7\pi}{4}$



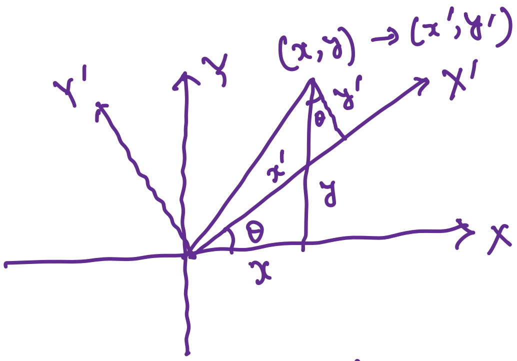
followed by a scaling through $\sqrt{8}$

- If we had taken $\lambda = 2-2i$ we would have obtained a different P and different B.

However $A = PBP^{-1}$ with these new P and B

Ref :

<u>coordinate geometry:</u>



Then $x = x'\cos\theta - y'\sin\theta$

$y = x'\sin\theta + y'\cos\theta$

So, $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$

and $\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}$

$\Rightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

<u>Ref</u> to the tutorial problem :

The Determinant of a Linear Operator

Definition: Let $T: V \to V$ be a linear operator where V is a vector space of finite dimension n.

Let α be any basis for V and let A be the matrix of T with respect to the (ordered) basis α. Then $\det T$ is defined as

$$\boxed{\det T = \det A}$$

Note: Suppose that β is any other basis for V and B is the matrix of T with respect to the basis β.

Now $B = PAP^{-1}$ where $P = P_{\alpha \to \beta}$ is the change of basis matrix.

Hence
$$\det B = \det(PAP^{-1}) = (\det P)(\det A)(\det P^{-1})$$
$$= \det A$$

Thus the definition given above is meaningful (independent of the basis taken)

Eigenvalues of Linear Operators

Definition: An eigenvector of a linear operator $T : V \longrightarrow V$ is a non-zero vector v such that $Tv = \lambda v$ for some scalar λ. Such a scalar is called an eigenvalue of the operator.

Note: If V is finite-dimensional, then the eigenvalues of a linear operator T coincide with the eigenvalues of the matrix of T with respect to any suitable basis of V

- This definition is useful for proving theoretical results and also for infinite-dimensional spaces.

MTH 100: Lecture 35

Inner Products:

Definition: An inner product on a (real) vector space V is a function, that to each pair of vectors u and v in V associates a scalar (real number) $\langle u, v\rangle$ and satisfies the following axioms:

1. $\langle u, v\rangle = \langle v, u\rangle \quad \forall\, u, v \in V$
2. $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle \quad \forall\, u, v, w \in V$
3. $\langle cu, v\rangle = c\langle u, v\rangle \quad \forall\, u, v \in V$ and $\forall\, c \in \mathbb{R}$
4. $\langle u, u\rangle \geq 0 \quad \forall\, u \in V$ and $\langle u, u\rangle = 0$ if and only if $u = 0$.

- A vector space with an inner product is called an inner product space.

Note: The above definition holds for real inner products. For complex inner products, the first axiom above becomes:

$$\langle u, v\rangle = \overline{\langle v, u\rangle} \quad \text{(the complex Conjugate)}$$

Definition: If we regard u, v in $\mathbb{R}^n$ as $n\times 1$ matrices (column vectors), then the transpose u^T is a $1\times n$ matrix (row vector). Then the matrix product $u^T v$ is a 1×1 matrix which is a real number. This real number is called the inner product or dot product and is written as $u\cdot v$

Examples of Inner product space:

(1) Let $V = \mathbb{R}^n$

For $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

define $\langle u, v\rangle = u\cdot v = u^T v$

$$= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Then the above is an inner product.

(1) $\langle u, v\rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$

$$= v_1 u_1 + v_2 u_2 + \cdots + v_n u_n$$

$$= [v_1 \; v_2 \; \cdots \; v_n] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \langle v, u\rangle$$

$\forall\, u, v \in \mathbb{R}^n$

(2) If $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ are in $\mathbb{R}^n$,

then $\langle u+v, w \rangle = \left\langle \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_n+v_n \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \right\rangle$

$$= (u_1+v_1)w_1 + (u_2+v_2)w_2 + \cdots + (u_n+v_n)w_n$$

$$= (u_1w_1 + u_2w_2 + \cdots + u_nw_n) + (v_1w_1 + v_2w_2 + \cdots + v_nw_n)$$

$$= \langle u, w \rangle + \langle v, w \rangle$$

(3) For $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$

and $c \in \mathbb{R}$,

$\langle cu, v \rangle = \left\langle c \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \right\rangle$

$$= \left\langle \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \right\rangle = (cu_1)v_1 + (cu_2)v_2 + \cdots + (cu_n)v_n$$

$$= c(u_1v_1) + c(u_2v_2) + \cdots + c(u_nv_n)$$

$$= c(u_1v_1 + u_2v_2 + \cdots + u_nv_n)$$

$$= c\langle u, v \rangle$$

(4) For any $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$,

$$\langle u, u \rangle = u_1 u_1 + u_2 u_2 + \cdots + u_n u_n$$
$$= u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0$$

Furthermore

$$\langle u, u \rangle = 0 \iff u_1^2 + u_2^2 + \cdots + u_n^2 = 0$$
$$\iff u_1 = 0, u_2 = 0, \ldots, u_n = 0$$
$$\iff u = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \bar{0} \text{ (zero vector)}$$

Therefore the product defined is an inner product.

<u>Example:</u>

The space $\mathbb{R}_n[t]$ of all polynomials of degree less than or equal to n can be made into an inner product space in the following way:

Let $t_0, t_1, t_2, \ldots, t_n$ be distinct real numbers (<u>note:</u> There are $(n+1)$ numbers).

For any two polynomials p and q in $\mathbb{R}_n[t]$, define $\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n)$

It can be shown that the four axioms for an inner product hold with the above definition. (exercise)

Why are $(n+1)$ points taken?

Note: The above inner product for polynomials is used when the values at specific points are important. (Interpolation problems)

Example: The space $C[a,b]$ of all continuous functions on the closed interval $[a,b]$ can be made into an inner product space with the following definition

$$\langle f, g \rangle = \int_a^b f(t)\, g(t)\, dt$$

It can be verified that the four axioms for an inner product hold with the above definition (exercise).

Note: The above inner product plays a very important role in the study of continuous functions and their applications in signals and systems.

Length and Distance in Inner Product Spaces

Definition: The <u>length</u> or <u>norm</u> of any vector u in an inner product space is the nonnegative number $\|u\| = \sqrt{\langle u,u\rangle}$

Note: In the case of $\mathbb{R}^n$, we get the length or norm as the nonnegative number

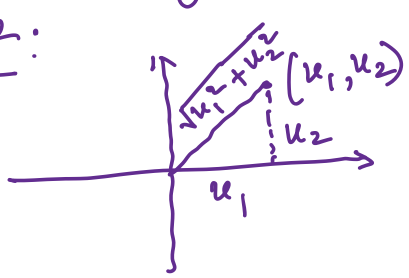
$$\|u\| = \sqrt{\langle u,u\rangle} = \sqrt{u_1^2+u_2^2+\cdots+u_n^2}$$

This coincides with the usual notion of length as the distance from the origin to the point (u_1,u_2) or (u_1,u_2,u_3) in $\mathbb{R}^2$ or $\mathbb{R}^3$.

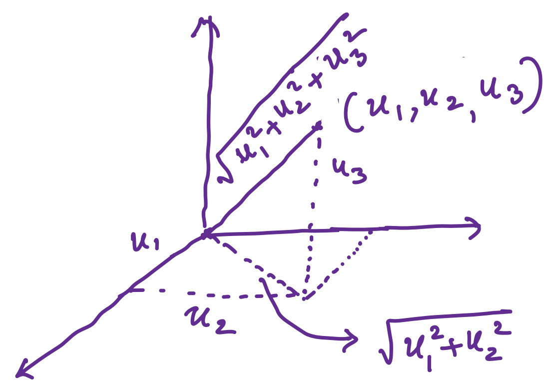
We can easily see that

for any scalar c, $\|cu\| = |c|\,\|u\|$

$\mathbb{R}^2$:



$\mathbb{R}^3$:



$$\|u\| \neq 0 \quad \left\| \frac{1}{\|u\|}\cdot u \right\| = \frac{1}{\cancel{\|u\|}}\cancel{\|u\|} = 1$$

- A vector whose length is <u>one</u> is called unit vector.

Given any non-zero vector u, the vector $\frac{u}{\|u\|}$ has norm <u>one</u> : This is called <u>normalizing</u>

- The distance between any two vectors u and v in V is defined as

$$\text{dist}(u,v) = \|u-v\| = \sqrt{\langle u-v, u-v\rangle}.$$

(check in $\mathbb{R}^2$ and $\mathbb{R}^3$)

<u>Proposition</u>: (The Cauchy-Schwarz inequality)

If V is an inner product space,

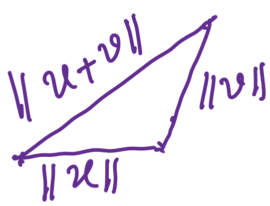
for all $u, v \in V$, $|\langle u,v\rangle| \leq \|u\|\,\|v\|$

(check in $\mathbb{R}^2$ and $\mathbb{R}^3$)

<u>Proposition</u> (The triangle inequality)

If V is an inner product space,

then for all $u, v \in V$, $\|u+v\| \leq \|u\| + \|v\|$



- The triangle inequality can be proved using Cauchy Schwarz inequality.

If $u, v \in V$, then

$$\|u+v\|^2 = \langle u+v, u+v\rangle$$

$$= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle$$

$$= \|u\|^2 + \langle u,v\rangle + \langle u,v\rangle + \|v\|^2$$

$$= \|u\|^2 + 2\langle u,v\rangle + \|v\|^2$$

$$\leq \|u\|^2 + 2|\langle u,v\rangle| + \|v\|^2$$

$$\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2$$ (By Cauchy-Schwarz inequality)

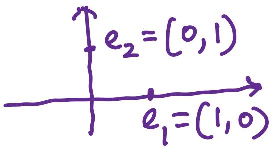
$$= (\|u\| + \|v\|)^2$$

$$\Rightarrow \boxed{\|u+v\| \leq \|u\| + \|v\|}$$

Orthogonality: Two vectors u, v in V are called orthogonal to each other if $\langle u, v \rangle = 0$

Notation for Orthogonality : $u \perp v$

Ex: In $\mathbb{R}^2$



$\langle e_1, e_2 \rangle = 1 \times 0 + 0 \times 1 = 0 + 0 = 0$

Hence $e_1 \perp e_2$

Note: The zero vector is orthogonal to every vector in V.

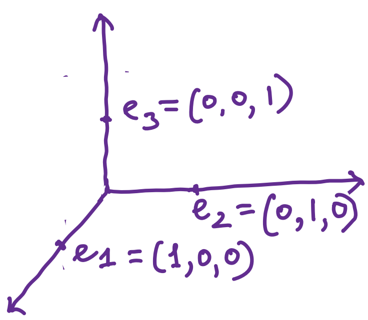
For every $u \in V$, $\langle \underset{\substack{\downarrow \\ \text{The zero vector}}}{0}, u \rangle = \langle \underset{\substack{\downarrow \\ \text{scalar zero}}}{0} \cdot u, u \rangle = 0 \langle u, u \rangle = 0$

$\Rightarrow \boxed{0 \perp u \ \forall u \in V}$

Definition: A set of vectors $\{u_1, u_2, \dots, u_p\}$ is said to be an Orthogonal set if any two distinct vectors in the set are orthogonal to each other

i.e. if $\langle u_i, u_j \rangle = 0$ whenever $i \neq j$

Ex: In $\mathbb{R}^3$



In $\mathbb{R}^3$, $\{e_1, e_2, e_3\}$ is an orthogonal set.

Proposition: An orthogonal set of <u>nonzero</u> vectors in V is linearly independent.

Proof: Suppose $S = \{u_1, u_2, \dots, u_p\}$ is an orthogonal set of nonzero vectors and suppose $c_1u_1 + c_2u_2 + \dots + c_pu_p = 0$

Then $\langle c_1u_1 + c_2u_2 + \dots + c_pu_p, u_1\rangle = \langle 0, u_1\rangle = 0$

$\Rightarrow c_1\langle u_1, u_1\rangle + c_2\langle u_2, u_1\rangle + \dots + c_p\langle u_p, u_1\rangle = 0$

$\Rightarrow c_1\langle u_1, u_1\rangle + c_2 \times 0 + \dots + c_p \times 0 = 0$

(Since $\langle u_i, u_1\rangle = 0$ for $i = 2, \dots p$)

$\Rightarrow c_1\langle u_1, u_1\rangle = 0$

Since $\langle u_1, u_1\rangle > 0$, we get $c_1 = 0$

Similarly $c_2 = c_3 = \dots = c_p = 0$

Therefore the set $S = \{u_1, u_2, \dots, u_p\}$ is linearly independent.

MTH 100 : Lecture 36

- Last time we defined inner product and inner product space. Gave various examples of inner product space.
- We have defined orthogonality of two vectors.

We have shown that an orthogonal set of nonzero vectors is linearly independent.

Definition: If W is a subspace of V, then a vector $v \in V$ is said to be orthogonal to W if v is orthogonal to every vector in W.
The set of all vectors orthogonal to W is called the orthogonal complement of W and is denoted by $W^{\perp}$ ('W-perp').

So, $W^{\perp} = \{ v \in V : v \perp w \ \forall w \in W \}$

Proposition:

(a) v belongs to $W^{\perp}$ if and only if v is orthogonal to every vector in a spanning set for W.

(b) $W^{\perp}$ is a subspace of V and $W \cap W^{\perp} = \{0\}$

Proof:

(a) $\Rightarrow$: If $v \in W^{\perp}$, v is actually orthogonal to every vector of W and hence orthogonal to every vector in a spanning set of W.

$\Leftarrow$: Suppose v is orthogonal to every vector in a spanning set K for W

Let w be any vector of W.

Then w can be written as

$$w = c_1 w_1 + c_2 w_2 + \cdots + c_p w_p$$

where $w_1, w_2, \ldots, w_p \in K$

Now $\langle w, v \rangle = \langle c_1 w_1 + c_2 w_2 + \cdots + c_p w_p , v \rangle$

$= c_1 \langle w_1, v \rangle + c_2 \langle w_2, v \rangle + \cdots + c_p \langle w_p, v \rangle$

$= 0$ (By hypothesis)

Therefore $v \in W^{\perp}$

(b) Since zero vector is orthogonal to every vector,

$\mathbf{0} \in W^{\perp}$
(zero vector)

Now let $v_1, v_2 \in W^{\perp}$

Then for any vector $w \in W$,

$$\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle = 0 + 0 = 0$$

and so $v_1 + v_2 \in W^{\perp}$

Now let $v \in W^{\perp}$ and $c \in \mathbb{R}$

Then for any vector $w \in W$,

$$\langle cv, w \rangle = c \langle v, w \rangle = c.0 = 0$$

and so $cv \in W^{\perp}$.

Therefore $W^{\perp}$ is a subspace of V.

If $\omega \in W \cap W^{\perp}$, then $\omega \in W$ and $\omega \in W^{\perp}$

$\Rightarrow \langle \omega, \omega \rangle = 0 \Rightarrow \omega = 0$

Thus $\boxed{W \cap W^{\perp} = \{0\}}$

<u>Note:</u> Actually if S is any subset of V, then $S^{\perp} = \{ v \in V : v \perp u \ \forall u \in S \}$ is a subspace of V (even if S is not a subspace).

<u>Proof:</u> Since zero vector is orthogonal to every vector, $0 \in S^{\perp}$ (zero vector)

Now let $v_1, v_2 \in S^{\perp}$

Then for any $u \in S$, $\langle v_1 + v_2, u \rangle = \langle v_1, u \rangle + \langle v_2, u \rangle$
$= 0 + 0 = 0$

$\Rightarrow v_1 + v_2 \in S^{\perp}$

Now let $v \in S^{\perp}$ and $c \in \mathbb{R}$

Then for any $u \in S$, $\langle cv, u \rangle = c \langle v, u \rangle = c.0 = 0$

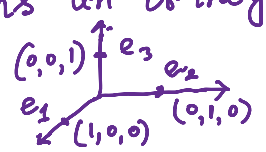
$\Rightarrow cv \in S^{\perp}$

Therefore $S^{\perp}$ is a subspace of V.

<u>Orthogonal Bases</u>

<u>Definition:</u>
- An orthogonal basis for a vector space V is a basis which is also an orthogonal set.
- An orthogonal basis for a subspace W is a <u>basis</u> which is also <u>an orthogonal set.</u>

<u>Example:</u> $\{e_1, e_2, e_3\}$ is an orthogonal basis for $\mathbb{R}^3$.



Note: • A set of vectors in a vector space is called Orthonormal if it is an orthogonal set and norm of each vector in the set is 1.

• A set of vectors in a vector space V is called an orthonormal basis if it is an orthogonal basis and norm of each vector in the basis is 1.

example: $\{e_1, e_2, e_3\}$ is an orthonormal basis for $\mathbb{R}^3$.

Proposition:

Let $\{u_1, u_2, \dots, u_p\}$ be an orthogonal basis for a subspace W. Then if $y = c_1u_1 + c_2u_2 + \dots + c_pu_p$ is any vector in W,

we have:
$$c_j = \frac{\langle y, u_j\rangle}{\langle u_j, u_j\rangle} \quad \text{for } j = 1, \dots, p$$

Proof: Since $y \in W$, y is uniquely expressible as a linear combination of the basis vectors.

i.e. $y = c_1u_1 + c_2u_2 + \dots + c_pu_p$

To determine the coefficients $c_1, c_2, \dots, c_p$ let us take the innerproduct with u_j for $1 \le j \le p$.

$$\text{Then } \langle y, u_j\rangle = c_1\langle u_1, u_j\rangle + c_2\langle u_2, u_j\rangle + \dots + c_j\langle u_j, u_j\rangle + \dots + \dots + c_p\langle u_p, u_j\rangle$$
$$= 0 + 0 + \dots + c_j\langle u_j, u_j\rangle + 0 + \dots + 0$$
$$= c_j\langle u_j, u_j\rangle$$

$$\Rightarrow \boxed{c_j = \frac{\langle y, u_j\rangle}{\langle u_j, u_j\rangle}}$$

Note: • The above proposition shows that it is easy to find the coordinates of a vector relative to an orthogonal basis it is only needed to take an inner product and divide by the inner product of the basis vector with itself.

• If it is an orthonormal basis, then the length of each basis vector is 1 and even the step of division is avoided.

Orthogonal Decomposition

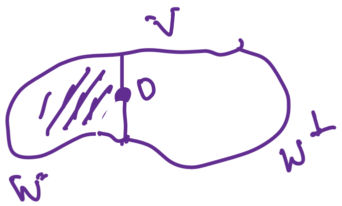
Theorem (Orthogonal Decomposition Theorem):

Let W be any finite-dimensional subspace of a vector space V. Then each vector y in V can be written uniquely in the form $y = \hat{y} + z$ where $\hat{y} \in W$ and $z \in W^{\perp}$

In fact, if $\{u_1, u_2, \dots, u_p\}$ is any orthogonal basis of W, then $\hat{y} = c_1 u_1 + c_2 u_2 + \cdots + c_p u_p$

where $c_j = \dfrac{\langle y, u_j\rangle}{\langle u_j, u_j\rangle}$ for $j = 1, 2, \dots, p$

and $z = y - \hat{y}$.



$y \in V$.

$y = \underbrace{\hat{y}}_{\in W} + z \in W^{\perp}$

Alternative Statement:

Given any finite-dimensional subspace W of V, we can then express $V = W + W^{\perp}$ with $W \cap W^{\perp} = \{0\}$

i.e. $V = W \oplus W^{\perp}$

Note: Thus every vector $v \in V$ can be uniquely expressed as a sum of a vector in W and a vector in $W^{\perp}$ i.e.. as the sum of two vectors which are orthogonal to each other.

Note: The vector $\hat{y}$ is called the orthogonal projection of y onto W and written as

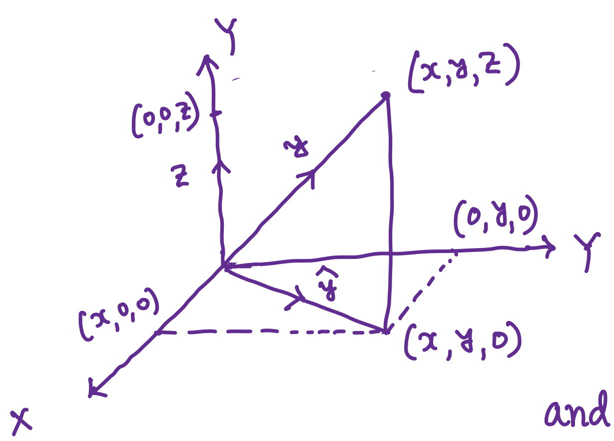
$\text{Proj}_W\, y = \hat{y}$ where $y = \hat{y} + z$ as in the Orthogonal Decomposition Theorem.

- In case $W = \text{Span}\{u\}$ is a one dimensional subspace of W, the expression for $\hat{y}$ is simplified to : $\hat{y} = \dfrac{\langle y, u\rangle}{\langle u, u\rangle} u$, which is simply called the orthogonal projection of y onto u.

Note: In case $y \in W$, its orthogonal projection onto W is y itself, i.e. $\boxed{\hat{y} = y \text{ for } y \in W}$

Example

Let $V = \mathbb{R}^3$, $W = \{(u, v, 0) : u, v \in \mathbb{R}\}$
i.e. W is the XY-plane.



For any $(x, y, z) \in \mathbb{R}^3$,

$(x, y, z) = (x, y, 0) + (0, 0, z)$

$= (x e_1 + y e_2) + z e_3$

Here $(x, y, 0) \in W$ and $(0, 0, z) \in W^\perp$

and $\boxed{\text{Proj}_W (x, y, z) = (x, y, 0)}$

Theorem (The Gram-Schmidt Process):

Given a basis $\{x_1, x_2, \dots, x_p\}$ for a subspace W of V, we can generate an orthogonal basis $\{v_1, v_2, \dots, v_p\}$ for W such that

$\text{Span}\{v_1, v_2, \dots, v_k\} = \text{Span}\{x_1, x_2, \dots, x_k\}$ for $k = 1, 2, \dots, p$.

In fact the vectors v_j are defined as follows:

$$v_1 = x_1$$

$$v_2 = x_2 - \left(\frac{\langle x_2, v_1\rangle}{\langle v_1, v_1\rangle}\right) v_1$$

$$v_3 = x_3 - \left(\frac{\langle x_3, v_1\rangle}{\langle v_1, v_1\rangle}\right) v_1 - \left(\frac{\langle x_3, v_2\rangle}{\langle v_2, v_2\rangle}\right) v_2$$

$$\vdots \qquad \vdots$$

$$v_p = x_p - \left(\frac{\langle x_p, v_1\rangle}{\langle v_1, v_1\rangle}\right) v_1 - \left(\frac{\langle x_p, v_2\rangle}{\langle v_2, v_2\rangle}\right) v_2 - \cdots\cdots \left(\frac{\langle x_p, v_{p-1}\rangle}{\langle v_{p-1}, v_{p-1}\rangle}\right) v_{p-1}$$

Note:

- At each stage, we subtract from the original basis vector x_i its projection onto the span of the previously obtained orthogonal vectors $v_1, v_2, \dots, v_{i-1}$.

- The process uses the idea we already used in Orthogonal Decomposition Theorem, of subtracting the orthogonal projection onto a subspace from the original vector.
- A formal proof that the vectors $\{v_1, v_2, \dots, v_k\}$ form an orthogonal set and that $\text{Span}\{v_1, v_2, \dots, v_k\} = \text{Span}\{x_1, x_2, \dots, x_k\}$ can be done by induction on k.
- We can obtain an <u>orthonormal basis</u> for every subspace W of V by normalizing each vector in an <u>orthogonal basis</u> (dividing each of the vectors by its norm). This step is usually left to the end because square roots can emerge.

<u>Ex:</u>

Construct an orthonormal basis for $\mathbb{R}^3$ starting with the basis using Gram-Schmidt process:

$$x_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

First put $v_1 = x_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

Then $v_2 = x_2 - \dfrac{\langle v_1, x_2 \rangle}{\langle v_1, v_1 \rangle} v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \dfrac{13}{14} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

$$= \begin{bmatrix} -\frac{6}{7} \\ \frac{15}{14} \\ \frac{3}{14} \end{bmatrix}$$

$$v_3 = x_3 - \frac{\langle v_1, x_3 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle v_2, x_3 \rangle}{\langle v_2, v_2 \rangle} v_2$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{6}{14} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \frac{-\frac{6}{7} + \frac{15}{14} + \frac{3}{14}}{\frac{12^2 + 15^2 + 3^2}{14^2}} \begin{bmatrix} -\frac{6}{7} \\ \frac{15}{14} \\ \frac{3}{14} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{7} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \frac{2}{9} \begin{bmatrix} -\frac{6}{7} \\ \frac{15}{14} \\ \frac{3}{14} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$$

<u>check</u>: v_1, v_2 and v_3 are orthogonal (whereas the original basis vectors x_1, x_2, x_3 were not).

$$\langle v_1, v_2 \rangle = \left\langle \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -\frac{6}{7} \\ \frac{15}{14} \\ \frac{3}{14} \end{bmatrix} \right\rangle = -\frac{12}{7} + \frac{15}{14} + \frac{9}{14}$$

$$= \frac{-24 + 24}{14} = 0$$

$$\langle v_2, v_3 \rangle = \left\langle \begin{bmatrix} -\frac{6}{7} \\ \frac{15}{14} \\ \frac{3}{14} \end{bmatrix}, \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \right\rangle = -\frac{6}{21} + \frac{15}{42} - \frac{3}{42}$$

$$= \frac{-12+15-3}{42} = 0$$

and $$\langle v_1, v_3 \rangle = \left\langle \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \right\rangle = \frac{2}{3} + \frac{1}{3} - 1$$

$$= 1 - 1 = 0$$

Note: If we want an orthonormal basis, we divide each vector v_i by its length $\|v_i\|$ to get:

$$v_1' = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad v_2' = \frac{1}{\sqrt{42}} \begin{bmatrix} -4 \\ 5 \\ 1 \end{bmatrix}, \quad v_3' = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Ex: Given an orthogonal basis $\mathcal{B}$:

$$v_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \text{ and } v_3 = \begin{bmatrix} -1 \\ 4 \\ -1 \end{bmatrix},$$

find the coordinate of $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ with respect to this basis β.

First check

$$\langle v_1, v_2 \rangle = 2 \times 1 + 1 \times 0 + 2 \times (-1) = 0$$

$$\langle v_2, v_3 \rangle = 1 \times (-1) + 0 \times 4 + (-1) \times (-1) = 0$$

$$\langle v_1, v_3 \rangle = 2 \times (-1) + 1 \times 4 + 2 \times (-1) = 0$$

Now if $v = c_1 v_1 + c_2 v_2 + c_3 v_3$

then $c_1 = \dfrac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} = \dfrac{2\times 1 + 1\times 2 + 2\times 3}{2^2 + 1^2 + 2^2}$

$= \dfrac{10}{9}$

$c_2 = \dfrac{\langle v, v_2 \rangle}{\langle v_2, v_2 \rangle} = \dfrac{1\times 1 + 0\times 2 + (-1)\times 3}{1^2 + 0^2 + 1^2} = -\dfrac{2}{2}$

$= -1$

$c_3 = \dfrac{\langle v, v_3 \rangle}{\langle v_3, v_3 \rangle} = \dfrac{1\times(-1) + 2\times 4 + 3\times(-1)}{(-1)^2 + 4^2 + (-1)^2}$

$= \dfrac{4}{18} = \dfrac{2}{9}$

So, $[v]_\beta = \begin{bmatrix} \frac{10}{9} \\ -1 \\ \frac{2}{9} \end{bmatrix}$

Check:

$$\frac{10}{9}\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} + (-1)\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \frac{2}{9}\begin{bmatrix} -1 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{20}{9} - 1 - \frac{2}{9} \\ \frac{10}{9} + \frac{8}{9} \\ \frac{20}{9} - \frac{2}{9} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Some additional Results

Proposition (Pythagorean Theorem):

u and v are orthogonal to each other if and only if $\|u+v\|^2 = \|u\|^2 + \|v\|^2$

Proof:

$$\|u+v\|^2 = \langle u+v, u+v\rangle$$
$$= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle$$
$$= \|u\|^2 + \|v\|^2 + 2\langle u,v\rangle$$

Therefore $\|u+v\|^2 = \|u\|^2 + \|v\|^2$

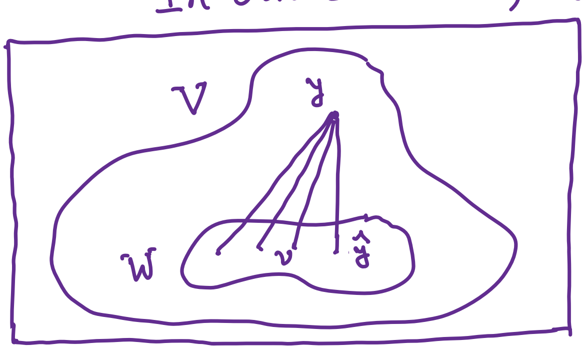
$$\Leftrightarrow \langle u,v\rangle = 0 \Leftrightarrow u \perp v$$

Proposition (Best Approximation Theorem):

Let W be any finite dimensional subspace of V, y any vector in V and $\hat{y}$ be the orthogonal projection of y onto W.

Then $\|y-\hat{y}\| < \|y-v\|$ for all $v \in W$ distinct from $\hat{y}$.

In otherwords, $\hat{y}$ is the closest vector (point) in W to y



Proof: Let $v \in W$ which is distinct from $\hat{y}$

Then $\|y-v\|^2 = \langle y-v, y-v\rangle$

$$= \langle y-\hat{y}+\hat{y}-v, y-\hat{y}+\hat{y}-v\rangle$$

$$= \langle (y-\hat{y})+(\hat{y}-v), (y-\hat{y})+(\hat{y}-v)\rangle$$

$$= \langle y-\hat{y}, y-\hat{y}\rangle + \langle \hat{y}-v, \hat{y}-v\rangle + \langle y-\hat{y}, \hat{y}-v\rangle + \langle \hat{y}-v, y-\hat{y}\rangle$$

$$= \|y-\hat{y}\|^2 + \|\hat{y}-v\|^2 + 2\langle y-\hat{y}, \hat{y}-v\rangle$$

Now $y-\hat{y} \in W^{\perp}$ (Note that $y = \hat{y} + (y-\hat{y})$ and $\hat{y} \in W$)

and $\hat{y}-v \in W$ (Since $\hat{y}, v \in W$ and W is a subspace of V)

So, $\langle \hat{y}-v, y-\hat{y}\rangle = 0$

Therefore $\|y-v\|^2 = \|y-\hat{y}\|^2 + \|\hat{y}-v\|^2$

Now if $y = v$, then $y \in W \Rightarrow y = \hat{y} = v$
which is not allowed as v is distinct from $\hat{y}$.

Hence $\|y-v\|^2 > 0$

and therefore $\|y-v\| > \|y-\hat{y}\|$

i.e. $\boxed{\|y-\hat{y}\| < \|y-v\|}$

Corollary: If y is any vector and W is a finite-dimensional subspace, then

$$\|\text{proj}_W y\| \leq \|y\|$$

Proof: We know that

$$y = \text{Proj}_W y + z \quad \text{where } z \in W^{\perp}$$

By Pythagorean Theorem,

$$\|y\|^2 = \|\text{Proj}_W y + z\|^2 = \|\text{Proj}_W y\|^2 + \|z\|^2$$

Since $\|z\|^2 \geq 0$, we get $\|y\|^2 \geq \|\text{Proj}_W y\|^2$

$$\Rightarrow \boxed{\|\text{Proj}_W y\| \leq \|y\|}$$

Proof of Cauchy-Schwarz Inequality (using the above corollary)

- Want to prove that $|\langle u,v\rangle| \leq \|u\|\|v\| \ \forall u,v \in V$.

Clearly the result holds if either $u=0$ or $v=0$

So, we may assume that both u and v are non-zero and apply the corollary above taking $W = \text{span}\{v\}$

Then $\|\text{Proj}_W u\| \leq \|u\|$

Now $\text{Proj}_W u = \dfrac{\langle u,v\rangle}{\langle v,v\rangle} v$

So, $\left\| \dfrac{\langle u,v\rangle}{\langle v,v\rangle} v \right\| \leq \|u\|$

$$\Rightarrow \frac{|\langle u,v\rangle| \|v\|}{\|v\|^2} \leq \|u\|$$

$$\Rightarrow \frac{|\langle u,v\rangle|}{\|v\|} \leq \|u\|$$

$$\Rightarrow \boxed{|\langle u,v\rangle| \leq \|u\|\,\|v\|}$$

Proof of Orthogonal Decomposition Theorem:

Note: In this proof we assume that any finite-dimensional subspace W of an inner product space has an orthogonal basis.

This assumption is Gram-Schmidt Process which has been covered later.

However the proof of Gram-Schmidt Process does not depend on Orthogonal Decomposition Theorem and so the assumption is logically valid.

- First we prove the uniqueness of the decomposition.

Suppose $y \in V$ and $y = \hat{y} + z$ and $y = \hat{y}_1 + z_1$ where $\hat{y}, \hat{y}_1 \in W$ and $z, z_1 \in W^{\perp}$

Subtracting, $0 = (\hat{y} - \hat{y}_1) + (z - z_1)$

$$\Rightarrow \hat{y}_1 - \hat{y} = z - z_1$$

Now $\hat{y}_1 - \hat{y} \in W$ and $z - z_1 \in W^{\perp}$

So, $\underset{(= z - z_1)}{\hat{y}_1 - \hat{y}} \in W \cap W^{\perp}$

Since $W \cap W^{\perp} = \{0\}$, we get $\left.\begin{array}{l}\hat{y}_1 - \hat{y} = 0 \Rightarrow \hat{y} = \hat{y}_1 \\ \text{and } z - z_1 = 0 \Rightarrow z = z_1\end{array}\right\}$

Thus the decomposition is unique.

- Now we will prove that such a decomposition exists.

i.e. any $y \in V$ can be written as
$y = \hat{y} + z$ where $\hat{y} \in W$ and $z \in W^{\perp}$.

Let $\{u_1, u_2, \dots, u_p\}$ be an orthogonal basis of W.

Let $\left.\begin{array}{l}\hat{y} = c_1 u_1 + \dots + c_p u_p \\ \text{where } c_j = \dfrac{\langle y, u_j \rangle}{\langle u_j, u_j \rangle}\end{array}\right\}$ for $j = 1, 2, \dots, p$.

Now $\hat{y} \in W$.

Let $z = y - \hat{y}$ Then $y = \hat{y} + z$

Now
$$
\begin{aligned}
\langle z, u_j \rangle = \langle y - \hat{y}, u_j \rangle &= \langle y, u_j \rangle - \langle \hat{y}, u_j \rangle \\
&= \langle y, u_j \rangle - \langle c_1 u_1 + \dots + c_p u_p, u_j \rangle \\
&= \langle y, u_j \rangle - c_1 \langle u_1, u_j \rangle - \dots - c_p \langle u_p, u_j \rangle \\
&= \langle y, u_j \rangle - c_j \langle u_j, u_j \rangle \\
&= \langle y, u_j \rangle - \frac{\langle y, u_j \rangle}{\langle u_j, u_j \rangle} \langle u_j, u_j \rangle \\
&= \langle y, u_j \rangle - \langle y, u_j \rangle = 0
\end{aligned}
$$
for $j = 1, 2, \dots, p$.

So, $z \perp u_j$ for $j = 1, 2, \ldots, p$

Since $\{u_1, u_2, \ldots, u_p\}$ is a basis of W,

$z \perp W \Rightarrow z \in W^{\perp}$

Therefore $\boxed{y = \hat{y} + z}$ where $\hat{y} \in W$ and $z \in W^{\perp}$.

$\boxed{\text{QED}}$

MTH 100 : Lecture 37

Diagonalization of Symmetric Matrices

Definition: A matrix A is said to be symmetric if $A = A^T$.

A symmetric matrix is necessarily square.

- For the time being we restrict ourselves to matrices and vectors with real entries.

Proposition: If A is symmetric, then any two eigenvectors from different eigenspaces (i.e. eigenvectors corresponding to different eigenvalues) are orthogonal.

Note: Earlier, we have shown that for any square matrix, eigenvectors from different eigenspaces (i.e. corresponding to different eigenvalues) are linearly independent.

For symmetric matrices, we have the stronger result above.

Proof: Let u_1 and u_2 be eigenvectors corresponding to different eigenvalues λ_1 and λ_2 respectively (for matrix A)

$$\lambda_1 \langle u_1, u_2 \rangle = \langle \lambda_1 u_1, u_2 \rangle$$

$$= \langle A u_1, u_2 \rangle = (A u_1)^T u_2$$

$$= (u_1^T A^T) u_2$$

$$= (u_1^T A) u_2 \quad (\text{since } A \text{ is symmetric})$$

$$= u_1^T (A u_2)$$

$$= u_1^T (\lambda_2 u_2)$$

$$= \langle u_1, \lambda_2 u_2 \rangle$$

$$= \lambda_2 \langle u_1, u_2 \rangle$$

$$\Rightarrow (\lambda_1 - \lambda_2) \langle u_1, u_2 \rangle = 0 \Rightarrow \langle u_1, u_2 \rangle = 0$$

$(\text{since } \lambda_1 - \lambda_2 \neq 0)$

So, $u_1 \perp u_2$

<u>Definition:</u> A square matrix P is said to be $\boxed{\text{orthogonal}}$ if its columns are <u>orthonormal</u>.

(Please note this slight inconsistency in terminology)

Proposition: An orthogonal matrix P is necessarily invertible and $P^{-1} = P^T$

Proof: Let $P = [v_1 \, v_2 \, \dots \, v_n]$ where v_i's are the orthonormal column vectors.

$$P^T P = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} [v_1 \, v_2 \, \dots \, v_n]$$

$$= \begin{bmatrix} v_1^T v_1 & v_1^T v_2 & \dots & v_1^T v_n \\ v_2^T v_1 & v_2^T v_2 & \dots & v_2^T v_n \\ \dots & \dots & \dots & \dots \\ v_n^T v_1 & v_n^T v_2 & \dots & v_n^T v_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

(since $v_i^T v_j = \delta_{ij}$ (kronecker delta)
$\delta_{ij} = 1$ if $i = j$
$= 0$ if $i \neq j$)

Therefore $P^T P = I$

$\boxed{\text{QED}}$

Definition: A square matrix A is said to be orthogonally diagonalizable if there is an orthogonal matrix P and a diagonal matrix D, such that $A = PDP^{-1} = PDP^T$
(ie. $AP = PD$)

Note: For an $n \times n$ matrix to be orthogonally diagonalizable, it should have n linearly independent and orthonormal eigenvectors. That happens only in the following case.

Proposition: If an $n \times n$ matrix A is orthogonally diagonalizable, then A is symmetric.

Proof: If A is orthogonally diagonalizable, then $A = PDP^{-1} = PDP^T$ where P is an orthogonal matrix.

Now $A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$

(Since D is a diagonal matrix)

$\Rightarrow A^T = A$ and so A is symmetric.

Definition: The set of eigenvalues of a matrix A is called the spectrum of A.

Theorem (Spectral Theorem for Symmetric Matrices):

An $n \times n$ symmetric matrix A has the following properties.

(a) The eigenspaces are mutually orthogonal.
(i.e. eigenvectors corresponding to different eigenvalues are orthogonal)

(b) A has n real eigenvalues, counting algebraic multiplicities.

(c) A is orthogonally diagonalizable.

(d) The dimension of the eigenspace for each eigenvalue λ equals the algebraic multiplicity of λ (as a root of the characteristic equation),
i.e. the geometric multiplicity is equal to the algebraic multiplicity.

Remarks:

(1) The proof of (a) is given in a previous proposition.
(2) The proof of (b) is an exercise.
(3) The proof of (c) is nontrivial and will be omitted.
(4) The statement (d) follows from the statement (c) using the Diagonalization Theorem.

Corollary: Taking statement (c) and previous proposition we have:
A is orthogonally diagonalizable if and only if A is symmetric.

The Spectral Theorem in Practice:

- In numerical examples, we first factorize the characteristic polynomial. We will always get as many real roots (counting multiplicities) as the dimension of the matrix, i.e. complex roots will not occur.

- While row reducing the matrix $(A-\lambda I)$ for any eigenvalue λ to solve the associated homogeneous system, we get as many free variables as the algebraic multiplicity of λ. Thus we get the desired number of basis vectors.
- For each eigenspace of dimension greater than one, we obtain an orthogonal basis by using the Gram-Schmidt process.
- Finally we normalize all the basis vectors.

Ex: <u>Diagonalization of a symmetric Matrix</u>

Given $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ we have to diagonalize A.

The characteristic Polynomial $= \det(A-\lambda I)$

$$= \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix}$$

$$= (2-\lambda)\{(2-\lambda)^2-1\} + 1\{1\times 1 - 1(2-\lambda)\} + 1\{1\times 1 - 1(2-\lambda)\}$$

$$= (2-\lambda)(\lambda^2-4\lambda+3) + (\lambda-1) + (\lambda-1)$$

$$= (2-\lambda)(\lambda-1)(\lambda-3) + (\lambda-1)+(\lambda-1)$$

$$= (\lambda-1)\left[(2\lambda-6-\lambda^2+3\lambda) + 1 + 1\right]$$

$$= (\lambda-1)\left(-\lambda^2+5\lambda-4\right) = -(\lambda-1)(\lambda^2-5\lambda+4)$$
$$= -(\lambda-1)(\lambda-4)(\lambda-1) = -(\lambda-1)^2(\lambda-4)$$

So, the eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 1$

Now for $\lambda_1 = 4$:

$$A-\lambda_1 I = \begin{bmatrix} 2-4 & 1 & 1 \\ 1 & 2-4 & 1 \\ 1 & 1 & 2-4 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow[R_3 \to R_3+2R_1]{R_2 \to R_2-R_1} \begin{bmatrix} 1 & 1 & -2 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{R_3 \to R_3+R_2} \begin{bmatrix} 1 & 1 & -2 \\ 0 & -3 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{R_2 \to -\frac{1}{3}R_2} \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 \to R_1-R_2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \text{RREF matrix}$$

So, the system of equation reduces to

$$\left.\begin{aligned} x_1 - x_3 &= 0 \\ x_2 - x_3 &= 0 \end{aligned}\right\} \Rightarrow \left.\begin{aligned} x_1 &= x_3 \\ x_2 &= x_3 \\ x_3 &= x_3 \end{aligned}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

By taking $x_3 = 1$, we get $u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ corresponding to $\lambda_1 = 4$

Normalising, we get $v_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

For $\lambda_2 = 1$: $(A-\lambda_2 I) = \begin{bmatrix} 2-1 & 1 & 1 \\ 1 & 2-1 & 1 \\ 1 & 1 & 2-1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

$$\text{RREF matrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \longleftarrow \begin{array}{l} R_2 \to R_2 - R_1 \\ R_3 \to R_3 - R_1 \end{array}$$

So, the system of equation reduces to

$$\left.\begin{array}{r} x_1 + x_2 + x_3 = 0 \\ x_2 = x_2 \\ x_3 = x_3 \end{array}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Taking $x_2 = 1, x_3 = 0$ and $x_2 = 0, x_3 = 1$ we get

$$u_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad u_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Note that $\langle u_1, u_2 \rangle = 0$, $\langle u_1, u_3 \rangle = 0$

but $\langle u_2, u_3 \rangle = (-1)(-1) = 1 \neq 0$

Therefore we need to apply Gram-Schmidt process to W = eigenspace corresponding to $\lambda_2 = 1$

Let $w_2 = u_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

Then $w_3 = u_3 - \dfrac{\langle u_3, u_2 \rangle}{\langle w_2, w_2 \rangle} w_2$

$$= \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

Note that $\langle w_2, w_3 \rangle = (-1)\left(-\frac{1}{2}\right) + 1\left(-\frac{1}{2}\right) = \frac{1}{2} - \frac{1}{2} = 0$

Now we normalise w_2 and w_3 to get

$$v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix} \quad \text{and} \quad v_3 = \frac{1}{\sqrt{\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2+1^2}}\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix}$$

$$= \frac{2}{\sqrt{6}}\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}-1\\-1\\2\end{bmatrix}$$

So desired P and D will be

$$P = \begin{bmatrix}\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix}4&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

Now we calculate AP and PD:

$$AP = \begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{bmatrix} = \begin{bmatrix}\frac{4}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{4}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{4}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{bmatrix}$$

$$\text{and} \quad PD = \begin{bmatrix}\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{bmatrix}\begin{bmatrix}4&0&0\\0&1&0\\0&0&1\end{bmatrix} = \begin{bmatrix}\frac{4}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{4}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{4}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{bmatrix}$$

Therefore $AP = PD \Rightarrow \boxed{A = PDP^{-1} = PDP^{T}}$

MTH 100 : Lecture 38

Singular Value Decomposition

Let A be an $m \times n$ matrix.

Then $(A^TA)^T = A^T(A^T)^T = A^TA$

Therefore A^TA is a symmetric $n \times n$ matrix.
and can be orthogonally diagonalized.

Let $\{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Then $\|Av_i\|^2 = \langle Av_i, Av_i \rangle$

$$= (Av_i)^T(Av_i) = (v_i^T A^T)(Av_i)$$

$$= v_i^T (A^TA) v_i$$

$$= v_i^T(\lambda_i v_i) = \lambda_i v_i^T v_i$$

$$= \lambda_i \|v_i\|^2$$

$$= \lambda_i \cdot 1 = \lambda_i$$

So, $\|Av_i\|^2 = \lambda_i \Rightarrow \|Av_i\| = \sqrt{\lambda_i}$ for $i = 1, 2, \ldots, n$

Thus $\lambda_i \geq 0$ for $i = 1, 2, \ldots, n$

Therefore all the eigenvalues of the matrix $A^T A$ are nonnegative.

Definition:

Let A be an $m \times n$ matrix.

The singular values of A are the square roots of the eigen values of $A^T A$ denoted by $\sigma_1, \sigma_2, \ldots, \sigma_n$ arranged in descending order

i.e. $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \ldots, n$

$$\sigma_1 \geq \sigma_2 \geq \cdots\cdots\cdots \geq \sigma_n$$

Note that the singular values are the lengths of the vectors

$$Av_1, Av_2, \ldots\ldots, Av_n.$$

Proposition: Suppose $\{v_1, \ldots, v_n\}$ is an **orthonormal basis** for $\mathbb{R}^n$ consisting of eigenvectors of A^TA with corresponding eigenvalues arranged so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

Suppose that A has r nonzero singular values.

Then $\{Av_1, Av_2, \ldots, Av_r\}$ is an orthogonal basis for $\operatorname{col} A$.

and $\operatorname{rank} A = r$

Proof: First note that for $j > r$, $\|Av_j\| = \sqrt{\lambda_j} = \sigma_j = 0$

$\Rightarrow Av_j = 0$

Now for $i, j \leq r$ $(i \neq j)$

we have $\langle Av_i, Av_j \rangle$

$$= (Av_i)^T (Av_j)$$

$$= (v_i^T A^T)(Av_j)$$

$$= v_i^T (A^T A) v_j$$

$$= v_i^T \lambda_j v_j$$

$$= \lambda_j v_i^T v_j$$

$$= \lambda_j \langle v_i, v_j \rangle = 0$$

(since $\{v_1, v_2, \dots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$)

So, $\{Av_1, Av_2, \dots, Av_r\}$ is an orthogonal set of nonzero vectors and are therefore linearly independent.

Next observe that the vectors $Av_1, Av_2, \dots Av_r, \dots Av_n$ belong to Col A (of course $Av_{r+1} = \dots = Av_n = 0$)

Now let $y \in \text{Col } A$

Then $y = Ax$ for some $x \in \mathbb{R}^n$

Since $\{v_1, v_2, \dots, v_n\}$ is a basis for $\mathbb{R}^n$, x can be expressed as

$$x = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$$

Note that if $A = [a_1\ a_2 \dots a_n]$ then

$$Ax = [a_1\ a_2 \dots a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 a_1 + x_2 a_2 + \dots + x_n a_n \in \text{Col} A$$

Then $y = Ax = A(c_1 v_1 + c_2 v_2 + \dots + c_n v_n)$

$$\Rightarrow \quad y = c_1 A v_1 + c_2 A v_2 + \cdots + c_r A v_r + c_{r+1} A v_{r+1} + \cdots + c_n A v_n$$

$$= c_1 A v_1 + c_2 A v_2 + \cdots + c_r A v_r$$

Thus the vectors $Av_1, Av_2, \ldots, Av_r$ span $\operatorname{Col} A$

Therefore the set of vectors $\{Av_1, Av_2, \ldots, Av_r\}$ forms an orthogonal basis for $\operatorname{col} A$

and $\operatorname{Rank} A = \dim(\operatorname{col} A) = r$

Singular Value Decomposition (SVD)

Theorem (Singular Value Decomposition of a matrix):

Let A be an $m \times n$ matrix with rank r.
Then A can be factored as a product $A = U \Sigma V^T$ as follows:

- Σ is an $m \times n$ matrix containing an $r \times r$ diagonal matrix D with the r non-zero singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, along the main diagonal. D is placed in the upper left corner of Σ. Remaining entries of Σ are zero.

- U is an $m \times m$ orthogonal matrix and V is an $n \times n$ orthogonal matrix.

- The matrix V has as its columns the orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of eigenvectors of $A^T A$.

- In order to obtain U, we take r vectors Av_i corresponding to the non-zero singular values, extend to an orthogonal basis of $\mathbb{R}^m$ using the Gram-Schmidt Process (This step is necessary only in case $r < m$) and finally normalize the vectors to obtain an orthonormal basis $\{u_1, u_2, \dots, u_m\}$ U has the vectors u_i as its columns.

Note: Any factorization $A = U \Sigma V^T$, with U and V as orthogonal matrices, Σ as described above is called a Singular Value Decomposition or SVD of A.
Note that U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A.

$$\boxed{\text{MTH 100 : Lecture 39}}$$

Example for SVD

Let $A = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}_{3\times 2}$

Then $A^T A = \begin{bmatrix} -3 & 6 & 6 \\ 1 & -2 & -2 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}$

$$= \begin{bmatrix} 81 & -27 \\ -27 & 9 \end{bmatrix}_{2\times 2} = B \text{ (say)}$$

(symmetric 2x2 matrix)

Characteristic Polynomial of B

$$= \det[B - \lambda I] = \begin{vmatrix} 81-\lambda & -27 \\ -27 & 9-\lambda \end{vmatrix}$$

$$= (81-\lambda)(9-\lambda) - (-27)(-27)$$

$$= \cancel{729} - 9\lambda - 81\lambda + \lambda^2 - \cancel{729}$$

$$= \lambda^2 - 90\lambda = \lambda(\lambda - 90)$$

Eigen values in descending order are

$\lambda_1 = 90$, $\lambda_2 = 0$

For $\lambda_1 = 90$

$$B - \lambda_1 I = \begin{bmatrix} 81-90 & -27 \\ -27 & 9-90 \end{bmatrix} = \begin{bmatrix} -9 & -27 \\ -27 & -81 \end{bmatrix}$$

$$\Big\downarrow R_2 \to R_2 - 3R_1$$

$$\text{RREF matrix} = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \xleftarrow{R_1 \to -\frac{1}{9}R_1} \begin{bmatrix} -9 & -27 \\ 0 & 0 \end{bmatrix}$$

So, the system of equation becomes:

$$\left.\begin{aligned} x_1 + 3x_2 &= 0 \\ x_2 &= x_2 \end{aligned}\right\} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Taking $x_2 = -1$ and normalising we get an eigenvector $v_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{bmatrix}$

For $\lambda_2 = 0$

$$B - \lambda_2 I = \begin{bmatrix} 81 & -27 \\ -27 & 9 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} -27 & 9 \\ 81 & -27 \end{bmatrix}$$

$$\Big\downarrow R_2 \to R_2 + 3R_1$$

$$\text{RREF matrix} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 0 \end{bmatrix} \xleftarrow{R_1 \to -\frac{1}{27}R_1} \begin{bmatrix} -27 & 9 \\ 0 & 0 \end{bmatrix}$$

So, the system of equations becomes $\left.\begin{array}{l} x_1 - \frac{1}{3}x_2 = 0 \\ x_2 = x_2 \end{array}\right\}$

$$\Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} \frac{1}{3} \\ 1 \end{bmatrix}$$

Taking $x_2 = 3$ and normalising we get an eigenvector $v_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

$$= \begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix}$$

Note that $\langle v_1, v_2 \rangle = 0$ (They are eigenvectors of distinct values.)

So, $V = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}_{2 \times 2}$

Note that $\sigma_1 = \sqrt{90} = 3\sqrt{10}$

$\sigma_2 = 0$

Now we will to compute U:

$$Av_1 = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} -\frac{10}{\sqrt{10}} \\ \frac{20}{\sqrt{10}} \\ \frac{20}{\sqrt{10}} \end{bmatrix}$$

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{3\sqrt{10}} \begin{bmatrix} -\frac{10}{\sqrt{10}} \\ \frac{20}{\sqrt{10}} \\ \frac{20}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

Note that $\|u_1\| = 1$

Now $Av_2 = 0.v_2 = 0$

So, we need to extend u_1 to an orthonormal basis of $\mathbb{R}^3$ by solving the system $\langle u_1, x\rangle = u_1 . x = 0$

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ be such that $\begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} . x = 0$

$$\Rightarrow \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$\Rightarrow -\frac{1}{3} x_1 + \frac{2}{3} x_2 + \frac{2}{3} x_3 = 0$$

$$\Rightarrow -x_1 + 2x_2 + 2x_3 = 0$$

$$\Rightarrow \left.\begin{aligned} x_1 &= 2x_2 + 2x_3 \\ x_2 &= x_2 \\ x_3 &= x_3 \end{aligned}\right\}$$

$$\Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

$$\left\{\begin{aligned} &\text{Taking } x_2 = 1, x_3 = 0 \text{ we get } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \\ &\text{Taking } x_2 = 0, x_3 = 1 \text{ we get } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \end{aligned}\right\}$$

Now $\left\langle \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\rangle = 2 \times 2 + 0 + 0 = 4 \neq 0$

Thus any two solution may not be orthogonal to each other.
If necessary we will have to use Gram-Schmidt orthonormalisation process.

In this problem, we will find the orthonormal vectors by inspection.

Taking $x_2 = \frac{2}{3}$ and $x_3 = -\frac{1}{3}$, we get $u_2 = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{bmatrix}$

Note that $\|u_2\| = 1$

Taking $x_2 = -\frac{1}{3}$ and $x_3 = \frac{2}{3}$, we get $u_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$

Note that $\|u_3\| = 1$

So, $$U = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}_{3\times 3}$$

$$\Sigma = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{3\times 2}$$

With this U, V and Σ we have

$$A = U\Sigma V^T = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}_{3\times 3} \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{3\times 2} \begin{bmatrix} \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}_{2\times 2}$$

<u>Check:</u> $A = U\Sigma V^T$ or equivalently <u>$AV = U\Sigma$</u>:

Now, $$AV = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}_{3\times 2} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}_{2\times 2} = \begin{bmatrix} -\frac{10}{\sqrt{10}} & 0 \\ \frac{20}{\sqrt{10}} & 0 \\ \frac{20}{\sqrt{10}} & 0 \end{bmatrix}_{3\times 2}$$

$$= \begin{bmatrix} -\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \end{bmatrix}_{3\times 2}$$

$$U\Sigma = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}_{3\times 3} \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{3\times 2}$$

$$= \begin{bmatrix} -\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \\ 2\sqrt{10} & 0 \end{bmatrix}_{3\times 2}$$

So, $AV = U\Sigma$

Proof of Singular Value Decomposition (SVD)

Theorem:

Suppose that $\{v_1, v_2, \dots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA with corresponding eigenvalues arranged so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. Let $\sigma_i = \sqrt{\lambda_i}$ for $i=1,2,\dots,n$

Suppose that A has r nonzero singular values so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$

Then $\{Av_1, Av_2, \dots, Av_r\}$ is an orthogonal basis for col A, (Thus rank $= r$) by a previous proposition.

Normalize each Av_i to obtain an orthonormal basis $\{u_1, \dots, u_n\}$ for col A by putting

$$u_i = \frac{Av_i}{\|Av_i\|} = \frac{1}{\sigma_i} Av_i$$

i.e. $\quad Au_i = \sigma_i u_i \quad$ (for $i = 1, 2, \dots, r$) $\quad \dots\dots (1)$

If $r < m$, extend $\{u_1, \dots, u_r\}$ to an orthonormal basis of $\mathbb{R}^m$ (Here we may use Gram-Schmidt Process)

Now let $U = [u_1 \; u_2 \dots u_n]$ and $V = [v_1 \; v_2 \dots v_n]$

U and V are orthogonal matrices by construction.

Now $AV = A[v_1 \; v_2 \dots v_n] = [Av_1 \; Av_2 \dots\dots Av_n]$

$$= [\sigma_1 u_1 \;\; \sigma_2 u_2 \dots\dots \sigma_r u_r \;\; 0 \dots\dots 0]$$

[By (1)]

Now let Σ be the $m \times n$ matrix containing an $r \times r$ diagonal matrix D with the r non-zero singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and we make D into an $m \times n$ matrix Σ (same size as A) by filling out with zeros

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \\ \rightarrow m-r \text{ rows} \end{array} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$\downarrow$ $n-r$ columns

Then $U\Sigma = [u_1 \ u_2 \ \cdots \ u_m] \begin{bmatrix} \sigma_1 & 0 \cdots 0 & \cdots & 0 \\ 0 & \sigma_2 & & \\ 0 & \cdots & \sigma_r & \cdots 0 \\ \vdots & & \vdots & \\ 0 & \cdots & 0 & \cdots 0 \end{bmatrix}$

$$= [\sigma_1 u_1 \ \sigma_2 u_2 \ \cdots \ \sigma_r u_r \ 0 \cdots 0] = AV$$

$\Rightarrow AV = U\Sigma$

Since V is orthogonal, $\boxed{A = U\Sigma V^T}$

$\boxed{\text{QED}}$