Linear Algebra and Matrix theory

Sai Krishna

1 Basic Matrix Theory

1.1 Basic operation performed on Matrices

Let A $\in \mathbb{R}^{m\times n}$, where $\mathbb{R}^{m\times n}$ is the space of all rectangular matrices with m rows and n columns.
Total number of elements in A is mn.

1.1.1 Matrix Addition

To add two or more matrices they must belong to the same space. i.e. they must have same number of rows and columns.
Let A,B $\in \mathbb{R}^{m\times n}$, then C = A+B is defined as

$$C_{ij} = A_{ij} + B_{ij} \qquad 1 \leq i \leq m, 1 \leq j \leq n$$

A_{ij} refers to the element in the i^{th} row at the j^{th} column. Let A,B,C $\in \mathbb{R}^{m\times n}$

- It is commutative A + B = B + A
- It is associative A + (B + C) = (A +B) + C

1.1.2 Scalar multiplication

Let A $\in \mathbb{R}^{m\times n}$ and k $\in \mathbb{R}$,

$$B = kA \implies B_{ij} = kA_{ij}$$

Let A $\in \mathbb{R}^{m\times n}$ and p,q $\in \mathbb{R}$

- pA = Ap
- (pA)q = (pq)A

1.1.3 Matrix Multiplication

To multiply two matrices A,B. The number of columns in A must be equal to number of rows in B.

$$A \in \mathbb{R}^{m \times n}$$
$$B \in \mathbb{R}^{p \times q}$$
$$C = AB \in \mathbb{R}^{m \times q}$$
$$\text{It requires n = p}$$

Multiplication of the above two matrices involves
m(n-1)q Additions and **mnq** Multiplications.

- A(BC) = (AB)C
- A(B + C) = AB + AC
- Matrix multiplication is not always commutative. However there are exceptions.
  Ex: AxA = A^2
- Row operations correspond to pre multiplication
- Column operation corresponds to post multiplication

1.2 Transpose of a Matrix

The transpose operation swaps the rows with the columns of the matrix.
$(A^T)_{ij} = A_{ji}$

If $A^T = A$, then A is **"Symmetric Matrix"**
If $A^T = -A$, then A is **"Skew Symmetric Matrix"**
If $A^T = A^{-1}$, then A is **"Orthogonal Matrix"**

- $A^T A$ is always symmetric
- $(A + B)^T = A^T + B^T$
- $(AB)^T = B^T A^T$
- $(A^{-1})^T = (A^T)^{-1}$

1.3 Determinants

The determinant is a number associated with a square matrix.

- Determinant of the identity matrix is 1. $|I| = 1$
- Determinant of permutation matrices is either +1 or -1 depending on the number of exchanges i.e one negative sign per exchange.

  - Even number of row/column exchanges, determinant remains same.
  - Odd number of row/column exchanges, determinant is multiplied by -1.

- If any single row/column is multiplied by a constant, the determinant also gets multiplied by the constant.
- If any single row/column is added with some numbers, the determinant will be equal to the sum of original determinant and determinant of matrix where the single row/column is replaced by the new numbers added to them.
- Row/column operations (i.e adding constant times one from another) will not change the determinant.
- If a row/column of zeros exists in the matrix, then it's determinant is equal to 0.
- If the same row/column is repeated in the matrix, then it's determinant is equal to 0.
- If the rows/columns of a matrix are linearly dependent, then it's determinant is equal to 0.
- The determinant of triangular matrices and diagonal matrices is equal to the product of principal diagonal elements (i.e the product of pivot elements).

  - If the determinant is zero, it means the matrix is singular i.e non-invertible.
  - If the determinant is not zero, it means the matrix is non-singular i.e invertible.

- Determinant of AB is equal to product of determinant of A and determinant of B. $|AB| = |A||B|$
- Determinant of inverse of A is equal to reciprocal of determinant of A. $|A^{-1}| = 1/A$
- Determinant of a scalar (k) times A (where A is nxn matrix) is equal to k^n times determinant of A. $|kA| = k^n|A|$
- Determinant of transpose of A is same as determinant of A. $|A^T| = |A|$
- Determinant of a skew symmetric matrix of odd order is equal to 0.

1.3.1 Formula for determinants

2x2 matrices : A $= \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
$|A| = ad - bc$
3x3 matrices : A $= \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

$|A| = a(ei - fh) - b(di - fg) + c(dh - eg) = aei + bfg + cdh - afh - bdi - cdh$

1.3.2 Geometric Picture

- In 2x2 case, if the rows of matrix A represent the 2 vectors that form edges of a parallelogram, then the absolute of value of determinant of A will give it's area.
  This can also be used to find area of triangle since area of triangle is half the area of parallelogram.

- In 3x3 case, if the rows of matrix A represent the 3 vectors that form edges of a parallelopiped, then the absolute of value of determinant of A will give it's volume.

1.4 Trace of a Matrix

The sum of diagonal elements of a matrix is called the trace of a matrix.
Let A $\in \mathbb{R}^{n \times n}$

$$Trace(A) = \sum_{j=1}^{n} A_{jj}$$

1.5 Minors, Co-Factors and Adjoint

Each element in a square matrix has it's own **minor** and **co-factor**.
The minor of an element is equal to the determinant of the matrix obtained by excluding the row and column of the element from the original matrix.

$M_{ij} = |A\ (with\ i^{th}\ row\ and\ j^{th}\ column\ erased)|$

The co-factor of an element is equal to $(-1)^{i+j}$ multiplied by the minor of that element.

- Co-factor is equal to minor is i+j is even

- Co-factor is equal to negative of minor if i+j is odd

$C_{ij} = (-1)^{i+j}\ M_{ij}$

Adjoint of a matrix is the transpose of co-factor matrix.
To find adjoint, replace each element of A with it's co-factor and take it's transpose.

$adj(A) = C^T$

- $|adj(A)| = |A|^{n-1}$
- $|adj(adj(A))| = |A|^{(n-1)^2}$
- adj(AB) = adj(B)adj(A)
- A.adj(A) = adj(A)A = $|A|$I

1.6 Rank of a Matrix

Different definitions of rank of a matrix are,

- The order of the biggest sub matrix with non-zero determinant
- Number of pivots in Echelon form
- Number of linearly independent vectors in Range of A

.

1.7 Inverse of a Matrix

A matrix that nullifies the effect of a given matrix is called the inverse matrix.

- $\mathrm{AA}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I}$
- Inverse exists only for full ranked matrices
- Inverse exists if $|A| \neq 0$
- $(\mathrm{AB})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$
- $A^{-1} = \frac{Adj(A)}{|A|}$

1.7.1 Left and Right Inverse

If A is not full ranked then it does not have 2 sided inverse, but if A is either full row-ranked or full column-ranked, then it has one sided inverse. Let A $\in \mathbb{R}^{m\times n}$

A is full column rank matrix; $r = n < m$
The null space is only 0, meaning there is either 1 solution or 0 solutions. A^TA is full rank and invertible.
$\implies$ **Left inverse** : $(A^TA)^{-1}A^T$ because $(A^TA)^{-1}A^TA = I$

A is full row rank matrix; $r = m < n$
The left null space is only 0, meaning there are infinite solutions. AA^T is full rank and invertible.
$\implies$ **Right inverse** : $A^T(AA^T)^{-1}$ because $AA^T(AA^T)^{-1} = I$

2 System of Linear Equations

Can be represented using Matrices and vectors.

$$\mathbf{A}\overrightarrow{x} = b$$

A is the co efficient matrix. $\mathbf{A} \in \mathbb{R}^{m \times n}$
$\overrightarrow{x}$ is the vector of unknowns. $x \in \mathbb{R}^n$
b is the vector of constants in the equations. $b \in \mathbb{R}^m$
The equations are coupled hence solving them directly is not possible, uncoupling the equations will make the task of finding $\overrightarrow{x}$ easy. Uncoupling the equations will transform **A** to a diagonal matrix.

Solution to the system of equations - Geometric Picture

- In an 'n' dimensional space, a plane will be of 'n-1' dimensions.
- A 2nd plane intersecting this plane will give a plane of 'n-2' dimensions and so on.
  Every new plane will reduce the dimension by 1.
- At the end, when 'n' planes are accounted for, the final intersection has zero dimensions, which is the solution.
  This will be the point that satisfies all 'n' equations, i.e the point that lies on all 'n' planes.

The possible solutions to the above system of equations are

1. **Unique solution**
   Planes intersect at one point

2. **Infinite solutions**
   Planes are co incident

3. **No solution**
   Planes are parallel

   b - **A**x $\neq \overrightarrow{0} \ \ \forall \overrightarrow{x} \in \mathbb{R}^n$
   The best possible solution is the one that minimizes the error,

$$\sum_{i=1}^{m} (b_i - (Ax)_i)^2$$

   This is the least square solution. Again we can have one/multiple least square solutions.

NOTE: All matrices cannot be diagonalized.
Ex:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

2.1 Inner product and tensor product

Let $\overrightarrow{x}, \overrightarrow{y} \in \mathbb{R}^n$
$(x, y) = y^T x = x^T y \in \mathbb{R}^1$
Is the inner product of x and y

$x \otimes y = xy^T \in \mathbb{R}^{n \times n}$
Is the tensor product of x and y
$x \otimes y \neq y \otimes x$
The resulting matrix has rank = 1

2.1.1 Norm of a vector

Norm is a function that assigns a strictly positive number to a vector, and has the below mentioned properties.
$\forall\ x \in \mathbb{R}^k$, the length of the vector or NORM is defined as,
$||x|| = \sqrt{(x, x)}$
Is also called **Euclidean Norm**

Properties

- $||x|| \geq 0$, equal to 0 $\iff x = \overrightarrow{0}$
- $\alpha||x|| = |\alpha|\ ||x||$
- **Cauchy-Schwarz Inequality**
  $||x + y|| \leq ||x||\ +\ ||y||$
- $||x + y|| = \sqrt{||x||^2 + ||y||^2 + 2(x.y)}$

2.1.2 Norm of a matrix

A matrix norm is a norm on the vector space $\mathbb{R}^{m \times n}$, and has the above mentioned properties.

2.2 Homogeneous System of Equations

$\mathbf{A}\overrightarrow{x} = \overrightarrow{0}$
Always has a solution, $\overrightarrow{x} = \overrightarrow{0}$ which is the trivial solution.
The homogeneous system may also have non trivial solutions. If x_H is a solution to the homogeneous system then kx_H is also a solution where k is a scalar.
Hence a homogeneous system can have only trivial solution or infinitely many solutions.

NOTE: If homogeneous system has non trivial solutions then the corresponding non homogeneous system will have infinitely many solutions.
Let x_H be a non trivial solution to homogeneous system. $\mathbf{A}\overrightarrow{x_H} = \overrightarrow{0}$
If the non homogeneous system is consistent, i.e. it has a solution u. $\mathbf{A}\overrightarrow{u} = \overrightarrow{b}$
then, $\mathbf{A}(\overrightarrow{u} + k\overrightarrow{x_H}) = \overrightarrow{b}$. Hence the corresponding non homogeneous system has infinitely many solutions. The converse is also true.
If the non homogeneous system has infinitely many solutions then the corresponding homogeneous system has a non trivial solution.

2.3 Elementary row operations

1. Row exchange
2. Add scaled version of one row to another row
3. Scale a row by a non zero value

The above operations are,
Invertible, the solution of corresponding homogeneous system is unaltered after this operation
The operations can be represented using matrices (Elementary matrices), and these have to be pre-multiplied to the matrix on which the operations have to be performed.

A $\xrightarrow{E_1} A_1 \xrightarrow{E_2} A_2 ... \xrightarrow{E_n} B$

where $E_1, E_2, ..$ are all elementary matrices. Then A and B are said to be Row equivalent.Row equivalence has Reflexitivity, Symmetry and transitivity property hence Row equivalence is an **Equivalence relation** on the set of all matrices.
Let E = $E_n ... E_2 E_1$, then $\mathbf{EA = B}$ and hence $\mathbf{A} = E^{-1}\mathbf{B}$

2.4 Gaussian Elimination

Since Elementary row operation do not change the solution, we can perform elementary operations on A, to convert it to a form where it is convenient to solve for the unknowns. This form is called **Echelon form**.

- Pivots should be 1.
  Pivots are the first non zero number in a row.
- All elements below and above the pivot must be zero.
- If pivot of i^{th} column is at k_i, pivot of j^{th} column is at k_j and $i < j \implies k_i < k_j$.
- All the zero rows should be at the bottom.

Express the pivotal elements interms of free variables.
Free variables are the columns corresponding to which there aren't any pivots.
Row Rank of a matrix is the number of nonzero rows in echelon form of that matrix, which is equal to number of pivotal variables.
Nullity of the matrix is equal to number of free variables.
Rank Nullity Theorem
Row rank + Nullity is equal to the number of variables/columns in the matrix.

Extending Gaussian Elimination to non homogeneous system
Repeat the above for the augmented matrix [A:b]. This is because elementary operations affect the solution of non homogeneous systems. The solution is, assign the RHS to pivotal variables, and zero to free variables.

2.5 Gauss - Jordan Method to obtain inverse of matrix

- Obtain Augmented matrix [A:I]
- Perform Gaussian Elimination on A to convert it to I.
- Perform the same operations on I. It will get converted to A^{-1}.

Performing Gaussian Elimination can be represented by Matrix multiplication.

$$QA = I$$

Assume A^{-1} exists and post multiply the above equation by A^{-1}

$$\begin{aligned} QAA^{-1} = IA^{-1} &= A^{-1} \\ QI &= A^{-1} \end{aligned}$$

Hence performing Gaussian elimmination operations of A on I, gives A^{-1}.

3 Vector Spaces

Let $\mathbb{V}$ be a non empty set in $\mathbb{R}^k$, and + be a rule of combining elements of this set. If the following properties are satisfied,

- Closure
- Commutativity: x + y = y + x
- Associativity: (x + y) + z = x + (y + z)
- Additive inverse: x + (-x) = θ_k
- Additive identity: x + θ_k = x

Then ($\mathbb{V}$,+) is an Abelian group.
Consider another rule of combination Scalar Multiplication.

- $\alpha, \beta \in \mathbb{R},\ x, y \in \mathbb{R}^k$, then $\alpha x \in \mathbb{R}^k$
- $(\alpha + \beta)x = \alpha x + \beta x$
- $\alpha\beta(x) = \beta\alpha(x)$
- $\alpha(x + y) = \alpha x + \alpha y$

If the elements of an non empty set in $\mathbb{R}^k$ are closed under addition and scalar multiplication and both, and if all the above properties hold then the set is called a **Vector Space**.

Consider a non empty infinite set of points $\mathbb{S} = \{s_1, s_2, ...\}$ (S can be $\mathbb{R}$ or $\mathbb{C}$)and a class of functions mapping from $\mathbb{S}$ to $\mathbb{F}$ where $\mathbb{F}$ is the set of all Real or Complex numbers, denoted by $F[S; \mathbb{F}]$. Then,

$$l^1[S; \mathbb{F}] = \{f \in F[S; \mathbb{F}] : \sum_{j=1}^{\infty} |f(s_j)| < \infty\}$$

It is the space of all functions whose sum of absolute values is bounded.
Ex: $f(s_j) = \frac{1}{j^2}$

$$l^2[S; \mathbb{F}] = \{f \in F[S; \mathbb{F}] : \sum_{j=1}^{\infty} |f(s_j)|^2 < \infty\}$$

It is the space of all functions whose sum of squared absolute values is bounded Ex: $f(s_j) = \frac{1}{j}$
Both l^1 and l^2 are vector space.

Let $\mathbb{I}$ be an interval on the real number line. Then $F[\mathbb{I};\mathbb{R}]$ is a vector space over $\mathbb{R}$ and $F[\mathbb{I};\mathbb{C}]$ is a vector space over $\mathbb{C}$ and S is non empty set in $\mathbb{I}$

$$L^1[S;\mathbb{R}] = \{f \in F[S;\mathbb{R}] : \int_{\mathbb{I}} |f(t)|dt < \infty\}$$

$$L^2[S;\mathbb{R}] = \{f \in F[S;\mathbb{R}] : \int_{\mathbb{I}} |f(t)|^2 dt < \infty\}$$

Are some of the standard vector spaces of functions over $\mathbb{R}$.
Similarly we can have a vector space of continuous functions over $\mathbb{R}$ and $\mathbb{C}$, vector space of polynomials over $\mathbb{R}$ and $\mathbb{C}$.

3.1 Linear Combination

Let $u_1, u_2, ..., u_k$ be a set of vectors, then $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k$ is the non trivial linear combination of the vectors even if one of the coefficients is not equal to 0.

- If any non trivial combination gives the zero vector, then the above vectors are said to be dependent.
  - Every vector which is obtained as a combination of linearly dependent vectors will atleast have 2 different representations.
    $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k = x$
    $\beta_1 u_1 + \beta_2 u_2 + ... + \beta_k u_k = 0$
    $(\alpha_1 + \beta_1)u_1 + (\alpha_2 + \beta_2)u_2 + ... + (\alpha_k + \beta_k)u_k = x$
- If only trivial combinatin leads to zero vector then the set of vectors are said to be linearly independent.

3.2 Subspace

Consider a vector space $\mathbb{V}$ over a field $\mathbb{F}$. Consider a finite set of vectors in $\mathbb{V}$. $S = \{u_1, u_2, ..., u_k\}$. Let $\mathbb{L}[S]$ denote the collection of all linear combinations of S vectors.
$\mathbb{L}[S] = \{x \in \mathbb{V} : x = \alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k; \alpha_j \in \mathbb{F}; 1 \leq j \leq k\}$
NOTE:

- $\theta_k \in \mathbb{L}[S]$ (trivial combination)
- $S \subset \mathbb{L}[S]$ (Only one non zero co efficient corresponding to that vector). Hence $\mathbb{L}[S]$ **is non empty**
- $\forall\ x\ \in \mathbb{L}[S]$ x also $\in \mathbb{V}$ since all the vectors of $S \in \mathbb{V}$
  $\implies$ the space $\mathbb{L}[S] \subset \mathbb{V}$
  If we take two vectors from $\mathbb{L}[S]$ and combine them using vector addition and/or scalar multiplication the resulting vector lies in $\mathbb{L}[S]$. Since linear combination of two linearly combined vectors is again a linear combination. Hence $\mathbb{L}[S]$ **is closed under vector addition and scalar multiplication**.

Hence $\mathbb{L}[S]$ is a **Subspace** of $\mathbb{V}$ spanned by vectors of S.

- θ_k is the smallest possible subspace of $\mathbb{V}$
- $\mathbb{V}$ is the largest possible subspace of $\mathbb{V}$
- An infinite set S in $\mathbb{V}$ is said to be linearly independent if every finite subset of S is linearly independent.
  Ex: $S = \{sin(nx)\}$ where n $\in \mathbb{Z}$
- Any non empty subset of linearly independent set is linearly independent. Any super set of linearly dependent set is linearly dependent.
- If S is a finite l.d. set then there exists a subset S_1 of S such that it is linearly independent and $\mathbb{L}[S] = \mathbb{L}[S_1]$, i.e. they span the same space.

3.3 Subspace associated with Matrix

Consider a matrix $A^{m \times n}$. $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$

- $W = \{x \in \mathbb{F}^n : Ax = \theta_m\}$
  A$\theta_n = \theta_m$
  Hence W is non empty subset of $\mathbb{F}^n$.

  Let x,y $\in$ W, $\implies Ax = \theta_m, Ay = \theta_m \implies A(x+y) = 0$
  Hence W is closed under vector addition.

  x $\in$ W, $\implies Ax = \theta_m \implies \alpha Ax = A(\alpha x) = \theta_m$
  Hence W is closed under scalar multiplication.

  This subspace is called the **Null space** of matrix A. Set of all solutions to homogenous equations involving A.

- $W = \{b \in \mathbb{F}^m : \text{there exists } x \ni Ax = b\}$
  If b = θ_m, then we know A$\theta_n = \theta_m$
  Hence W is non empty subset of $\mathbb{F}^m$

  If $b_1, b_2 \in W \implies Ax_1 = b_1$ and $Ax_2 = b_2$ then
  $Ax_1 + Ax_2 = b_1 + b_2$
  $A(x_1 + x_2) = b_1 + b_2$
  Hence W is closed under vector addition.

  $b_1 \in W \implies Ax_1 = b_1$
  $\alpha Ax_1 = \alpha b_1$
  $A(\alpha x_1) = \alpha b_1$
  Hence W is closed under scalar multiplication.

  This subspace is called the **Range/Columnn Space** of A. Set of all b for which **A**x = b is consistent.

- $W = \{x \in \mathbb{F}^m : A^T x = \theta_n\}$
  W is the Null space of A^T in $\mathbb{F}^m$
  This subspace is called **Left Null Space** of A.

- $W = \{b \in \mathbb{F}^n : \text{there exists } x \ni A^T x = b\}$
  W is the Column space of A^T in $\mathbb{F}^n$
  This subspace is called **Row space** of A.

$A^{m \times n}$ has
Row space and Null space in $\mathbb{F}^n$
Column space and Left null space in $\mathbb{F}^m$
Are the 4 important sub spaces associated with a matrix.

3.4 Basis

Let $\mathbb{W}$ be a subspace of $\mathbb{V}$. Then $\mathbb{B} \subset \mathbb{W}$ is called a basis for $\mathbb{W}$

- If $\mathbb{B}$ is linearly independent
- If $\mathbb{L}[\mathbb{B}] = \mathbb{W}$ i.e. every vector in $\mathbb{W}$ can be expressed as a linear combination of vectors in $\mathbb{B}$
- Iff $\mathbb{B}$ is Maximal Linearly independent set, i.e. any set bigger than the basis must be linearly dependent.

Hence $\mathbb{B}$ is a basis for the subspace $\mathbb{W}$.
It is this interpretation for basis together with **Zorn's lemma** that assures every vector space has a basis.
Basis is not unique

- Consider a set of k linearly independent vectors $S = u_1, u_2, ..., u_k$ in $\mathbb{R}^n$ where $n > k$. By appending (n-k) linearly independent vectors from $\mathbb{R}^n - \mathbb{L}[S]$. S can be extended to be a basis.

A set of linearly independent vectors can be extended to be a basis.

Ordered basis: A basis in which the vectors are arranged in a fixed order. Given an ordered basis the representation of a vector is unique. Hence any vector x in $\mathbb{V}$[dim = n] can be converted to a vector x_b in $\mathbb{F}^n$ where b is the ordered basis. Every component of x_b is the weight given to the corresponding basis vector in b.

3.5 Dimension of a Space

- If $\mathbb{V}$ has a basis $\mathbb{B}$ having n elements, then any set with n+1 elements is a linearly dependent set.
- All the basis for a given space will always have the same number of elements. .
- In a n-Dimensional space($\mathbb{V}$), any set of n vectors($\mathbb{B}_1$) may not form a basis because $\mathbb{L}[B_1] \neq \mathbb{V}$,
  but any set of n linearly independent vectors will form a basis.

. The number of vectors in the basis determines the **Dimension** of the space. Space of all polynomials is an infinite dimensional vector space, and the corresponding basis is, $1, x, x^2, x^3, ...$

4 Linear Transformation

Based on the ordered basis used the representation for a vector in $\mathbb{F}^n$ changes, but since they represent the same vector in $\mathbb{V}$ they should be related. And this is obtained by a **Linear Transformation** of the basis vectors.

Consider two set of basis vectors B and B'.
$B = u_1, u_2, ...u_n \quad x \xrightarrow{\text{B}} x_B$
$B' = v_1, v_2, ...v_n \quad x \xrightarrow{\text{B'}} x_{B'}$
Represent $u_1, u_2, ...u_n$ in terms of B', i.e. $[u_1]_{B'}, [u_2]_{B'}, ...[u_n]_{B'}$
Construct a matrix using these as columns.

$$\begin{aligned}
\big[B\big]_{B'} &= \big[[u_1]_{B'}\ [u_2]_{B'}\ ...\ [u_n]_{B'}\big] \\
[u_1]_{B'} &= b_{11}v_1 +\ b_{21}v_2 +\ ...\ + b_{n1}v_n \\
[u_j]_{B'} &= b_{1j}v_1 +\ b_{2j}v_2 +\ ...\ + b_{nj}v_n \\
x &= \sum_{j=1}^{n} x_j u_j \\
x &= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} b_{ij}v_i = \sum_{i=1}^{n}\sum_{j=1}^{n}(b_{ij}x_j)v_i \\
x &= \sum_{i=1}^{n} \alpha_i v_i \\
x_{B'} &= \big[B\big]_{B'} x_B
\end{aligned}$$

$[\alpha_1\ \alpha_2\ ..\ \alpha_n]^T = x_{B'}$
To get $x_{B'}$ from x_B we first need the representation of basis vectors in B in terms of B'.

Properties

- $x \xrightarrow{\text{B}} x_B$
  $y \xrightarrow{\text{B}} y_B$
  $x + y \xrightarrow{\text{B}} x_B + y_B$

- $\alpha x \xrightarrow{\text{B}} (\alpha x)_B = \alpha x_B$

The identification of x in $\mathbb{V}$ with x_B in $\mathbb{F}^n$ preserves vector addition and scalar multiplication. Hence the transformation is linear.
Thus every ordered basis B in an n dimensional vector space $\mathbb{V}$ over $\mathbb{F}$ produces a linear transformation T_B from $\mathbb{V}$ to $\mathbb{F}^n$

- Any m $\times$ n matrix A $\in \mathbb{F}^{m\times n}$ produces a linear transformation
  $T_A : \mathbb{F}^n \longrightarrow \mathbb{F}^m$ as $\quad T_A(x) = Ax$

- A linear transformation from $\mathbb{V}$ to $\mathbb{V}$ is called a **Linear operator**.
  $\therefore$ Every Square matrix is a linear operator
- T: $\mathbb{V} \longrightarrow \mathbb{W}$
  T$(\theta_V) = \theta_W$
  A linear transformation always maps a zero vector to another zero vector.
- $N_T = \{x \in \mathbb{V} \ : Tx = \theta_W\}$
  N_T is the set of all vectors that are mapped to the zero vector and is a subspace of $\mathbb{V}$ it is called the **Null space** of T. Dimension of Null space is the **Nullity** of T.
- $R_T = \{y \in \mathbb{W} \ : \ni \ x \in \mathbb{V} \ni Tx = y\}$
  R_T is a subspace of $\mathbb{W}$ and is called the **Range** of the transformation. The dimension of R_T is called **Rank** of T.
- $\mathbb{V} = \mathbb{F}[x]$ space of polynomials,[p is the basis]

  Let $D(p) = \frac{dp}{dx} \ \in \ \mathbb{F}[x] \ \ \forall p \ \in \ \mathbb{F}[x]$
  Hence differentiation is a linear operator on the space of all polynomials.
  Similarly integration is also a linear operator

4.1 Similar matrices

Let **X,Y** $\in \mathbb{F}^{n\times n}$ then **X** is said to be similar to **Y** if there exists a linear transformation $\mathbf{P} \in \mathbb{F}^{n\times n}$ such that $T_P(\mathbf{X}) = \mathbf{P}^{-1}\mathbf{X}\mathbf{P} = \mathbf{Y}$.
Y is a coded version of **X** where the code is provided by **P**. Such $T_P(\mathbf{X})$ are called **Similarity transformations**.

4.2 Rank Nullity Theorem

$T : \mathbb{V} \longrightarrow \mathbb{W}$
dim($\mathbb{V}$) = n
dim($\mathbb{W}$) = m
$\therefore$ Nullity $\leq$ n
Rank $\leq$ m
Let Nullity be equal to k, and $u_1, u_2, ...u_k$ be the basis for Null space.
And $v_1, v_2, ...v_{n-k}$ be the remaining linearly independent vectors in $\mathbb{V}$ that form the basis for $\mathbb{V}$.
Consider any vector x in $\mathbb{V}$, then

$$\begin{aligned}
x &= \alpha_1 u_1 + \alpha_2 u_2 + .. + \alpha_k u_k + \beta_1 v_1 + \beta_2 v_2 + ... + \beta_{n-k} v_{n-k} \\
Tx &= T\big(\alpha_1 u_1 + \alpha_2 u_2 + .. + \alpha_k u_k + \beta_1 v_1 + \beta_2 v_2 + ... + \beta_{n-k} v_{n-k}\big) \\
Tx &= \beta_1 T v_1 + \beta_2 T v_2 + ... + \beta_{n-k} T v_{n-k}
\end{aligned}$$

Since the other vectors are in Null space and map to the zero vector.From the definition of Range space we see that $Tv_1, Tv_2, .., Tv_{n-k}$ is the spanning set of the range space. Hence the rank = n-k.
Rank + Nullity = n. $\implies$ Rank of $T \leq$ dim($\mathbb{V}$), dim($\mathbb{W}$).

4.3 One-One Linear Transformation

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces, then the transformation $T : \mathbb{V} \longrightarrow \mathbb{W}$ is one-one if
$Tx = Ty \iff x = y$
Advantage of having one-one transformation is that, once transformed the original vector can be obtained by applying the inverse transform.

Properties

- Null space only has the zero vector. Nullity = 0
- From Rank-Nullity theorem, the Transformation is full ranked, i.e. $\dim(\mathbb{V})$
- From Rank-Nullity theorem, Rank $\leq \dim(\mathbb{W}) \implies \dim(\mathbb{V}) \leq \dim(\mathbb{W})$
  Intuitively, if the mapping is from higher dimension to a lower dimension then information from an entire dimension is lost. And hence the inverse transform cannot obtain the original vector.
- A set of linearly independent vectors in $\mathbb{V}$ are transformed to a set of linearly independent vectors in $\mathbb{W}$.

4.4 Onto Linear Transformation

Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector space. Then $T : \mathbb{V} \longrightarrow \mathbb{W}$ is onto iff
$\forall\, w \in \mathbb{W} \ni v \in \mathbb{V} \ni Tv = w$
i.e. Range of $T = \mathbb{W}$.

Properties

- Rank of $T = \dim(\mathbb{W})$
- From Rank-Nullity theorem, Rank of $T \leq \dim(\mathbb{V})$
  $\implies \dim(\mathbb{W}) \leq \dim(\mathbb{V})$

4.5 Isomorphism

Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces over $\mathbb{F}$ then
$T : \mathbb{V} \longrightarrow \mathbb{W}$ is called an Isomorphism if $\mathbb{V}$ onto $\mathbb{W}$ is both one-one and onto.
If such a T exists then $\mathbb{V}$ is said to be isomorphic to $\mathbb{W}$, and $\dim(\mathbb{W}) = \dim(\mathbb{V})$

T is onto
$\iff$ Rank $= \dim(\mathbb{W})$
$\iff$ $\dim(\mathbb{W}) = \dim(\mathbb{V})$
$\iff$ Nullity $= 0$
$\iff$ T is one-one.
Hence every Linear transformation that is onto is also one-one and vice versa.
$\therefore$ to check if T is an isomorphism it is sufficient to check for either one-one or onto.

Consider two vector spaces of same dimension $\mathbb{V}$,$\mathbb{W}$ a transformation T that leads to isomorphism of $\mathbb{V}$ onto $\mathbb{W}$ is defined as follows,
Let $v_1, v_2, ...v_n$ be the basis for $\mathbb{V}$
$w_1, w_2, ...w_n$ be the basis for $\mathbb{W}$
the transformation $Tv_j = w_j \quad j = 1, 2, ..., n$.
Hence any two vector spaces having same dimension are isomorphic.
To construct the isomorphic transformation we require the basis for $\mathbb{V}$ and $\mathbb{W}$.

$\therefore$ An n-dimensional vector space $\mathbb{V}$ and the corresponding representation in $\mathbb{F}^n$ is an isomorphism. This method of coding a member from $\mathbb{V}$ to $\mathbb{F}^n$ is one-one and onto. Hence operations can be done in $\mathbb{V}$ or $\mathbb{F}^n$ interchangeably based on convenience.

Thus the only meaningful n-dimensional vector space is $\mathbb{F}^n$ and any other vector space of dimension n over $\mathbb{F}$ is essentially $\mathbb{F}^n$ spoken in a different language.

5 Orthogonality

If $x, y \in \mathbb{R}^k$ we say x is orthogonal to y iff $(x, y) = 0$

NOTE: θ_k is the only vector orthogonal to all vectors. Because only $(0, 0) = 0$

Let S be a non empty subset of $\mathbb{R}^k$, then the set of all vectors in $\mathbb{R}^k$ orthogonal to S is denoted by $S^\perp$.

$$S^\perp = \{u \in \mathbb{R}^k : (s, u) = 0 \; \forall s \in S\}$$

- $\theta_k \in S^\perp$ hence the set is non empty.
- Let x, y be in $S^\perp$, then $(s, (x + y)) = 0$. Hence $(x + y) \in S^\perp$. Hence it is closed under vector addition.

- Let $x \in S^{\perp}$ then $(s, \alpha x) = \alpha(s, x) = 0$. Hence it is closed under scalar multiplication.

Hence $S^{\perp}$ is a subspace in $\mathbb{R}^k$, irrespective of S being a subspace or not.
$S^{\perp}$ is the orthogonal complement of S

- To check if a given vector $y \in S^{\perp}$, it is sufficient to check if $(y, \mathbb{B}) = 0$ where $\mathbb{B}$ is the set of basis vectors of S.

$$\begin{aligned}
\mathbb{B} &= b_1, b_2, ..., b_k \\
(y, b_1) + (y, b_2) + ... + (y, b_k) &= 0 \\
(y, \alpha_1 b_1) + (y, \alpha_2 b_2) + ... + (y, \alpha_k b_k) &= 0 \\
(y, (\alpha_1 b_1 + \alpha_2 b_2 + ... + \alpha_k b_k)) &= 0
\end{aligned}$$

  Hence if y is orthogonal to basis then it is perpendicular to every vector in the space.

5.1 Orthonormal Sets

In $\mathbb{R}^k$ a set $S = u_1, u_2, ..., u_k$ of vectors is said to be orthonormal if

$$(u_i, u_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Distinct vectors are orthogonal and every vector has magnitude of 1.

- A set of k orthonormal vectors in $\mathbb{R}^k$ is linearly independent.

$$\begin{aligned}
U &= \{u_1, u_2, ..., u_k\} \\
\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k &= \theta_k
\end{aligned}$$

  We can choose α_i such that the above is true.
  The dot product of the above set with any vector in $\mathbb{R}^k$ should be 0.

$$((\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k), x) = 0 \quad \forall x$$

  Let $x = u_1$

$$\alpha_1(u_1, u_1) + \alpha_2(u_2, u_1) + ... + \alpha_k(u_k, u_1) = 0$$

  Hence, $\alpha_1 = 0$.
  Similarly repeating the above procedure taking $x = u_2, u_3, ..., u_k$ we get $\alpha_2 = \alpha_3 = ... = \alpha_k = 0$.
  Hence set of orthonormal vectors is linearly independent.

5.2 Orthonormal Basis

A set of k vectors S in $\mathbb{R}^k$ such that it is
$i)$ Orthonormal
$ii)$ $\mathbb{L}[S] = \mathbb{R}^k$
is called an **Orthonormal basis**.
If we have orthognal basis divide every vector by its length to get the corresponding orthonormal basis.

- Advantage of having orthonormal basis is that, it is easy to find the co efficients.

$$x = (x, u_1)u_1 + (x, u_2)u_2 + ... + (x, u_k)u_k$$

  The above expansion is called **Fourier Expansion** of x w.r.t the orthonormal basis $\mathbb{B} = u_1, u_2, ..., u_k$.
  (x, u_j) is the j^{th} **Fourier co efficient** of x w.r.t $\mathbb{B}$.

- $\mathbb{S} = e_1, e_2, ...e_k$ is the standard orthonormal basis. Dot product of two vectors is equal to the sum of product of the Fourier co efficients corresponding to $\mathbb{S}$.

$$\begin{aligned} x &= x_1e_1 + x_2e_2 + .. + x_ke_k \\ y &= y_1e_1 + y_2e_2 + .. + y_ke_k \\ (x, y) &= x_1y_1 + x_2y_2 + ... + x_ky_k \end{aligned}$$

  In case of a general orthonormal basis

$$\begin{aligned} x &= x_1\phi_1 + x_2\phi_2 + .. + x_k\phi_k \\ y &= y_1\phi_1 + y_2\phi_2 + .. + y_k\phi_k \\ (x, y) &= (x_1, \phi_1)(y_1, \phi_1) + (x_2, \phi_2)(y_2, \phi_2) + ... + (x_k, \phi_k)(y_k, \phi_k) \\ (x, y) &= x_1y_1 + x_2y_2 + ... + x_ky_k \end{aligned}$$

  Dot product of two vectors is equal to the sum of product of the Fourier co efficients.

$$(x, y) = \sum_{j=1}^{k}(x, \phi_j)(y, \phi_j)$$

  is called **Plancheral's formula**.
  If $x = y$ in the above formula we get,
  Length of a vector is equal to sum of squares of Fourier co efficients.This is called **Parseval's identity**.

5.3 Gram Schimdt Orthonormalization

GOAL: Given $S = u_1, u_2, ..., u_r$ a linearly independent set in $\mathbb{R}^k$,
to produce $O = \phi_1, \phi_2, ..., \phi_r$ an orthonormal set in $\mathbb{R}^k$ such that $\mathbb{L}[S] = \mathbb{L}[O]$.
The above is achieved in a recursive manner such that

$$\begin{aligned} S_j &= u_1, u_2, ..u_j \qquad 1 \leq j \leq r \\ O_j &= \phi_1, \phi_2, ..\phi_j \\ \mathbb{L}[S_j] &= \mathbb{L}[O_j] \end{aligned}$$

Procedure:

Given $u_1, u_2, ...u_k$ a set of linearly independent vectors.

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= u_2 - (u_2, v_1)\frac{v_1}{||v_1||^2} \\ &. \\ &. \\ &. \\ v_j &= u_j - \sum_{i=1}^{j-1}(u_j, v_i)\frac{v_i}{||v_i||^2} \end{aligned}$$

$v_1, v_2, ..., v_j$ are the set of orthogonal vectors. Further dividing the orthogonal vectors by their length, we get a set of orthonormal vectors.

$$\begin{aligned} \phi_1 &= \frac{v_1}{||v_1||^2} \\ \phi_2 &= \frac{v_2}{||v_2||^2} \\ &. \\ &. \\ \phi_j &= \frac{v_j}{||v_j||^2} \end{aligned}$$

$O = \phi_1, \phi_2, ..., \phi_r$ is the required orthonormal set of vectors such that $\mathbb{L}[S_j] = \mathbb{L}[O_j]$.

Hence given any set of linearly independent vectors can be extended to an orthonormal basis.

5.4 Orthogonal complement

Consider a subspace $\mathbb{W}$ of dim = d in $\mathbb{R}^k$, with an orthonormal basis $B_W = \phi_1, \phi_2, .., \phi_d$. This can be extended to be an orthonormal basis for the entire space. Let the vectors to be appended be $\psi_1, \psi_2, ..., \psi_{k-d}$, which are orthonormal as well.

$\phi_1, \phi_2, .., \phi_d, \psi_1, \psi_2, ..., \psi_{k-d}$ forms an orthonormal basis for $\mathbb{R}^k$.

Any vector in $\mathbb{R}^k$ can be expressed as

$$x = \sum_{j=1}^{d}(x, \phi_j)\phi_j \ + \ \sum_{j=1}^{k-d}(x, \psi_j)\psi_j$$

We know that any vector that is perpendicular to all basis vectors of $\mathbb{W}$ belongs to $\mathbb{W}^\perp$. Hence the vectors $\psi_1, \psi_2, ..., \psi_{k-d} \in \mathbb{W}^\perp$
Consider a vector $y \in \mathbb{W}^\perp$, and since $\mathbb{W}^\perp$ is a subspace of $\mathbb{R}^k$ it can be expressed by a linear combination of basis vectors of $\mathbb{R}^k$, and by definition of $\mathbb{W}^\perp$ we have $(y, \phi_j) = 0$. Hence we have

$$y = \sum_{j=1}^{k-d}(x, \psi_j)\psi_j$$

$\therefore$ any vector in $\mathbb{W}^\perp$ can be expressed interms of linear combination of $\psi_1, \psi_2, ..., \psi_{k-d}$. Hence it forms an orthonormal basis for $\mathbb{W}^\perp$.

dim($\mathbb{W}$) + dim($\mathbb{W}^\perp$) = dimension of the space.

$\mathbb{B}_W \cup \mathbb{B}_{W^\perp} = \mathbb{B}$ **Basis of the space.**
Any vector in $\mathbb{R}^k$ can be written as a sum of $x_W \in \mathbb{W}$ and $x_{W^\perp} \in \mathbb{W}^\perp$

- The decomposition of x into x_W and $x_{W^\perp}$ is unique.
  Let

$$\begin{aligned} x = x'_W + x'_{W^\perp} &= x_W + x_{W^\perp} \\ x'_W - x_W &= x_{W^\perp} - x'_{W^\perp} \\ X_W &= X_{W^\perp} \end{aligned}$$

The only vector common to $\mathbb{W}$ and $\mathbb{W}^\perp$ is θ_k.

$$\begin{aligned} X_W = X_{W^\perp} &= \theta_k \\ \implies x_W = x'_W, x_{W^\perp} &= x'_{W^\perp} \end{aligned}$$

x_W is the **orthogonal projection** of x onto $\mathbb{W}$,
$x_{W^\perp}$ is the **orthogonal projection** of x onto $\mathbb{W}^\perp$.

$$\begin{aligned}||x||^2 &= (x,x) = (\ (x_W + x_{W^\perp}),\ (x_W + x_{W^\perp})\ ) \\ &= (x_W, x_W) + (x_W, x_{W^\perp}) + (x_{W^\perp}, x_W) + (x_{W^\perp}, x_{W^\perp}) \\ &= ||x_W||^2 + 0 + 0 + ||x_{W^\perp}||^2 \\ &= ||x_W||^2 + ||x_{W^\perp}||^2\end{aligned}$$

Which is nothing but the **Pythagorus Theorem**.

- $(\mathbb{W}^\perp)^\perp = \mathbb{W}$

5.5 Orthogonal subspace and Matrices

Let $A \in \mathbb{R}^{m\times n}$

$$\begin{aligned}N_A &= \{x \in \mathbb{R}^n\ :\ Ax = \theta_m\} \\ R_{A^T} &= \{x \in \mathbb{R}^n\ :\ x = A^T b\ \forall\ b \in \mathbb{R}^m\} \\ N_{A^T} &= \{x \in \mathbb{R}^m\ :\ A^T x = \theta_n\} \\ R_A &= \{b \in \mathbb{R}^m\ :\ b = Ax\ \forall\ x \in \mathbb{R}^n\}\end{aligned}$$

Let $x \in N_A$

$$\begin{aligned}Ax &= \theta_m \\ \iff (Ax, b) &= 0\ \forall\ b \in \mathbb{R}^m \\ \iff b^T(Ax) &= 0\ \forall\ b \in \mathbb{R}^m \\ \iff (b^T A)x &= 0\ \forall\ b \in \mathbb{R}^m \\ \iff (A^T b)^T x &= 0\ \forall\ b \in \mathbb{R}^m \\ \iff (x, A^T b) &= 0\ \forall\ b \in \mathbb{R}^m \\ \implies x &\perp R_{A^T} \\ \implies N_A &\perp R_{A^T}\end{aligned}$$

Null space and Row space are orthogonal complements in $\mathbb{R}^n$
Left Null space and Column space are orthogonal complements in $\mathbb{R}^m$

- From Rank-Nullity theorem, for $A \in \mathbb{R}^{m\times n}$
  $\implies$ Rank(A) + Nullity(A) = n
  Null space and Row space are orthogonal complements in $\mathbb{R}^n$
  $\implies$ Rank(A^T) + Nullity(A) = n
  $\implies$ **Rank(A) = Rank(A^T)**.

- **Best Approximation problem**
  Consider a subspace W in $\mathbb{R}^k$ with dim(W) = d, find w_0 in W for a given x such that

$$||x - w_0||^2\ \leq\ ||x - w||^2\ \forall w \in W$$

  i.e. w_0 is the best possible approximation of x from the vectors in W.

Let $\phi_1, \phi_2, ..., \phi_d$ be the orthonormal basis of W.

$$
\begin{aligned}
w &= \alpha_1\phi_1 + \alpha_2\phi_2 + .. + \alpha_d\phi_d \\
||x - w||^2 &= \big((x - (\alpha_1\phi_1 + \alpha_2\phi_2 + .. + \alpha_d\phi_d), (x - (\alpha_1\phi_1 + \alpha_2\phi_2 + .. + \alpha_d\phi_d)\big) \\
&= (x, x) - \alpha_1(x, \phi_1) - \alpha_2(x, \phi_2) - \alpha_1(x, \phi_1) - \alpha_2(x, \phi_2) - .... - \alpha_d(x, \phi_d) - \alpha_d(x, \phi_d)
\end{aligned}
$$

$+\alpha_1^2 + \alpha_2^2 + ... + \alpha_d^2$

$$
\begin{aligned}
&= ||x||^2 + \sum_{j=1}^{d}\{\alpha_j^2 - 2\alpha_j(x, \phi_j)\} \\
&= ||x||^2 + \sum_{j=1}^{d}\{\alpha_j^2 - 2\alpha_j(x, \phi_j) + (x, \phi_j)^2 - (x, \phi_j)^2\} \\
&= ||x||^2 + \sum_{j=1}^{d}\{\big(\alpha_j - (x, \phi_j)\big)^2\} - \sum_{j=1}^{d}(x, \phi_j)^2
\end{aligned}
$$

To minimize the above expression we need the middle term to be zero

$$
\alpha_j = (x, \phi_j)
$$

$$
\therefore w_0 = (x, \phi_1)\phi_1 + (x, \phi_2)\phi_2 + .. + (x, \phi_d)\phi_d
$$

The corresponding error(Least square error) is $||x_{W^\perp}||^2$

The best approximation for a vector in a given subspace is the orthogonal projection of that vector onto the subspace. This is the idea behind the least squares solution.

6 Eigen Decomposition - Diagonalisation

Problem: Given $\mathbf{A} \in \mathbb{R}^{n \times n}$ find an invertible matrix $\mathbf{P}$ such that,

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

D is a diagonal matrix.

Let the columns of $\mathbf{P}$ be $p_1, p_2, .., p_n$

$$P = [p_1 \ p_2 \ .... \ p_n]$$
$$AP = PD$$
$$[Ap_1 \ Ap_2 \ ... \ Ap_n] = [p_1 \ p_2 \ .... \ p_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

$$[Ap_1 \ Ap_2 \ ... \ Ap_n] = [\lambda_1 p_1 \ \lambda_2 p_2 \ .... \ \lambda_n p_n]$$

$\implies$ There exists n real numbers $\lambda_1, \lambda_2, ..., \lambda_n$ (not necessarily distinct) and n linearly independent vectors $p_1, p_2, ..., p_n$ such that

$$Ap_j = \lambda_j p_j \quad \forall \, 1 \leq j \leq n \tag{1}$$

- For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ to be diagonalizable there exists n real numbers and n linearly independent vectors such that (1) is satisfied.
- Conversely, if n real numbers and n linearly independent vectors satisfy (1), then a diagonalizable matrix $\mathbf{A}$ constructed by these is given by

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

Theorem: A $\in \mathbb{R}^{n \times n}$ is diagonalizable $\iff Ap_j = \lambda_j p_j \quad \forall \, 1 \leq j \leq n$

6.1 Eigen Values and Eigen Vectors

A real number $\lambda \in \mathbb{R}$ is said to be an **eigen value** of the matrix $\mathbf{A}$ if there exists a non zero vector ϕ such that

$$A\phi = \lambda\phi$$

ϕ is the **eigen vector** of $\mathbf{A}$ corresponding to the **eigen value** λ.
(λ, ϕ) is called **eigen pair** or **characteristic pair**.

$\therefore$ For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ to be diagonalizable requires the existance of n Eigen pairs.

Suppose we know the eigen values(λ_j), we can find the corresponding eigen vectors using (1)

$$\begin{aligned} A\phi_j &= \lambda_j \phi_j \\ (A - \lambda_j I)\phi_j &= \theta_n \\ A_{\lambda_j}\phi_j &= \theta_n \end{aligned}$$

This is similar to solving a system of homogeneous linear equations.
Eigen vector cannot be a zero vector, hence the above system of homogeneous linear equations has a non trivial solution.
$\implies$ A_{λ_j} is not invertible (Dimension of Null space is > 0)
$\implies$ Determinant of A_{λ_j} is 0.

$$\begin{aligned} det(A_{\lambda_j}) &= 0 \\ det(\lambda I - A) &= 0 \\ C_A(\lambda) &= 0 \end{aligned}$$

The roots of the above **Characteristic Equation** are the eigen values of the Matrix. $C_A(\lambda)$ is the **Characteristic Polynomial** of the matrix.

NOTE: The leading co efficient in a polynomial is 1 $\rightarrow$ "Monic polynomial". The characteristic equation is a monic polynomial in λ

- The roots of characteristic polynomial of $\mathbf{A} \in \mathbb{R}^{n \times n}$ need not be real always. Because for a polynomial with real co-efficients the root can be real or complex conjugates. Hence eigen values can be real or complex and the corresponding eigen vectors can also be real or complex.

6.1.1 Algebraic Multiplicity

The characteristic equation can be factorized using the eigen values.

$$\begin{aligned} C_A(\lambda) \;=\; (\lambda - \lambda_1)^{a_1}(\lambda - \lambda_2)^{a_2}...(\lambda - \lambda_k)^{a_k} \\ a_1 + a_2 + .. + a_k = n \\ a_1 \geq 1, a_2 \geq 1, ...\; a_k \geq 1 \end{aligned}$$

$\lambda_1, \lambda_2, .., \lambda_k$ are the distinct eigen values of A, and $a_1, a_2, .., a_k$ are the algebraic multiplicity of the corresponding eigen values.

6.1.2 Significance of eigen value = 0

- Determinant is 0
- Matrix is not invertible
- Algebraic multiplicity of zero eigen values is equal to Nullity of the matrix.
- If all eigen values are zero, then the matrix is a **Nilpotent Matrix**.
- Geometrically it means, there is zero information pertaining to the matrix along the eigen vector corresponding to zero eigen value.

6.1.3 Geometric Multiplicity

$$A\phi = \lambda_j \phi$$
$$(A - \lambda_j I)\phi = 0$$

Hence eigen vector corresponding to λ_j lies in the Null space of the matrix $(A - \lambda_j I)$.

$$W_j = \{x \in \mathbb{C}^n \ : \ (A - \lambda_j I)x = \theta_n\}$$

W_j is the Null space of $(A - \lambda_j I)$, and is called the **eigen space** corresponding to eigen value λ_j. Every non zero vector in W_j is an eigen vector for λ_j. Dimension of W_j is called **Geometric Multiplicity** of the eigen value λ_j.

$$\dim(W_j) = g_j$$
$$g_j \geq 1 \quad \forall \ 1 \leq j \leq k$$

Where k is the number of distinct roots of the characteristic equation.

Remark: For every eigen value λ_j of A,

$$1 \ \leq \ g_j \ \leq \ a_j \ , \quad j = 1, 2, ..., k$$
$$g_j \text{ is the geometric multiplicity}$$
$$a_j \text{ is the algebraic multiplicity}$$

- **All Eigen vectors corresponding to an eigen value are linearly independent**
  Let $\phi_1, \phi_2, .., \phi_{g_j}$ be the basis for W_j. Since these are the non zero vectors in W_j, they are the eigen vectors corresponding to λ_j. Hence eigen vectors form a basis for W_j and are linearly independent.

6.1.4 Lagrange Interpolation polynomial

Let $\lambda_1, \lambda_2, .., \lambda_k$ be the distinct eigen values of A. Consider the follwing polynomials

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)...(\lambda - \lambda_k)$$
$$\lambda_1, \lambda_2, .., \lambda_k, \text{are the roots of the polynomial}$$

$$p_1(\lambda) = (\lambda - \lambda_2)(\lambda - \lambda_3)...(\lambda - \lambda_k)$$
$$p_2(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_3)...(\lambda - \lambda_k)$$
$$.$$
$$.$$
$$p_k(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_{k-1})$$
$$p_j(\lambda) = \prod_{r=1, r\neq j}^{k} (\lambda - \lambda_r) \quad j = 1, 2, .., k$$

$p_j(\lambda)$ is a monic polynomial of degree (k-1).
Define

$$l_j(\lambda) = \frac{p_j(\lambda)}{p_j(\lambda_j)} = \prod_{r=1, r\neq j}^{k} \frac{(\lambda - \lambda_r)}{(\lambda_j - \lambda_r)}$$

$$l_j(\lambda_r) = \begin{cases} 0 & r \neq j \\ 1 & r = j \end{cases}$$

$l_1(\lambda), l_2(\lambda), .., l_k(\lambda)$ are called the **Lagrange Interpolation Polynomials** corresponding to the distinct points $\lambda_1, \lambda_2, .., \lambda_k$.

$$\begin{aligned} Au &= \lambda u \\ A(Au) &= A(\lambda u) \\ A^2 u &= \lambda^2 u \\ &\text{Similarly,} \\ A^r u &= \lambda^r u \end{aligned}$$

Consider a polynomial evaluated on the matrix

$$\begin{aligned} p(A) &= a_0 I + a_1 A + a_2 A^2 + .. + a_r A^r \\ p(A)u &= (a_0 I + a_1 A + a_2 A^2 + .. + a_r A^r)u \\ p(A)u &= a_0 u + a_1 Au + a_2 A^2 u + .. + a_r A^r u \\ p(A)u &= (a_0 + a_1 \lambda + a_2 \lambda^2 + .. + a_r \lambda^r)u \\ p(A)u &= p(\lambda)u \end{aligned}$$

Suppose $u_1, u_2, .., u_k$ are the eigen vectors corresponding to distinct eigen values $\lambda_1, \lambda_2, .., \lambda_k$ respectively. Suppose,

$$\alpha_1 u_1 + \alpha_2 u_2 + .. + \alpha_k u_k = \theta_n \tag{1}$$

$$\begin{aligned} l_j(A)[\alpha_1 u_1 + \alpha_2 u_2 + .. + \alpha_k u_k] &= l_j(A)\theta_n \\ \alpha_1 l_j(A)u_1 + \alpha_2 l_j(A)u_2 + .. + \alpha_k l_j(A)u_k &= \theta_n \\ \alpha_1 l_j(\lambda_1)u_1 + \alpha_2 l_j(\lambda_2)u_2 + .. + \alpha_k l_j(\lambda_k)u_k &= \theta_n \end{aligned}$$

From the definition of Lagrange interpolation polynomial we get,

$$\begin{aligned} \alpha_j l_j(\lambda_j)u_j &= \theta_n \\ \alpha_j u_j = \theta_n &\implies \alpha_j = 0 \end{aligned}$$

Repeating the above procedure by pre mutiplying $l_1(A), l_2(A), ..., l_k(A)$ we can show that $\alpha_1, \alpha_2, .., \alpha_k$ are equal to 0.
Hence the only solution to (1) is the trivial solution. Hence $u_1, u_2, .., u_k$ are linearly independent.

Eigen vectors corresponding to distinct eigen values are linearly independent.

CASE 1: Algebraic Multiplicity = Geometric multiplicity

$$\begin{aligned} a_j &= g_j \\ a_1 + a_2 + .. + a_k &= n \\ \implies g_1 + g_2 + .. + g_k &= n \end{aligned}$$

Hence we have n Eigen vectors, which implies we have n eigen pairs. Hence the corresponding matrix can be diagonalized.

Let $u_1^j, u_2^j, .., u_{a_j}^j$ be the eigen vectors corresponding to λ_j.
Then the required P matrix for diagonalisation is

$$P = [u_1^1 \ u_2^1 \ .. \ u_{a_1}^1 \ u_1^2 \ u_2^2 \ .. \ u_{a_2}^2 \ ..... \ u_1^k \ u_2^k \ .. \ u_{a_k}^k]$$

Since all the eigen vectors are linearly independent, P is invertible. The resulting Diagonal matrix contains the corresponding eigen values along the main diagonal.

For a matrix to be diagonalizable it requires Algebraic multiplicity to be equal to geometrix multiplicity

- Since all the eigen vectors are linearly independent and they span the entire space, they qualify to be a basis. And this is called **Eigen basis**.

CASE 2: Algebraic Multiplicity > Geometric multiplicity

We use the concept of generalized eigen vectors and Jordan form.

Cannot be diagonalised

Alternatively we can look at Singular Value Decomposition, in which the matrix is converted to an essentially diagonal matrix, which means the significant part of the block matrix is the leading diagonal block.

6.2 Additional Properties of Eigen Values and Eigen vectors

- Sum of eigen values of a matrix is equal to Trace of the matrix
- Product of eigen values of a matrix is equal to determinant of the matrix
- A and A^T have the same eigen values
- $$\begin{aligned} Ax &= \lambda x \\ A^k x &= \lambda^k x \\ kAx &= k\lambda x \end{aligned}$$

- A $\pm$ kI has an eigen value $\lambda \pm k$
- Eigen value of Orthogonal matrix has magnitude 1, $|\lambda| = 1$
- Eigen value of Projection matrix is 0,1
- Eigen value of Rotation matrices and Skew symmetric matrices can be complex
- Eigen values of diagonal and triangular matrices are the principle diagonal elements

When is matrix multiplication commutative
Let two square matrices A,B have dimensions such that matrix multiplication is compatible.
For AB = BA, the matrices A and B must be simultaneously diagonalizable.
i.e. A and B share the same eigen vectors.
Let Eigen values of A be $\lambda_1, \lambda_2, ..., \lambda_n$
Eigen values of B are $\beta_1, \beta_2, .., \beta_n$
Eigen vectors of A,B are $v_1, v_2, ..., v_n$

D_A = diag$[\lambda_1, \lambda_2, ..., \lambda_n]$
D_B = diag$[\beta_1, \beta_2, .., \beta_n]$

$$T = [v_1 \; v_2 \; .. \; v_n]$$
$$TD_AT^{-1} = A$$
$$TD_BT^{-1} = B$$
$$AB = TD_AT^{-1}TD_BT^{-1} = TD_AD_BT^{-1}$$

Diagonal matrices are commutative.

$$AB = TD_BD_AT^{-1} = TD_BT^{-1}TD_AT^{-1} = BA$$

7 Hermitian and Symmetric matrices

The class of matrices for which Algebraic multiplicity is always equal to Geometric multiplicity. Hence can always be diagonalised.
Consider a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, and vectors $x, y \in \mathbb{C}^n$

$$
\begin{aligned}
(Ax)_j &= \sum_{k=1}^{n} a_{jk}x_k \\
(Ax, y) &= \sum_{j=1}^{n} (Ax)_j \overline{y_j} \\
(Ax, y) &= \sum_{j=1}^{n} \sum_{k=1}^{n} (a_{jk}x_k)\ \overline{y_j} \\
(Ax, y) &= \sum_{k=1}^{n} x_k \sum_{j=1}^{n} (a_{jk}\overline{y_j}) \\
(Ax, y) &= \sum_{k=1}^{n} x_k \overline{\sum_{j=1}^{n} \overline{a_{jk}} y_j} \\
(Ax, y) &= \sum_{k=1}^{n} x_k \overline{\sum_{j=1}^{n} a^*_{kj} y_j} \\
(Ax, y) &= \sum_{k=1}^{n} x_k \overline{(A^*y)}_k \\
(Ax, y) &= (x, A^*y)
\end{aligned}
$$

Defining $A^* = \overline{A^T}$ we get, $\quad (Ax, y) = (x, A^*y)$
A^* is called **Hermitian Conjugate** of A.
If $\mathbf{A} \in \mathbb{R}^{n\times n}$, then the above reduces to, $\quad (Ax, y) = (x, A^T y)$

From the above it can be observed that, when the matrix A has to move from one variable to another within an inner product, it moves as A^*. The computation of moving A from one variable to another within an inner product becomes easier if $A = A^*$.

Hermitian Matrix
A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is said to be Hermitian if $A^* = A$.

Real Symmetric Matrix
A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be Hermitian if $A^T = A$
Real symmetric matrices $\subset$ Hermitian matrices

Let H_n denote the class of Hermitian matrices
Properties:

- A $\in H_n \implies$ (Ax,y) = (x,Ay)
- $A^* = A \implies$ Diagonal entries are real
- (Ax,x) = (x,Ax) = $\overline{(Ax, x)} \implies$ (Ax,x) is real, $\forall x \in \mathbb{C}^n$ and $\forall A \in \mathbb{C}^{n\times n}$
- A,B $\in H_n \implies$ (A + B) $\in H_n$
- A $\in H_n$ then $\alpha A \in H_n$ if and only if α is real.
  $\therefore$ H_n is a vector space over $\mathbb{R}$ and not $\mathbb{C}$.
- A,B $\in H_n$ then AB $\in H_n \iff AB = BA$

Unitary Matrix:
$U \in \mathbb{C}^{n\times n}$ is said to be a unitary matrix if UU* = U*U = I $\implies$ U* = U^{-1}
All the columns of an Unitary matrix are orthonormal vectors.

Orthogonal Matrix:
$O \in \mathbb{R}^{n\times n}$ is said to be an orthogonal matrix if $O^T O = I \implies O^T = O^{-1}$
All the columns of an orthogonal matrix are orthonormal vectors.

7.1 Eigen values and eigen vectors of Hermitian matrices

- Let (λ, u) be an eigen pair for A $\in H_n$.
  λ(u,u) = (λu,u) = (Au,u) $\implies \lambda = \frac{(Au,u)}{(u,u)}$
  $\therefore$ λ is real.
  Every eigen value of Hermitian matrix/Real symmetric matrix is real.
- A $\in H_n$ with distinct eigen values λ, β and corresponding eigen vectors u, v

$$\begin{aligned}
\lambda(u, v) &= (\lambda u, v) \\
&= (Au, v) \\
&= (u, Av) \\
&= (u, \beta v) \\
&= (u, v)\beta^* = (u, v)\beta \qquad \text{since } \beta \text{ is real} \\
(\lambda - \beta)(u, v) &= 0 \\
\implies (u, v) &= 0 \\
\therefore u &\perp v
\end{aligned}$$

Eigen vectors corresponding to distinct eigen values of Hermitian matrices/Real symmetric matrices are orthogonal to each other.

- Let W_j be the null space of $(A - \lambda_j I)$, hence every non zero vector in W_j is an eigen vector of A corresponding to λ_j. Using Gram-Schmidt orthonormalization we can obtain an orthonormal set of eigen vectors
  We can repeat the same for other eigen values, and since eigen vectors corresponding to distinct eigen values are orthogonal, we can obtain an orthonormal basis of eigen vectors for the entire space.
  $B = \{\phi_1^1\ \phi_2^1\ ..\ \phi_{a_1}^1\ \phi_1^2\ \phi_2^2\ ..\ \phi_{a_2}^2\ .....\ \phi_1^k\ \phi_2^k\ ..\ \phi_{a_k}^k\}$
  Stacking the eigen vectors as columns we get an Unitary matrix.

- U*AU = diagonal matrix
  A $\in H_n \implies$ **A is unitarily diagonalizable**

  For the real case,
  O^{T}AO = diagonal matrix
  A is real symmetric matrix $\implies$ **A is orthogonally diagonalizable**

- PRODUCT DECOMPOSITION of Hermitian matrix
  A = UDU*

- SUM DECOMPOSITION (SPECTRAL DECOMPOSITION)
  Let the eigen vectors of A $\in H_n$ of rank ρ be

$$\phi_1^1\ \phi_2^1\ ..\ \phi_{a_1}^1\ \phi_1^2\ \phi_2^2\ ..\ \phi_{a_2}^2\ .....\ \phi_1^k\ \phi_2^k\ ..\ \phi_{a_k}^k$$

$$x = \sum_{j=1}^{k}\sum_{r=1}^{a_j}(x, \phi_r^j)\phi_r^j$$

$$Ax = \sum_{j=1}^{k}\sum_{r=1}^{a_j}(x, \phi_r^j)\ A\phi_r^j$$

$$Ax = \sum_{j=1}^{k}\sum_{r=1}^{a_j}(\phi_r^j)^* x\ \lambda_j \phi_r^j$$

$$Ax = \sum_{j=1}^{k}\sum_{r=1}^{a_j}\lambda_j \phi_r^j\ (\phi_r^j)^* x$$

$$\implies A = \sum_{j=1}^{k}\sum_{r=1}^{a_j}\lambda_j \phi_r^j\ (\phi_r^j)^*$$

$$\implies A = \sum_{j=1}^{k}\sum_{r=1}^{a_j}\lambda_j (\phi_r^j \otimes \phi_r^j)$$

  Tensor product results in matrix of rank 1, adding n such matrices($a_1 + a_2 + .. + a_k = n$) results in matrix of rank n. This is true only because the eigen vectors included in the summation of such tensor products are linearly independent.
  This is the required decomposition of A as a sum of ρ rank 1 matrices.

Tensor product $\rightarrow$ Matrix of rank 1
Proof:
Let T = u $\otimes$ u,
For any x $\in \mathbb{R}^n$ we have
Tx = (uu*)x = u(u*x) = α u
If x = u $\implies$ Tu = (uu*)u = u
$\therefore$ u $\in$ Range of T. And every other vector in the range of T is a scalar multiple of u. Hence the number of linearly independent vectors in the range of T is 1. Which is why the dimension of Range of T is 1 i.e. rank of T = (uu*) = 1.

- Tensor product of any vector in $\mathbb{C}^n$ with itself results in a Hermitian matrix of rank 1. $\therefore$ SUM DECOMPOSITION expresses the given matrix as a sum of ρ rank 1 Hermitian matrices.

7.2 Positive Semi Definite Matrices

A set of matrices **A** in H_n for which (Ax,x) ≥ 0 $\forall x \in \mathbb{C}^n$.
if further (Ax,x) > 0 $\forall x \neq \theta_n$, then A is said to be **positive definite matrix**.
Properties:

- All properties of Hermitian matrices apply to positive semi definite matrices.
- All eigen values are real and non negative.
  Au = λu
  (Au,u) = (λu,u) = λ(u,u) $\implies \lambda = \frac{(Au,u)}{(u,u)} \geq 0$
- Eigen vectors corresponding to zero eigen values, forms an orthonormal basis for Null space of A.
  Eigen vectors corresponding to positive eigen values, forms an orthonormal basis for Range of A.

7.2.1 Transforming a given matrix to a positive semi definite matrix

Given $\mathbf{A} \in \mathbb{R}^{m\times n}$ construct a positive semi definite matrix $\mathbf{L} \in \mathbb{R}^{n\times n}$ such that the analysis of **L** reflects in the analysis of **A**. Where Rank(A) = ρ

$$\mathbf{L} = \mathbf{A}^T\mathbf{A}$$

- $x \in \mathbb{R}^n$ (Lx,x) = (A^{T}Ax,x) = (Ax,Ax) = $||Ax||^2 \geq 0$
  $\implies$ L is a positive definite matrix in $\mathbb{R}^{n\times n}$
- $\mathbf{M} = \mathbf{A}\mathbf{A}^T$ is a positive semi definite matrix in $\mathbb{R}^{m\times m}$
- Null space of L is same as Null space of A. Nullity(L) = Nullity(A)
  Range of L is same as Range of A^T. Rank(L) = Rank(A)
- Similarly Range of M = Range of A, and
  Null space of A^T = Null space of M

1. Orthonormal basis for Null space of A, are the eigen vectors of L corresponding to zero eigen values. $[\phi_1, \phi_2, .., \phi_{\nu_A}]$
2. Orthonormal basis for Null space of A^T, are the eigen vectors of M corresponding to zero eigen values. $[\psi_1, \psi_2, .., \psi_{\nu_{A^T}}]$
3. Orthonormal basis for Range of A^T, are the eigen vectors of L corresponding to positive eigen values. $[v_1, v_2, .., v_\rho]$
4. Orthonormal basis for Range of A, are the eigen vectors of M corresponding to positive eigen values.$[u_1, u_2, .., u_\rho]$

   Alternatively, Orthonormal basis for Range of A, can also be obtained from the eigen vectors of L corresponding to positive values i.e. Range of A^T
   Let λ_r be the eigen values of L. Consider,

$$\begin{aligned} Av_j &= w_j \\ (w_j, w_r) &= (Av_j, Av_r) = (v_j, A^T Av_r) \\ &= (v_j, Lv_r) = (v_j, \lambda_r v_r) \\ &= (v_j, v_r)\lambda_r \end{aligned}$$

$$(w_j, w_r) = \begin{cases} 0 & j \neq r \\ \lambda_r & j = r \end{cases}$$

$\therefore$ $w_1, w_2, .., w_\rho$ are a set of orthogonal vectors. Dividing by their length we get an orthonormal set of vectors. Moreover since Rank(A^T) = Rank(A). This orthonormal set of vectors forms an orthonormal basis. Let $u_1, u_2, .., u_\rho$ represent the basis

$$u_j = \frac{w_j}{||w_j||} = \frac{w_j}{\sqrt{\lambda_j}}$$
$$\text{Let } \; s_j = \sqrt{\lambda_j}$$

s_j are called the singular values of A

$$\begin{aligned} u_j &= \frac{Av_j}{s_j} \\ Av_j &= s_j u_j \quad j = 1, 2, ..., \rho \qquad (1) \\ A^T u_j &= \frac{A^T Av_j}{s_j} = \frac{Lv_j}{s_j} = \frac{\lambda_j v_j}{\sqrt{\lambda_j}} \\ A^T u_j &= s_j v_j \qquad (2) \end{aligned}$$

From (1) and (2) it can be observed that, the basis vectors of range of A and A^T are connected such that they point to the same direction with a scaling of the corresponding singular value.

7.3 Singular Value Decomposition (product version)

Consider the product decomposition of a mxn matrix. Using the orthonormal bases of L and M as defined above,

$$U = [u_1, u_2, .., u_\rho, \psi_1, \psi_2, .., \psi_{\nu_{A^T}}] \in \mathbb{R}^{m \times m}$$
$$V = [v_1, v_2, .., v_\rho, \phi_1, \phi_2, .., \phi_{\nu_A}] \in \mathbb{R}^{n \times n}$$

U,V are orthogonal matrices.

$$AV = A[v_1, v_2, .., v_\rho, \phi_1, \phi_2, .., \phi_{\nu_A}]$$
$$AV = [Av_1, Av_2, .., Av_\rho, A\phi_1, A\phi_2, .., A\phi_{\nu_A}]$$
$$AV = [s_1 u_1, s_2, u_2, .., s_\rho u_\rho, \theta_m, \theta_m, .., \theta_m]$$
$$U^T AV = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_\rho \\ \psi_1 \\ \psi_2 \\ . \\ . \\ \psi_{\nu_{A^T}} \end{bmatrix} [s_1 u_1, s_2, u_2, .., s_\rho u_\rho, \theta_m, \theta_m, .., \theta_m]$$
$$U^T AV = \begin{bmatrix} s_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & s_\rho & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$U^T AV = \left[\begin{array}{c|c} S_{\rho \times \rho} & \mathbf{0}_{\rho \times (n-\rho)} \\ \hline \mathbf{0}_{(m-\rho) \times \rho} & \mathbf{0}_{(m-\rho) \times (n-\rho)} \end{array}\right]$$
$$S_{\rho \times \rho} = diag(s_1, s_2, .., s_\rho) = S_A$$
$$U^T AV = S_A$$
$$A = US_A V^T$$

Thus we have a factorization of A into 3 factors where the extreme matrices are orthogonal and the middle one is essentially a diagonal matrix. (which means the significant part of the matrix is the leading diagonal block).

SVD gives a general diagonalization/factorization theorem for any mxn real matrix

7.4 Singular Value Decomposition (sum version)

Consider any mxn matrix,
$u_1, u_2, .., u_\rho, \psi_1, \psi_2, .., \psi_{\nu_{A^T}}$ forms an orthonormal basis in $\mathbb{R}^m$
$v_1, v_2, .., v_\rho, \phi_1, \phi_2, .., \phi_{\nu_A}$ forms an orthonormal basis in $\mathbb{R}^n$

$$\therefore x \in \mathbb{R}^n \implies x = \sum_{j=1}^{\rho}(x, v_j)v_j + \sum_{r=1}^{\nu_A}(x, \phi_r)\phi_r$$

$$Ax = \sum_{j=1}^{\rho}(x, v_j)Av_j + \sum_{r=1}^{\nu_A}(x, \phi_r)A\phi_r$$

$$Ax = \sum_{j=1}^{\rho}(x, v_j)s_j u_j + 0$$

$$Ax = \sum_{j=1}^{\rho} s_j u_j v_j^T x$$

$$Ax = \sum_{j=1}^{\rho} s_j (u_j \otimes v_j)x$$

$$\implies A = \sum_{j=1}^{\rho} s_j (u_j \otimes v_j)$$

Each term in the above sum is an mxn matrix of rank 1.
Thus any A $\in \mathbb{R}^{m\times n}$ of rank ρ can be expressed as the sum of ρ one ranked matrices in $\mathbb{R}^{m\times n}$

7.5 Back to System of Linear Equations

$$Ax = b$$

A is the co efficient matrix. $\mathbf{A} \in \mathbb{R}^{m \times n}$
$\overrightarrow{x}$ is the vector of unknowns. $x \in \mathbb{R}^n$
b is the vector of constants in the equations. $b \in \mathbb{R}^m$

$$b = \sum_{j=1}^{\rho} (b, u_j) u_j + \sum_{r=1}^{\nu_{A^T}} (b, \psi_r) \psi_r \tag{1}$$

$$x = \sum_{j=1}^{\rho} (x, v_j) v_j + \sum_{r=1}^{\nu_A} (x, \phi_r) \phi_r$$

$$x = \sum_{j=1}^{\rho} \alpha_j v_j + \sum_{r=1}^{\nu_A} \beta_r \phi_r$$

$$Ax = \sum_{j=1}^{\rho} \alpha_j A v_j + \sum_{r=1}^{\nu_A} \beta_r A \phi_r$$

$$Ax = \sum_{j=1}^{\rho} \alpha_j s_j u_j + 0 \tag{2}$$

Equating (1) and (2) we get

$$\sum_{j=1}^{\rho} \alpha_j s_j u_j = \sum_{j=1}^{\rho} (b, u_j) u_j + \sum_{r=1}^{\nu_{A^T}} (b, \psi_r) \psi_r$$

$$\sum_{j=1}^{\rho} (\alpha_j s_j - (b, u_j)) u_j = \sum_{r=1}^{\nu_{A^T}} (b, \psi_r) \psi_r$$

u_j spans the Range of A, ψ_r spans the null space of A^T, since these two are orthogonal complements in $\mathbb{R}^n$, the only common vector is the zero vector.

$$\begin{aligned}
\sum_{r=1}^{\nu_{A^T}} (b,\psi_r)\psi_r = \theta_m \iff (b,\psi_r) = 0 \quad 1 \le r \le \nu_{A^T} \\
\implies b \in \text{Range of A} \qquad (3) \\
\sum_{j=1}^{\rho} (\alpha_j s_j - (b,u_j))u_j = \theta_m \\
\implies \alpha_j s_j = (b,u_j) \\
\therefore \alpha_j = \frac{(b,u_j)}{s_j} \\
\implies x = \sum_{j=1}^{\rho} \frac{(b,u_j)}{s_j} v_j + \sum_{r=1}^{\nu_A} \beta_r \phi_r \qquad (4) \\
\beta_1, \beta_2, .., \beta_{\nu_A} \text{ can be chosen arbitrarily}
\end{aligned}$$

(3) is the consistency condition to be satisfied for a system of Linear Equations to have a solution. And the solution is given by (4).

Further,
we get an unique solution only when Nullity = 0 i.e. A is full ranked

$$x = \sum_{j=1}^{\rho} \frac{(b,u_j)}{s_j} v_j$$

we get infinite solutions when Nullity $\neq 0$ i.e. A is not full ranked

$$\begin{aligned}
x_r = \sum_{j=1}^{\rho} \frac{(b,u_j)}{s_j} v_j \\
x_n = \sum_{r=1}^{\nu_A} \beta_r \phi_r \\
\text{From Pythagorus theorem} \\
||x||^2 = ||x_r||^2 + ||x_n||^2
\end{aligned}$$

Hence if there are infinite solutions, the solution with least length is obtained when $x_n = \theta_n \implies \beta_1 = \beta_2 = .. = \beta_{\nu_A} = 0$.
This solution is called **optimal solution**.

$$x_{opt} = \sum_{j=1}^{\rho} \frac{(b,u_j)}{s_j} v_j \qquad \text{where } \rho < n$$

7.5.1 Least Squares Solution to System of Linear Equations

If the consistency condition is not satisfied, i.e. b $\notin$ Range of A.
$\therefore$ b - Ax $\neq \theta_m \quad \forall x \in \mathbb{R}^n \implies ||b - Ax||^2 > 0$
The error in taking x as a solution when b does not satisfy the consistency condition is

$$e_b(x) = ||b - Ax||^2$$

The best possible solution x_l should minimize the above error, and this solution is called the **Least squares solution**

$$||b - Ax_l||^2 < ||b - Ax||^2 \quad \forall x$$

Hence A$x_l \in R_A$ (Range of A), is the closest vector to b in R_A.

$$b = b_r + b_n$$
$$b_r = \sum_{j=1}^{\rho}(b, u_j)u_j$$
$$b_n = \sum_{j=1}^{\nu_A}(b, \psi_j)\psi_j$$

b_r is the projection of b onto R_A.
b_n is the projection of b onto N_{A^T}.
And we know that the closest vector to b in R_A is the orthogonal projection of b onto R_A i.e. b_r. So the least square solutions of the system are solutions of

$$Ax = b_r$$

We get an unique least squares solution when A is full ranked

$$x_l = \sum_{j=1}^{\rho}\frac{(b_r, u_j)}{s_j}v_j = \sum_{j=1}^{\rho}\frac{(b, u_j)}{s_j}v_j$$

We get infinite least squares solution when A is not full ranked i.e. Nullity $\neq 0$

$$x_l = \sum_{j=1}^{\rho}\frac{(b, u_j)}{s_j}v_j + \sum_{k=1}^{\nu_A}\beta_k\phi_k$$

$\beta_1, \beta_2, .., \beta_{\nu_A}$ can be choosen arbitrarily
The **optimal least square solution** has the least length.
$\implies \beta_1 = \beta_2 = .. = \beta_{\nu_A} = 0$

$$x_{l_{opt}} = \sum_{j=1}^{\rho}\frac{(b, u_j)}{s_j}v_j \quad \rho < n$$

7.6 Conclusions about system of linear equations using rank of coefficient matrix

Consider a system of linear equations with ‘m’ equations and ‘n’ unknowns.
The coefficient Matrix A (mxn) has rank 'r'.
The existence of solution to the linear system of equations can be analysed using rank and dimensions of the coefficient matrix A, the constant vector b.

1. **if r = m i.e A is full column rank matrix.**
   There will be n - m free variables.
   The null space of A consists of some subspace in R^n
   Meaning, there will be infinite solutions.

2. **if r = n i.e A is full row rank matrix.**
   There will be no free variables.
   The null space of A consists of only the zero vector!
   Meaning, there will be either 0 solutions or 1 solution.

   - 0 solutions if the constant vector b is not present in the column space of A.
   - 1 solution if the constant vector b is present in the column space of A.

3. **if r = m = n i.e A is invertible matrix.**
   There will be no free variables.
   The null space of A consists of only the zero vector!
   Meaning, there will be 1 unique solution.

4. **if r < m and r < n i.e A is non-invertible matrix.**
   There will be n - r free variables.
   The null space of A may or may not consist of some subspace in R^n.
   Meaning, there will be either 0 or infinite solutions.

   - 0 solutions if the constant vector b is not present in the column space of A.
   - Infinite solution if the constant vector b is present in the column space of A.

7.7 Pseudo Inverse of a Matrix

We observe that,

$$x = \sum_{j=1}^{\rho} \frac{(b, u_j)}{s_j} v_j$$

is the unique solution/ unique least square solution when A is full ranked
is the optimal solution/optimal least square solution when A is not full ranked

$$x = \sum_{j=1}^{\rho} \frac{1}{s_j} v_j (b, u_j)$$

$$x = \sum_{j=1}^{\rho} \frac{1}{s_j} v_j u_j^T b$$

$$\text{Let } A^{\dagger} = \sum_{j=1}^{\rho} \frac{1}{s_j} v_j u_j^T$$

$$x = A^{\dagger} b$$

$A^{\dagger} \in \mathbb{R}^{n \times m}$ is called PesudoInverse of a matrix A.

If m = n, then $A^{\dagger} = A^{-1}$
Proof:
Inverse of a matrix exists if the matrix is full ranked. Hence there exists an unique solution given by $A^{-1}b$. But we know that the solution to a system of equations is given by $A^{\dagger}b$, which is unique in this case.

$$A^{-1}b = A^{\dagger}b$$

$$\implies A^{-1} = A^{\dagger}$$

7.7.1 Computation of PseudoInverse

$$Ax = b$$

$$A^T A x = Lx = A^T b$$

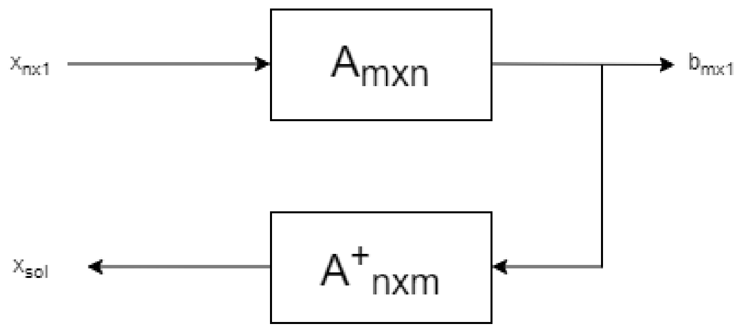
$A^T b$ lies in the range of A^T, so by premultiplying b with A^T,we have projected b onto Range of A^T. Since the range of A^T is same as range of L, the above system of equations always has a solution.

$$x = L^{-1} A^T b = (A^T A)^{-1} A^T b$$

but we know that the solution is $A^{\dagger}b \implies A^{\dagger} = (A^T A)^{-1} A^T$

7.8 Linear Systems picture

Solving Ax = b
Can be viewed as finding the input that gives the observed output from the system. To achieve this we use another system $A^{\dagger}$ whose output is x_{sol}



If there actually exists an input x which corresponds to output b then $x_{sol} = x$
If there is no input that corresponds to output b exactly, then x_{sol} is such that Ax_{sol} is closest to b.

8 Useful Matrix Decompositions

8.1 LU Decomposition

Expressing a matrix A as the product of Lower Triangular Matrix (L) and Upper Triangular Matrix (U).
Using elementary row operations, A can be converted to a upper triangular matrix.

$$\begin{aligned} EA &= U \\ A &= E^{-1}U \\ \text{Let } L &= E^{-1} \\ A &= LU \end{aligned}$$

The above decomposition assumes that no row exchanges were necessary to reduce A to U.

8.1.1 LU factorization with partial pivoting

If the decomposition of A to U, requires row swapping, then it is referred to as LU factorization with partial pivoting.

$$PA = LU$$

where P is a permutation matrix.

8.1.2 LU factorization with full pivoting

If the decomposition of A to U, requires both row and column swapping, then it is reffered to as LU factorization with full pivoting.

$$PAQ = LU$$

P is a permutation matrix that reorders rows of A
Q is a permutation matrix that reorders columns of A.

8.1.3 LDU decomposition

It is of the form

$$A = LDU$$

L is a unit lower triangular matrix, (diagonal entries are 1)
U is a unit upper triangular matrix, (diagonal entries are 1)
D is a diagonal matrix.

NOTE: LDU decomposition is unique for a given matrix.

8.2 QR Decomposition

The decomposition of A into the product of an orthogonal matrix(Q) and an upper triangular matrix(R). It is useful to solve least squares problem.
Procedure

- Apply Gram-Schmidt process on the columns of a full ranked matrix A, to get the matrix Q
- R = Q^TA
- Pre multiply by Q on both sides to get QR = A

9 Special Matrices

9.1 Permutation Matrices

These matrices, when multiplied with another matrix will simply exchange some rows or columns without modifying their values.
A permutation matrix will simply be the identity matrix with rows or columns interchanged.
These matrices are used in order to account for row exchanges in LU decomposition.
For nxn matrices, the number of permutation matrices possible = n!
All permutation matrices satisfy $P^{-1} = P^T$

9.2 Projection Matrix

A projection matrix is a nxn matrix, that projects any vector from $\mathbb{R}^n$ onto a given subspace. The columns of the projection matrix are the projections of the standard basis vectors.

NOTE: P is a projections matrix $\iff P^2 = P$

Consider 2 vectors a,b. p is the projection of b on a.
Since p is along the direction of a, it will be equal to a scalar times a. Let this scalar be denoted by x.
$e \ = \ b \ - \ p$
$p \ = \ ax$
$e \ = \ b \ - \ ax$
Since a is perpendicular to e, $a.e \ = \ 0$
$a.(b \ - \ ax) \ = \ 0 \ \ i.e \ \ a^T \ (b \ - \ ax) \ = \ 0$
Hence, $x \ = \ \frac{a^T b}{a^T a} \quad \implies p \ = \ \frac{a \ (a^T b)}{a^T a} \ = \ Pb$
where $P \ = \ \frac{aa^T}{a^T a}$ is the **Projection Matrix**.
P projects any vector onto the vector a.

9.3 Rotation Matrices

10 Keypoints

1. Addition of matrices of rank n, can result in a matrix with rank greater than or lesser than or equal to n.
2. Det(A) = 0 $\implies$ one of the Eigen Value = 0 $\implies$ Not full rank
3. If X and Y are subspaces, then
   (a) X $\cup$ Y does not give a subspace (Unless X and Y lie on the same plane)
   (b) X $\cap$ Y always gives a subspace