Linear Algebra

Dhruva Hegde

1 System of Linear Equations

A system of 'm' linear equations with 'n' unknowns can be represented using matrices and vectors as,

$$AX = b$$

A is the co-efficient matrix, of size mxn
X is the vector of unknowns, of size nx1
b is the vector of constants in the equations, of size mx1

Note that 'm' is equal to number of row of A and 'n' is equal to number of columns of A.

1.0.1 Row picture:

Consider the following example with 2 equations and 2 unknowns.
$x + 2y = 5$
$2x - y = 1$
The row picture describes intersection of hyper-planes (lines in this case)

1.0.2 Column picture:

The same equation can also be represented as a combination of columns.

$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ x + $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ y = $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$

While the row picture is a more commonly used representation, understanding and using the column picture is important in applied linear algebra.

1.0.3 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.

1.1 The Singular Case

Scenario in which a system of linear equations does not have unique solution.

1. **No solution** - there is no point of intersection between said planes
2. **Infinite solutions** - the planes are linear combinations of one another

In the column picture, if ‘n’ columns have no point in common or have infinite points in common, then these ‘n’ columns must lie on the same plane.

- **If** $\mathbf{m} < \mathbf{n}$: Unique solution does not exist; infinitely many solutions exist.
- **If** $\mathbf{m} > \mathbf{n}$: Unique solution exists only if ‘b’ can be expressed as a linear combination of ‘m’ columns of A. Otherwise no solution.
- **If** $\mathbf{m} = \mathbf{n}$: Unique solution exists unless The Singular Case occurs.

1.2 Gaussian Elimination

Converting the given system of equations in matrix form to a triangular matrix and then using back substitution to solve the system.

This conversion is done by row operations (or column operations).

1.2.1 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

(similarly column operations can be defined)

1.2.2 Echelon Form

Elementary row operations do not change the solution for the system of equations. Hence, row operations can be performed on A, to convert it to a form where it is convenient to solve for the unknowns.
Echelon form is obtained when the operations are performed such that-

- Pivot elements must be non-zero
- All elements below the pivot elements must be zero

Breakdown of elimination -
Non-singular case: Can be avoided by swapping the rows
Signular case: Can't be avoided

Cost of elimination -
Left side (AX) : number of operations = $(n^3 - n)/3$
Can be approximated to $n^3/3$ when n is sufficiently large
Right side (b) : number of operations = n^2

2 Lengths and Dot Products

Dot product / Inner product
Consider 2 vectors v and u, $v = [v_1 \ v_2 \ .... \ v_n] and w = [w_1 \ w_2 \ ... \ w_n]$
The dot product is given by,

$$v.w = v_1w_1 + v_2w_2 + ... + v_nw_n$$

In matrix notation, the dot product is represented as $v^T w$.

If dot product of 2 vectors result in zero, then the vectors are perpendicular or orthogonal.

Length of vector
It is the square root of dot product of a vector with itself.
Also called as magnitude of the vector.
$||v|| = \sqrt{(v.v)} = \sqrt{v^T v}$

Unit vector is any vector whose length is 1.

Angle between two vectors (T)
$cosT = (v.w)/(||v|| \; ||w||)$

Properties of dot product
u, v and w are 3 vectors and c is a scalar

- $u.v = v.u$
- $u.(v + w) = u.v + u.w$
- $u.(v.w) = (u.v).w$
- $(cv).u = c(u.v)$
- Schwarz Inequality :- $v.w \leq ||v||||w||$
- Triangle Inequality :- $||v + w|| \leq ||v|| + ||w||$
- If $u = v + w$, then $||u|| = ||v + w|| = \sqrt{(||v|| + ||w|| + 2(v.w))}$

3 Matrix Multiplication

A is mxn matrix and B is pxq matrix
For C = AB to exist, n = p must be satisfied.
Dimensions of C will be mxq.

Note that every row/column operation can be represented in terms of matrix multiplication.

- Column operation corresponds to post-multiplication
- Row operation corresponds to pre-multiplication

Properties of matrix multiplication

- Associative Law : A(BC) = (AB)C
- Distributive Law: A(B+C) = AB + AC

Commutative Law is not always obeyed!!
(AB $\neq$ BA ; barring some exceptions)

Different ways of performing matrix multiplication

1. Dot product of Rows of A with Columns of B
   A = $\begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$ B = $\begin{pmatrix} 4 & 0 \\ -1 & 1 \end{pmatrix}$
   C = AB = $\begin{pmatrix} 1 \times 4 + 3 \times -1 & 1 \times 0 + 3 \times 1 \\ 2 \times 4 - 1 \times -1 & 2 \times 0 - 1 \times 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 9 & -1 \end{pmatrix}$

2. By columns: A = $\begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$
   B = $\begin{pmatrix} 4 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 \end{pmatrix}$ where $b_1 = \begin{pmatrix} 4 \\ -1 \end{pmatrix}$ and $b_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
   C = AB = A$\begin{pmatrix} b_1 & b_2 \end{pmatrix} = \begin{pmatrix} Ab_1 & Ab_2 \end{pmatrix}$

3. By rows: A = $\begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ where $a_1 = \begin{pmatrix} 1 & 3 \end{pmatrix}$ and $a_2 = \begin{pmatrix} 2 & -1 \end{pmatrix}$
   B = $\begin{pmatrix} 4 & 0 \\ -1 & 1 \end{pmatrix}$
   C = AB = $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$B = $\begin{pmatrix} a_1 B \\ a_2 B \end{pmatrix}$

4. Block multiplication:
   Matrices are split as smaller matrices called blocks and multiplied individually as shown. The blocks must be compatible for multiplication.

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \quad B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}$$

$$C = AB = \begin{pmatrix} A_1B_1 + A_2B_3 & A_1B_2 + A_2B_4 \\ A_3B_1 + A_4B_3 & A_3B_2 + A_4B_4 \end{pmatrix}$$

Cost of multiplication
Number of additions = $m(n-1)q$
Number of multiplications = mnq

4 Inverse of a Matrix

If A is a non-singular square matrix, then $A^{-1}A = AA^{-1} = I$
(where I is identity matrix).
If A is singular, then it is non-invertible and hence A^{-1} does not exist.
This is because, if A is singular, then there exists a non-zero vector X such that AX = 0, which means inverse of A can not be obtained.

4.1 Gauss Jordan method to find inverse of a matrix

- Obtain augmented matrix [A:I]
- Perform Gaussian elimination on A to convert it to I
- Perform the exact same operations on I, it will be converted to A^{-1}

Why does this work?
All the row operations performed on A can be represented by a pre-multiplication matrix Q. Hence, $Q[A : I]$ is to be obtained.
Since $QA = I$ and $QI = Q$, it is obvious that $Q = A^{-1}$.

Inverse of product of matrices
$(AB)^{-1} = B^{-1}A^{-1}$
$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

5 Transpose of a Matrix

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

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

Properties of Transpose operation

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

6 L-U Decomposition

Expressing a matrix A as the product of Lower Triangular Matrix (L) and Upper Triangular Matrix (U).
For L, all elements above diagonal are 0 and for U, all elements below diagonal are 0.
This is done by using Gaussian Elimination and finding inverse of the operation matrix.
Assuming no row exchanges are necessary, A = LU is obtained by the following steps.

- Perform QA = U where Q represents row operations to get U.
- Find L = Q^{-1} and multiply L on both sides to get A = LU.

Further decomposition A = LDU can also be obtained (D is diagonal matrix i.e all elements except principal diagonal elements are 0).

If A = LDU and $A = L_1 D_1 U_1$, then $L = L_1$, $D = D_1$ and $U = U_1$. Meaning, LDU decomposition is unique for a given matrix A.

Note that if row operation necessary is $R_i = R_i - aR_j$, then it's equivalent inverse operation will be $R_i = R_i + aR_j$.

6.1 Permutation Matrix

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 $P^{-1} = P^T$

Example for 3x3 permutation matrix - P = $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

If row exchanges are necessary for LU decomposition, then the exchange is done using permutation matrix. Hence, PA = LU (P is the required permutation matrix)

7 Vector Space

R^n is a vector space if all vectors in R are closed under scalar multiplication, addition and both.
Basically means that a linear combination of any vectors in the space should also be part of the same space.
Meaning, if X and Y belong to R, then aX, bY, X + Y, aX + bY should all belong to R.

For example,
R^2 - The entire X-Y plane
R^3 - The entire X-Y-Z 3D space

7.1 Subspace

A part of a vector space which satisfies the same criteria.

Subspaces in R^2:

1. X-Y plane
2. Any line that passes through origin (0,0)
3. Only the zero vector (0,0)

Subspaces in R^3:

1. X-Y-Z 3D space
2. Any plane that pass through origin (0,0,0)
3. Any line that passes through origin (0,0,0)
4. Only the zero vector (0,0,0)

Any vector space will always contain the origin i.e the zero vector!

If P and Q are two subspaces,

- P $\cup$ Q does not give a subspace (unless P and Q lie on the same plane)
- P $\cap$ Q always gives a subspace (in worst case scenario, the intersection is just the zero vector, which is a subspace)

7.2 The Column Space, C(A)

C(A) consists of all linear combinations of the columns of a matrix A.
For example, take a system of 'm' linear equations and 'n' unknowns where n<m, represented by A(mxn) X(nx1) = b(mx1).
This system can be solved uniquely if and only if the vector b lies in the column space of A.

Linear Independence
The columns of a matrix are linearly independent if none of those columns

can be expressed as a linear combination of other columns.
If a column can be expressed as a linear combination of other columns, then they are linearly dependent.
Dependent columns are redundant and contribute nothing new to the column space.
The number of linearly independent columns in the matrix A will give the dimension of the column space.
Hence, if A(mxn) has p linearly independent columns, then the column space is a 'p' dimensional subspace of R^n.

7.3 The Null Space, N(A)

N(A) consists of all solutions X for the system of equations AX = 0.
The null space is an 'n-p' dimensional subspace of R^n.
It is also called as Nullity of the matrix.

8 Rank

Rank of a matrix is the number of pivot elements in the row echelon form of a matrix.
Rank will be equal to the number of linearly independent vectors in the matrix.
For a matrix A(mxn) with rank r, $r \leq min(m, n)$

8.1 Solution to AX = 0

Half-way method

- Obtain row reduced echelon form of the matrix A, named U.
- Identify the columns containing pivots and the ones not containing pivots (called free columns).
- In the vector X, choose any values (generally some combinations of 0's and 1's) for the elements corresponding to the free columns.

- Solve the equation UX = 0 for the elements corresponding to pivot columns.

Fully reduced method

- Perform elimination further to obtain row reduced echelon form of the matrix A, named R.
- Identify the columns containing pivots and the ones not containing pivots (called free columns).
- Write the matrix R in the form $\begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$ in block matrix representation, which can be obtained by choosing proper values for free columns.
- Here, I = Identity matrix representing Pivot columns, F = Matrix of elements corresponding to free columns
- Solve RX = 0; X = $\begin{pmatrix} -F \\ I \end{pmatrix}$

  The null space matrix (or vector) X_n will be c$\begin{pmatrix} -F \\ I \end{pmatrix}$ i.e it describes all the solutions for AX = 0

8.2 Solution to AX = b

- Perform Gaussian Elimination on augmented matrix [A:b].
- This will either give conditions on b for existence of the solution or the solution directly.
- In case solution does not come directly, substitute 0 for free variables and solve for pivot variables to get the particular solution X_p.

8.3 Complete Solution

$A\ X_n\ =\ 0 \textit{ and } A\ X_p\ =\ b$
$A\ (X_n\ +\ X_p)\ =\ b$
$X\ =\ X_n\ +\ X_p$
Note that X will not form a subspace (since X is a shifted version of X_n, so it does not contain the zero vector).

8.4 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.

More formal definition of Independence

A set of vectors are linearly independent if there exists no linear combination of said vectors which can result in 0 (expect for all 0 combinations).

A set of vectors are linearly dependent if there exists at least one linear combination of said vectors which can result in 0.

If the zero vector is one of the vectors in consideration, then the vectors are always dependent.
In a space of n dimensions, any n+k number of vectors will be dependent. (k is positive integer)

If the vectors under consideration are the columns of a matrix, then the vectors are linearly independent if:

- The rank of the matrix is equal to min(m,n).
- The matrix is non-singular i.e invertible.
- The null space consists only of the zero vector. (otherwise the vectors are linearly dependent)

9 Basis and Dimension

A set of vectors is said to span a space if all combinations of the said vectors exist in the space.

The **Basis of a vector space** is the set of vectors which are linearly independent and span the vector space.
Meaning, the basis will give the minimum set of vectors that are required to fully fill the vector space upon linear combinations.

Given a vector space, it can have multiple basis. But the number of vectors in the basis will always be the same.
The number of vectors in a basis of a vector space is called it's **Dimension**.

The rank of a matrix A is the dimension of the columns space of A.
r = dim[C(A)]
Dimension of null space is the difference between number of columns of A and the rank of A.
n - r = dim[N(A)]

Rank-Nullity Theorem
Dim[C(A)] + Dim[N(A)] = n (or rank + nullity = number of columns)

The dimension of a subspace that has only the zero vector is 0, hence it has no basis.

9.1 The 4 fundamental subspaces

A is (mxn) matrix, rank r;

1. **Column Space of A**: in R^m
   The combinations of linearly independent columns.
   Dimension = r
   Basis = Pivot columns in row reduced echelon form of A

2. **Row Space of A or Column Space of A^T**: in R^n
   The combinations of linearly independent rows.
   Dimension = r
   Basis = First r rows in row reduced echelon form of A

3. **Null Space of A**: in R^n
   All solutions to AX = 0
   Dimension = n-r
   Basis = Special solutions i.e combination of columns that gives 0

4. **Null Space of A^T or Left Null Space of A**: in R^m
   All solutions to A^Tx = 0
   Dimension = m-r
   Basis = Special solutions i.e combination of rows that gives 0

Note that row operation will change the column space but will preserve the row space and vice versa.

Dim[$C(A^T)$] + Dim[$N(A^T)$] = m (or rank + left nullity = number of rows)

10 Matrix Spaces

A group of matrices can be treated as a space if they can be multiplied with scalars and added such that the result will still be in the same space.
Examples - All 3x3 matrices, All 4x4 symmetric matrices, All 2x2 diagonal matrices, etc..
Take 3x3 as the standard/general case.

- Space of all matrices, M:
  Basis = matrices with one elements as 1 and the rest 0
  Dimension = 9
- Space of all symmetric matrices, S:
  Dimension = 6
- Space of all upper triangular matrices, U:
  Dimension = 6
- Space of all lower triangular matrices, L:
  Dimension = 6
- Space of all diagonal matrices, D:
  Dimension = 3

Note that, $S \cap U = D and S + U = M$.
Hence, if A and B are two different matrix spaces, then $Dim(A)+Dim(B) = Dim(A+B) + Dim(A \cap B)$.

Rank 1 matrices
A is (mxn) matrix with rank, r = 1.
A can always be expressed as, $A = u\ v^T$ where u is column vector and v^T is row vector.
For example, A = $\begin{bmatrix} 1 & 2 \\ -1 & -2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$
Also, any matrix (with rank > 1) can be expresses as a sum of rank 1 matrices.

$$B = \begin{bmatrix} 1 & 2 \\ -1 & -2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}$$

11 Orthogonality

Orthogonal = Perpendicular

If x and y are two vectors such that $x^T\ y = 0$ (i.e x.y = 0), they are called **'Orthogonal Vectors'**.
In this case, $||x||^2 + ||y||^2 = ||x + y||^2$

The zero vector is orthogonal to every other vector.

Subspaces S and T are orthogonal if every vector in S is orthogonal to every vector in T.

For two subspaces to be orthogonal, they should not meet at any point apart from the origin. Meaning, the two subspaces must have nothing in common except for the zero vector.

Orthogonal subspaces S and T are Orthogonal Complements if the sum of their dimensions is equal to the dimension of the whole vector space.

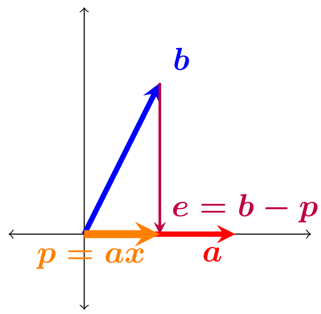
Row Space and Null Space are orthogonal complements in R^n
Column Space and Left Null Space are orthogonal complements in R^m

If a bunch of vectors are orthogonal, then they are also independent.
If a bunch of unit vectors are orthogonal, they are called **"Orthonormal vectors"**.

Therefore, the vectors q_i and q_j are orthonormal vectors if,

$$q_i^T\ q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

11.1 Projections



p is the projection of b on a; meaning, p is the closest point from b to a.
$e = b - p$
$p = ax$
$e = b - ax$
Since a is perpendicular to e, $a.e = 0$
$a.(b - ax) = 0$ $i.e$ $a^T (b - ax) = 0$
Hence, $x = \frac{a^T b}{a^T a} \implies p = \frac{a\ (a^T b)}{a^T a} = Pb$
where $P = \frac{aa^T}{a^T a}$ is the **Projection Matrix**.

Column space, C(P) = line through A
Rank of P, r = 1
$P^T = P$ (P is symmetric)
$P^2 = P$ (Projecting on the same vector twice will be same as projecting once)

11.2 "Solving" Ax = b when it is unsolvable!

Since AX = b is unsolvable i.e b is not in the column space of A, the best solution is obtained by solving for AY = p where p is the projection of b onto the column space of A.
Using the concept of projection, find Y using p = AY.
b - AY is perpendicular to the matrix A.
$A^T (b - AY) = 0$ where $b - AY = e$.
Meaning, e is in the null space of A^T and hence, it is perpendicular to the columns space of A.
Hence, $A^T A Y = A^T b \implies Y = (A^T A)^{-1} A^T b$
$\therefore p = A (A^T A)^{-1} A^T b = P b$

where $P \;=\; A\,(A^T\;A)^{-1}\;A^T$ is the Projection Matrix.

Note that A is not a square matrix!
If A was a square invertible matrix, then the projection matrix would simply be the identity matrix.

- If b is in the column space of A, then Pb = b
- If b is perpendicular to A, then Pb = 0

11.2.1 Method of Least Squares

The square of the error term (e) is minimized in order to get the best possible solution.
AX = b is not solvable since b is not in the column space of A.
AY = p is solvable since p is the projection of b onto the column space of A.

The method of least squares will find the best possible values of p; which is obtained by solving $(A^TA)Y = A^Tb$.
This can be derived from calculus by partially differentiating the sum of squares of error terms and equating them to 0.

- A is usually a rectangular matrix, but A^TA is always a square symmetric matrix.
- Null Space of A = Null Space of A^TA
- Rank of A = Rank of A^TA
- A^TA is invertible if and only if the columns of A are linearly independent.

11.3 Orthogonal Matrices

If $q_1,\; q_2, ...,\; q_n$ are orthonormal vectors and Q is a matrix such that $Q \;=\; [q_1\; q_2\; ....\; q_n]$ (the vectors are it's columns), then $Q^TQ = I$.
If such as matrix Q is a square matrix, is called **Orthogonal matrix**.
Note that the $Q^TQ = I$ will be true even if Q is a rectangular matrix.

For an orthogonal matrix, $Q^TQ = Q^{-1}Q$
$\therefore Q^T = Q^{-1}$

An orthogonal matrix will make computations easier while finding projection matrix and while solving using method of least squares.
$P = Q(Q^TQ)^{-1}Q^T \implies P = QQ^T$
$(Q^TQ)Y = Q^Tb \implies Y = Q^Tb$

Multiplying a vector X with an Orthogonal Matrix Q will not changed the length of X.
$||QX||^2 = (QX)^TQX = X^TQ^TQX = X^TX = ||X||^2$

11.3.1 Examples for Orthogonal Matrices

Rotation Matrix

Q $= \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$

Reflection Matrix

Q $= \begin{bmatrix} cos\theta & sin\theta \\ sin\theta & -cos\theta \end{bmatrix}$

Householder Reflections
$H = I - 2uu^T$ where $uu^T = 1$
$H^TH = (I - 2uu^T)^T(I - 2uu^T) = ||(I - 2uu^T)||^2 = I$

Hadmard Matrices
$H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
$H_4 = \frac{1}{2} \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}$
and so on...
Hadmard Matrices exist for all square matrices of dimensions 4n (n is a natural number).

Other examples are Fourier Matrices and Wavelet Matrices which will be dealt with in more detail later.

11.4 Gram Schmit Orthogonalization

Process in which the columns of a matrix are made orthonormal (where the columns are independent) i.e to make the basis of the matrix orthonormal.
General process flow

- a,b,c... are the independent columns/vectors
- Find best possible A,B,C.... from the original vectors such that they are orthogonal
- Divide each of them with their length to convert to unit vectors, hence making them orthonormal

Finding orthogonal vectors from independent vectors.
Let a,b and c be 3 independent vectors.

- $A = a$
- $B = b - ax = b - \frac{A^T b}{A^T A} A$
- $C = c - ay - bz = c - \frac{A^T c}{A^T A} A - \frac{B^T b}{B^T B} B$
  (This process can be continued for more vectors)

The above process is removing the projection components of the vector onto the vectors it has to be orthogonal to. It's same as finding the error term (e).
Finally, to find orthonormal vectors q1, q2 and q3.
$q1 = \frac{A}{||A||}; \quad q2 = \frac{B}{||B||}; \quad q3 = \frac{C}{||C||}$

The column space of the original matrix will be same as the column space of the orthogonal matrix Q!

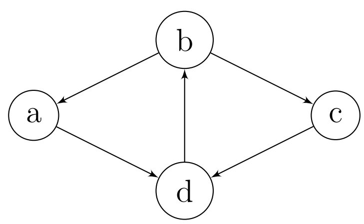
Since Gram Schmit orthogonalization is converting matrix A to Q, it can be expressed in terms of matrix multiplication.
A = QR where R will turn be an Upper Triangular Matrix.
The reason becomes evident on reviewing how orthogonal vectors are found from the independent vectors.

12 Graph Theory

Graph is a combination of Nodes and Edges.
Matrix representation of a graph is called **Incidence Matrix**, A (mxn) where n = number of nodes and m = number of edges.
For an incident matrix, rank r = n-1 (because 1 node is considered as ground).

Each edge will have a direction.
While creating the incidence matrix, an edge leaving a node will correspond to -1 and an edge entering a node will correspond to +1.

Loops are parts of the graph that start and end at the same node.
Loops correspond to linearly dependent rows in the incidence matrix.



In the above graph, a, b, c and d are nodes; ba (e1), ad (e2), db (e3), bc (e4) and cd (e5) are edges; bcd and bad are loops.

The corresponding incidence matrix, A = $\begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & -1 \end{pmatrix}$

A graph which has no loops is called as a Tree.
A tree will have only linearly independent rows.

Dim[$C(A^T)$] (or the rank of the incident matrix, r) will give the number of linearly independent rows.
Dim[$N(A^T)$] will give the number of loops (since Dim[$N(A^T)$] = m - r).
From the above facts, loops (l) = edges (m) - nodes (n) - 1.
l = m - n + 1

For the above example, m = 5; n = 4; l = 2.

13 Determinants

The determinant is a number associated with a square matrix.

13.1 Properties of determinants

1. Determinant of the identity matrix is 1. $|I| = 1$
2. 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.
3. If any single row/column is multiplied by a constant, the determinant also gets multiplied by the constant.
4. 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.
5. Row/column operations (i.e adding constant times one from another) will not change the determinant.
6. If a row/column of zeros exists in the matrix, then it's determinant is equal to 0.
7. If the same row/column is repeated in the matrix, then it's determinant is equal to 0.
8. If the rows/columns of a matrix are linearly dependent, then it's determinant is equal to 0.
9. 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.

10. Determinant of AB is equal to product of determinant of A and determinant of B. $|AB| = |A||B|$

11. Determinant of inverse of A is equal to reciprocal of determinant of A. $|A^{-1}| = 1/A$

12. Determinant of a scalar (k) times A (where A is nxn matrix) is equal to k^n times determinant of A. $|kA| = k^n|A|$

13. Determinant of transpose of A is same as determinant of A. $|A^T| = |A|$

13.2 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. i.e $adj(A) = C^T$

13.3 Formula for determinant

2x2 matrices : A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$
$|A| = ad - bc$

3x3 matrices : A = $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$
$|A| = a(ei - fh) - b(di - fg) + c(dh - eg) = aei + bfg + cdh - afh - bdi - cdh$

General procedure for finding determinant

- Choose any single row or column.
- Multiply the elements of the chosen row/column with the co-factors of the respective elements.
- Add all the values obtained algebraically to find the determinant.

$\therefore |A| = a_{11}C_{11} + a_{12}C_{12} + ... + a_{1n}C_{1n}$
(if chosen to calculate along the 1st row)

13.4 Applications of Determinants

1. **To find inverse of a matrix**
   General formula to find the inverse of a square matrix A is given by, $A^{-1} = \frac{adj(A)}{|A|}$
2. **Cramer's Rule**
   Since $AX = b, X = A^{-1}b = \frac{C^T\ b}{|A|}$
   If $X = [x_1, x_2, ..., x_n]$, then
   $xi = \frac{|Bi|}{|A|}$ where B_i is the matrix A with i_{th} column replaced with b.
3. **Geometric applications (Volume and Area)**
   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.

14 Eigen Values and Eigen Vectors

If A is a square matrix and X represents vector(s), then X is an Eigen Vector of A if AX will be in the same direction as X (i.e $AX||X$).
This means, $AX = \lambda X$ where λ is a scalar called the Eigen Value of A.

If A is nxn matrix, then it will have 'n' Eigen Vectors and corresponding Eigen Values (which can be distinct or same).

If A is singular, then AX = 0, meaning the Eigen Vector X is in the nullspace of A, meaning $\lambda = 0$.

If A is projection matrix, then the Eigen Values of A will be either 1 or 0 (1 for X in the column space and 0 for X in the null space).

If A is permutation matrix, then the Eigen Values of A will be either 1 or -1.

- Sum of Eigen Values of a matrix is equal to the Trace of the matrix. (Trace is the sum of main diagonal elements) $\Sigma\lambda = \Sigma A_{ii}$
- Product of Eigen Values of a matrix is equal to the Determinant of the matrix. $\Pi\lambda = |A|$

Rewriting the equation, $(A - \lambda I)X = 0$ is obtained.
A - λI is a Singular Matrix, hence it's Determinant is 0.

Characteristic Equation of a square matrix : $|A - \lambda I| = 0$

For a 2x2 matrix, the Characteristic Equation will be,

$$\lambda^2 - trace(A)\ \lambda + det(A) = 0$$

For a 3x3 matrix, the Characteristic Equation will be,

$$\lambda^3 - trace(A)\ \lambda^2 + (C_{11} + C_{22} + C_{33})\ \lambda - det(A) = 0$$

Eigen Values of rotation matrices or anti-symmetric matrices will be complex! Complex eigen values exist as conjugate pairs. Meaning, there can't be odd number of complex eigen values.

Eigen Values of symmetric matrices will always be real.

Eigen Values of some matrices can be repeated. Such matrices will have shortage of linearly independent Eigen Vectors.
Meaning, Eigen Vectors of distinct Eigen Values are Linearly Independent.

For an nxn matrix A, if the number of distinct eigen values is less than n, it is called a Degenerate Matrix.

Properties of Eigen Values
If AX = λX, then

- $(A + kI)X = (\lambda + k)X$
- $(A^{-1})X = X/\lambda$
- $(A^n)X = \lambda^n X$

14.1 Diagonization

Suppose matrix A has n independent eigen vectors $(x_1, x_2, ...x_n)$ and corresponding eigen values $(\lambda_1, \lambda_2, ...\lambda_n)$.
Matrix S is such that it's columns are the eigen vectors i.e S is the Eigen Vector Matrix.

AS = A$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{bmatrix}$
$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ = SΛ
where Λ is the diagonal eigen value matrix

$$\therefore A = S\Lambda S^{-1} \quad or \quad S^{-1}AS = \Lambda$$

Note that S^{-1} exists only if the columns of S are linearly independent, which is why the eigen vectors of A have to be independent.

Matrix A is sure to be diagonalizable (i.e have independent eigen vectors) if all it's eigen values are distinct.

If eigen values are repeated, then matrix A may or may not have independent eigen vectors.

For triangular matrices and diagonal matrices, the eigen values are the main diagonal elements, or the **principal diagonal elements**.

14.2 Solution for Difference Equations

Consider the difference equation system $U_{k+1} = AU_k$ with U_0 being the initial conditions vector.
$U_1 = AU_0$, $U_2 = AU_1 = A^2U_0$, ... $U_k = AU_{k-1} = A^kU_0$.

The n eigen vectors of A must be unique, meaning they must form a basis for R^n. Hence, U_0 can be expressed as a combination of eigen vectors of A.
$U_0 = c_1x_1 + c_2x_2 + ... + c_nx_n$

$\implies U_k = A^k\ (c_1x_1 + c_2x_2 + ... + c_nx_n) = \begin{bmatrix} c_1\lambda_1^k x_1 & c_2\lambda_2^k x_2 & \cdots & c_n\lambda_n^k x_n \end{bmatrix}$
$\therefore U_k = S\Lambda^k c$

14.3 Solution for Differential Equations

Consider a first order differential equation with constant coefficient matrix, u'(t) = Au(t) where initial conditions is given by vector u(0).

General solution is give by,
$u(t) = c_1e^{\lambda_1 t}x_1 + c_2e^{\lambda_2 t}x_2 + ... + c_ne^{\lambda_1 n}x_n$
$\therefore u(t) = Se^{\Lambda t}c$

The initial conditions can be plugged in to evaluate the values of constants to find particular solution.

14.4 Stability Analysis using Eigen Values

Difference Equations

- Magnitudes of all Eigen Values are less than 1 $\implies$ Stable System since steady state output dies down to 0.

- One Eigen Value has magnitude 1 and rest have magnitude less than 1 $\implies$ Steady state output will be some constant value.
- One or more Eigen Values have magnitude greater than 1 $\implies$ Unstable System since steady state value blows up i.e never stops growing.

Differential Equations

- All Eigen Values are negative $\implies$ Stable System since steady state output dies down to 0.
- One Eigen Value is 0 and the rest are negative $\implies$ Steady state output will be some constant value.
- One or more Eigen Values are positive $\implies$ Unstable System since steady state value blows up i.e never stops growing.

If Eigen Values are complex, perform same analysis on the real part.

14.5 State Transition Matrix

For u'= Au, if u is expressed in terms of the eigen vector matrix as u = Sv, then $Sv' = ASv, \implies v' = S^{-1}ASv = \Lambda v$

This way, if the eigen vectors are chosen as the basis, solving the differential equation will be simpler.

Solution : $v(t) = e^{\Lambda t}v(0)u(t) = Se^{\Lambda t}S^{-1}u(0)$, where $e^{At} = Se^{\Lambda t}S^{-1}$
$\therefore u(t) = e^{At}u(0)$

Result that will make calculations more efficient: $e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix}$

Note that n^{th} order differential equation can be converted to n 1^{st} order differential equations (i.e nxn coefficient matrix).

14.6 Markov Matrix

Markov matrices have the following features

- All entries are greater than or equal to 0
- All columns add to 1

Markov matrices have one of the eigen value equal to 1.
This is because A - I (i.e $\lambda = 1$) will have linearly dependent columns since the sum of columns add up to 0, meaning $|A - I| = 0$ or A - I is singular.

All other eigen values have magnitude lesser than or equal to 1.

The eigen vector corresponding to the unit eigen value will have all positive terms (or 0) and it gives the steady state value.

$U_{k+1} = AU_k$ where A is a Markov Matrix is used in probability. Steady state value of U will give final probabilities at which the process settles down to.

15 Symmetric Matrices

A symmetric matrix has **Real Eigen Values** and **Orthogonal Eigen Vectors**.
Note that the matrices being dealt with here are real matrices.

Diagonalization of a symmetric matrix A will lead to $A = S\Lambda S^{-1}$ where S is formed out of orthonormal vectors, hence it can be represented as Q.
Since vectors of Q are orthonormal, $Q^{-1} = Q^T$
$\therefore A = Q\Lambda Q^T$
This is called the 'Principle Axis Theorem'.
The eigen vectors tell the direction of the principle axes and the eigen values tell the magnitude of the axes in their respective directions.

Every symmetric matrix is a combination of perpendicular projection matrices.

The signs of pivots of a symmetric matrix are same as signs of eigen values.

- number of pivots lesser than 0 = number of eigen values lesser than 0
- number of pivots greater than 0 = number of eigen values greater than 0

If a matrix a satisfies $A^T = -A$, then it is called 'Skew-symmetric matrix' and it has imaginary Eigen values.

Complex Cases

If the matrix A is complex, then it is called 'Hermitian matrix' if it satisfies $\bar{A}^T = A$ or $A^H = A$ where 'H' is the Hermitian operator which means taking transpose of conjugate.

For a real vector x, the length is $x^T x$, but for a complex vector x, the length will be $x^H x$.

For a complex matrix to be Orthogonal, it must satisfy $A^H = A^{-1}$.

Hermitian matrices will have real eigen values, Skew-Hermitian matrices will have imaginary eigen vales.

15.1 The Fourier Matrix

The Fourier matrix is a complex square matrix used to compute Discrete Fourier Transform (DFT) of a given signal.
Discrete signal of n elements (treated as nx1 vector) multiplied with nxn Fourier matrix will give n-point DFT of the discrete signal.

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^n \\ 1 & w^2 & w^4 & \cdots & w^{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)(n-1)} \end{bmatrix} \text{ where } w = e^{i(2\pi/n)}$$

The columns of the Fourier matrix are orthogonal. Hence, the Fourier matrix is an orthogonal matrix i.e $F^H = F^{-1}$ (makes finding Inverse DFT easier).

Fast Fourier Transform (FFT) is a matrix technique to reduce the number of computations required to find the n-point DFT of a signal.

Note that direct multiplication of a discrete signal with the Fourier matrix needs n^2 multiplications.

Larger order Fourier matrices are connected to a smaller order matrices as-
$F_n = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} \frac{F_n}{2} & 0 \\ 0 & \frac{F_n}{2} \end{bmatrix}$ P
where D is a diagonal matrix block and P is odd-even permutation matrix.

Exploiting the structures of these matrices, this operation be done in $\frac{n}{2}log_2n$ multiplications, which is significantly less than before.

16 Positive Definiteness

Positive definite matrix is a symmetric matrix where

- All eigen values are positive
- All pivots are positive
- All sub-determinants are positive

Test for positive definiteness of a matrix A: $x^TAx > 0$ for all vectors x (except at 0 vector)

The function $f(x) = x^TAx$ of a positive definite matrix has a minima at the origin i.e 0 vector.
This is because if A is positive definite, then $f(x) = x^TAx$ can be expressed as a sum of squares, which is always positive.
If A was not positive definite, then $f(x) = x^TAx$ at the 0 vector is a saddle point (minima in some ways and maxima in some ways), meaning the function is not always positive.

If A is positive definite, then A^T (same as A^{-1}) is also positive definite since $1/\lambda$ is always positive given λ is positive.

If A and B are positive definite, then A+B is also positive definite.

For a rectangular matrix A, then A^TA is always positive definite.

17 Similar Matrices

A and B are similar matrices if there exists a matrix M such that $B = M^{-1}AM$ Hence, $S^{-1}AS = \Lambda$ implies that the matrix A and it's diagonal eigen value matrix are similar.

If$Ax = \lambda x$, then $AMM^{-1}x = \lambda x$
$\implies$ $(M^{-1}AM)M^{-1}x = \lambda M^{-1}x$ i.e $Bx_1 = \lambda x_1$
$\therefore$ λ is also an eigen value of B.

Similar matrices have the same eigen values. Meaning, the matrices that have the same eigen values are similar.
Similar matrices also have the same number of eigen vectors.

The only exception here is when all the eigen values are equal.
For example, if for a 2x2 matrix, the eigen values are 3 and 3, then there are 2 families of similar matrices.
One family consists only of 3I i.e $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$ and the other family consists of all other matrices that satisfy similarity.
This is because if A = 3I, then $(M^{-1}AM) = A$ always and no other matrix B can satisfy the similarity condition.

17.1 Jordan Form

The simplest non-diagonal matrix that satisfies similarity for a given matrix A is called it's Jordan Form.
For the above example, the Jordon Form would be $\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$

Jordon's Theorem
Every matrix A is similar to a Jordan matrix J, where J consists of Jordan blocks as diagonal blocks.
$\implies$ number of Jordan blocks in J = number of eigen vectors of A

In case there are no repeated eigen values, then the Jordon matrix is simply the eigen value diagonal matrix (each diagonal element is a Jordon block).

In case there are repeated eigen values, then the Jordan blocks are matrices

consisting of eigen values on it's diagonals, 1's directly above the diagonals and 0's everywhere else.
Meaning, for an eigenvalue λ_i repeated r times, the Jordan block is an rxr matrix with λ_i on the diagonals,and 1's on the above diagonal with rest of the entries 0.

18 Singular Value Decomposition

The SVD is a process that decomposes any matrix (rectangular matrices included) into a product of an orthogonal matrix, a diagonal matrix and another orthogonal matrix.
For any matrix A of rank r, there exists orthonormal vectors $v_1, v_2, ..v_r$ such that
$Av_1 = \sigma_1 u_1$, $Av_2 = \sigma_2 u_2$, .. $Av_r = \sigma_r u_r$ where $u_1, u_2, .., u_r$ are also orthonormal vectors. Here, $\sigma_1, \sigma_2, .., \sigma_r$ are called 'Singular Values' of A.
$\implies AV = U\Sigma$
$\therefore A = U\Sigma V^T$
where U and V are orthogonal matrices and Σ is diagonal matrix consisting of singular values.

- V is an orthonormal basis for the row space of A.
- U is an orthonormal basis for the column space of A.

Note that there will be 0's in the diagonal singular value matrix which will correspond to the null spaces of A and A^T.

Singular Values and Eigen Values

Consider the matrix A^TA, which is always square, symmetric and positive definite. Using SVD, $A^TA = (U\Sigma V^T)^T U\Sigma V^T \implies A^TA = V\Sigma U^T U\Sigma V^T$
$\therefore A^TA = V\Sigma^2 V^T$
Similarly, it can be deduced that $AA^T = U\Sigma^2 U^T$
This means that the eigen values of A^TA and AA^T are given by the diagonal matrix Σ^2.

If complex numbers are involved, then perform Hermitian operation (conjugate transpose) instead of just transpose.

19 Linear Transformations

A transformation is mapping of all vectors in a subspace to some other subspace.
Example, $T : R^3 \rightarrow R^2$ projects every vector in 3D space to 2D space.

A transformation is linear when the following 2 properties are satisfied

- T(u + w) = T(u) + T(w)
- T(cv) = c T(v)

In a linear trasformation, the 0 vector always maps to itself [i.e T(0) = 0]

Any linear transformation can be represented by matrix multiplication.
T(v) = A v where A is the transformation matrix. Consider the example $T : R^3 \rightarrow R^2$ and let A represent the corresponding transformation matrix. Input will be a 3x1 vector, output will be a 2x1 vector and hence A should be 2x3 matrix.

The information needed to know what the linear transformation does to any vector in the input space is the linear transformation output for the basis of the input space.
Meaning, by knowing what the transformation does to every vector in the input basis, using linear combinations, the output for any input in the input space can be found.

The basis of the input space will give the co-ordinate system and their linear combination gives the required co-ordiantes of an arbitrary input.
The vector $\begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$ in 3D can be represented as $2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ where $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ represent the x,y and z axes of cartesian co-ordinate system respectively and (2,3,5) are the co-ordinates of the vector.

The matrix A multiplies the input co-ordinates and gives the output co-ordinates. For example, if the above vector was to be projected into 2D space with x and y axes only, then the output would be $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ And the matrix that does that is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

Note that the standard basis is not always the best choice of basis. It is better to choose eigen vectors of the projection as the basis.

Also, the inverse of the transformation matrix if exists, will give inverse transformation. Meaning, if A is singular, then the transformation is not invertible.

19.1 Change of Basis

The basis of a subspace can be changed simply by performing the necessary linear transformation on the basis.
If V is the basis for a given subspace, U = AV will give a different basis for the same subspace.
Suppose a linear transformation T is defined for 2 different basis V and W such that T(V) = AV and T(W) = BW.
Here, A and B will be similar matrices.
Eigen vector basis is the best basis because for the eigen vector matrix S, T(S) = ΛS where Λ is the daigonal eigen value matrix. Meaning, the change of basis matrix is a diagonal matrix.

Image Compression
Change of basis is used extensively for image compression. It is not efficient to represent an image using the standard basis.
Hence the basis is usually changed to either the Fourier basis or the Wavelet basis.

20 Left, Right and Pseudo-Inverses

The 2 sided inverse of a square matrix A is the usual inverse i.e where $A^{-1}A = AA^{-1} = I$. This inverse exists only when A is full rank matrix i.e rank, r = m = n.
The cases where A is not a full rank matrix, but is either full row rank matrix or full column rank matrix will not have 2 sided inverses.

A is full column rank matrix; $r = n < m$
The null space is only 0, meaning there is either 1 solution or 0 solutions.
$A^T A$ is full rank and invertible.
$\implies$ **Left inverse** : $(A^T A)^{-1} A^T$ because $(A^T A)^{-1} A^T A = I$
Note that multiplying the left inverse from the right will give A $(A^T A)^{-1} A^T$ which is the projection matrix onto the row space!

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 $AA^T (AA^T)^{-1} = I$
Note that multiplying the right inverse from the left will give $A^T (AA^T)^{-1} A$ which is the projection matrix onto the column space!

A is neither full row rank nor full column rank matrix.
Even if a matrix A is not invertible, all the vectors in the row space will have a 1-1 mapping with all the vectors in the column space.
Meaning, by taking any vector in the row space of A and multiplying it with A, a unique vector in the column space is obtained.
Hence, the row space and column space share an invertible relationship.
The matrix that undoes the mapping from the row space of A to the column space of A i.e maps back from the column space of A to the row space of A is called the **'Pseudoinverse'** of A and is represented by A^+.

20.1 Pseudo-Inverse using SVD

SVD of A gives $A = U \Sigma V^T$ where Σ is the diagonal matrix of singular values. The pseudoinverse Σ^+ is obtained by taking reciprocals of the non-zero singular values of Σ (the singular values that are 0 correspond to the null spaces and are hence ignored).
$\therefore A^+ = V \Sigma^+ U^T$
The pseudoinverse is used in finding least squares solution when A is not a full rank matrix.