Week 1

INTRODUCTION

Mathematical optimization / Programming is basically a tool which is used to solve quantitative problems in various fields like Engineering, economics, biology, physics, finance, and so on.

The idea here is that, we express a Real-world problem as an optimization model.

- Any optimization problem would have an objective or a goal which needs to be accomplished such as to minimize / maximize something.
- It may have some design variables or decision variables, based on what action to take or what to design like length or measurement, ...
- It may have constraints or restrictions on those design variables.

Goal

The Goal of this course is to appreciate the role of Convexity in Optimization theory. There are different kinds of optimization problems, and Convexity plays a very special role.

We will learn to formulate problems in Signal processing and communication, as optimization problems.

We will learn about the Watershed (convexity) between easy and hard problems.

We will see, how seemingly Hard problems can be written as easy problems.

We will also look at the General approaches to solve these problems approximately.

Let's try to motivate the role of optimization in Convex Optimization.

- What an optimization problem is ?

$$\boxed{\begin{array}{l} \min \; f_0(x) \\ \text{subject to the constraint } \; f_i(x) \le 0 \end{array}}$$

where, $f_0(x)$ - is the Objective

$f_i(x)$ is the constraint

- we may also have equality constraint, say $f_i(x) = 0$

x is the decision variable, which can be a vector.

- $x \in \mathbb{R}^n$
- $x_{n \times 1}$
- $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

x is always column vector, NOT row vector

This is an example of an Optimization problem.

Let's look at a concrete example.

① - $n = 1$ (i.e) (1-Dimensional vector) (or) (Scalar)

We can have problem of the form

$\min \; (x-1)^2$, subject to the constraint $x \ge 0$.

The Solution is $x^* = 1$ optimum value

② $n = 2$ (i.e) 2-Dimensional vector, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{2\times 1}$, $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{2\times 1}$

$\min \; \|x - a\|^2$, subject to the constraint $x \ge 0$.

$\Rightarrow \min \; (x_1 - 1)^2 + (x_2 - 2)^2$, s.t $x_1 \ge 0$, $x_2 \ge 0$.

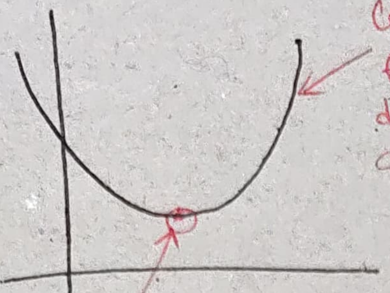
The Solution is $x^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

These are 2 examples of simple optimization problem.

- Now, let us come to Convex Optimization.

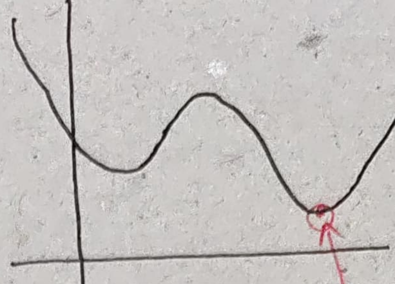
What is convexity ?

It is the depiction of a convex function.



Convexity for one dimensional case.

plot 1: Convex Function



Plot 2: Non-convex function

Note the single point where the function is minimum locally and globally.

There is still single point where the function is minimized, but there are two points which looks like minima atleast locally.

So, what is about convex optimization that we prefer?

1. Convex functions are easy to optimize over. (ie) in plot 1, we see that is single point which is minimum locally and globally, which apparently helps a lot in designing algorithms which are optimum.

2. There are some Mathematical properties which help in the optimization process. One of them is "Local minima = Global minima", which is most important property.

3. Usually, when we design Convex Optimization algorithms, we are able to guarantee convergence

The algorithms are iterative. Each iteration tend to optimum solution. So, most of the optimum algorithms are very Predictable.

Say for example!, Least Squares problem

$$x^* = \arg\min \underbrace{\|Ax - b\|^2}$$

$$(Ax-b)^T (Ax-b)$$

$$\underbrace{[\;\cdots\cdots\;]}_{m \text{ entries}} \left.\begin{bmatrix} \vdots \\ \vdots \end{bmatrix}\right\} m \text{ entries}$$

where,

$x \in \mathbb{R}^n$

$b \in \mathbb{R}^m$

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

$(Ax) \in \mathbb{R}^m$

$(Ax - b) \in R^m$

Special case:

(i) If A is Square Matrix (ie) $n = m$, and A is invertible (ie) A^{-1} exists

Then the solution is given by

$$x^* = A^{-1}b \quad , \text{ so that } \|Ax^* - b\| = 0 .$$

(ii) If A is not Square Matrix (ie) $n \neq m$, and A^TA is invertible (ie) $(A^TA)^{-1}$ exists, Then the solution is given by

$$x^* = (A^TA)^{-1}A^Tb$$

In general, x^* is a CLOSED FORM solution (ie) there exist formula for the solution. But it is not necessary that all optimization problems can be solved in closed form.

Following MATLAB Commands will give solution of the corresponding LS problem for all possible n and m.

- A \ b
- lscov(A, b)

Example 2. Linear programming:

The Linear Programs are the first optimization problems that were not solvable in closed form. They were studied in depth since 1940s. The Linear programs takes the form

$$\min \; c^Tx \qquad\qquad c, x \in \mathbb{R}^n$$
$$\text{s.t } Ax \leq b \qquad\qquad b \in \mathbb{R}^m$$

A is of the form,

$$A = \begin{bmatrix} \text{---} \; a_1^T \; \text{---} \\ \text{---} \; a_2^T \; \text{---} \\ \vdots \\ \text{---} \; a_m^T \; \text{---} \end{bmatrix} , \text{ where, } \begin{array}{l} a_1^Tx \leq b_1 \\ a_2^Tx \leq b_2 \\ \vdots \\ a_m^Tx \leq b_m . \end{array} \qquad a_i^Tx \leq b_i \quad i = 1, 2, \dots, m.$$

Another way of writing this is,

$$\sum_j a_{ij} x_j \leq b_i \quad , \quad \begin{array}{l} i = 1, 2, \dots m \\ j = 1, 2, \dots n \end{array}$$

$$[A]_{ij} = a_{ij}.$$

This is the Linear Program.

Both Least squares and Linear Programming are well studied area in optimization problems.

Let us see the complexity of solving these problems.

(ii) How many computations does it take to solve the problems to some accuracy?

One way of measuring the complexity is to measure how many floating point operations (flops) eg. $(\times, \div, +, -)$.

In case of Least Squares and Linear Programming, the complexity is in the order of

$$\sim O(mn^2)$$

No. of flops $\leq C mn^2$
where C is a constant.
$m > n$.

In general, for Convex optimization problems, the complexity is atmost in the order of

$$\sim O(n^3) \quad , \text{ not beyond this.}$$

where,

$$n = \max(\#\text{ variables}, \#\text{ constraints}).$$

For non-convex optimization / General optimization problem, the complexity is in the order of

$$O(e^n).$$

where, $e^n >> n^3$ or n^5.

Eg $e^{10} \sim 22026$
$n^3 \sim 1000$

Notations

① Vector $\underline{x} \in \mathbb{R}^n$

The underline may not be used in some places where it should be clear from the context.

② Scalar $x \in \mathbb{R}$

③ Matrix $X \in \mathbb{R}^{m \times n}$

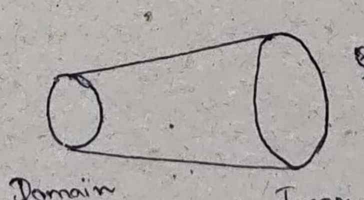
④ Set $\mathcal{X} \in \mathbb{R}^n$

calligraphic $\mathcal{X}$

- Can be set of vectors, set of matrices, and so on..

⑤ Function. $f : \mathcal{A} \rightarrow \mathcal{B}$

- Function goes from one set to Another set.



Eg. Function goes from the domain to image.

- So, generally, the function domain also needs to be specified.

function takes value in $\mathbb{R}$

Output value in $\mathbb{R}$

<u>Example</u>: $f(x) = \frac{1}{x^2}$

function $f : \mathbb{R} \rightarrow \mathbb{R}$

More precisely, the domain of f needs to be explicitly specified as

$$\text{dom. } f = \underbrace{\mathbb{R} \setminus \{0\}}_{} = \{x \in \mathbb{R},\ x \neq 0\}$$

Everything in R, except 0

$$\text{im. } f = \underbrace{\mathbb{R}_{++}}_{} = \{x \in \mathbb{R},\ x > 0\}$$

Set of x in R, such that x is strictly +ve.

Matrix function

<u>Example</u>: $f(x) = \log(\det(X))$

This is applicable for function for which determinant can be Calculated.

function $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$

$$\text{dom. } f = \{x \in \mathbb{R}^{n \times n},\ \det(X) > 0\}$$

$$\text{im. } f = \mathbb{R}$$

<u>Some Important Concepts of Linear Algebra</u>.

<u>INNER PRODUCT</u>.

$$\underbrace{\langle \underline{x}, \underline{y} \rangle}_{} = x^T y = \sum_{i=1}^{n} x_i y_i$$

where $\underline{x}, \underline{y} \in \mathbb{R}^n$

x and y are vectors of same size.

<u>EUCLIDEAN NORM</u>.

$$\|\underline{x}\|^2 = \langle \underline{x}, \underline{x} \rangle = x^T x = \sum x_i^2$$

<u>CAUCHY - SCHWARZ INEQUALITY</u>

The inequality that relates inner product with Euclidean norm.

$$\boxed{-\|x\| \|y\| \leq \langle x, y \rangle \leq \|x\| \|y\|}$$

This becomes EQUALITY if and only if x and y are scalar multiples.

i.e) $\underline{x} = \alpha \underline{y}$

Scalar

Vectors.

$$\Rightarrow \langle x, y \rangle = \alpha \langle x, x \rangle = \alpha \|x\| \|x\|$$

ANGLE

Angle is defined only for non-zero vectors.

$x, y \neq 0$, then $\theta = \cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\| \|y\|}\right)$.

The vectors x and y are said to be ORTHOGONAL if $\langle x, y\rangle = 0$.

The above concepts are described for vectors. In similar fashion, these concepts can be extended to Matrices as well.

Let us look at an example of vectors in n-dimension, which are all orthogonal.

- Consider vector, $e_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}$ ← i^{th} location

which is 0 everywhere, but 1 at i^{th} location.

$$\langle e_i, e_j\rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

$e_1, e_2, \ldots, e_n$ are orthogonal (basis vectors).

- Consider Matrix X and Y.

$$\langle X, Y\rangle = Tr(X^T Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} Y_{ij}$$

where, $X, Y \in \mathbb{R}^{m \times n}$

Norm : $\|X\|_F^2 = Tr(X^T X) = \langle X, X\rangle = \sum_{i,j} X_{ij}^2$

← Frobenius Norm.

Symmetric matrices :

$$S^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^T\}$$

Set of Symmetric Matrices

Norm.

Norm is basically, any function that satisfies certain properties.

- Represented as $\| . \|$.
- It should not have any domain restriction.
  (i) should be defined $\forall x$.
  Eg. Vector Norm $\underline{x} \in \mathbb{R}^n$, then $\|x\|$ defined $\forall x \in \mathbb{R}^n$
- Norm has to be <u>non- negative</u>.
  $$\|x\| \geq 0 \quad \forall \underline{x}$$
- Norm has to be <u>definite</u>
  $$\|x\| = 0 \text{ only if } x = 0$$
- Homogeneity
  $$\|t \, \underline{x}\| = |t| \, \|x\| \quad , \; t \in \mathbb{R}$$
  Scalar
  vector
  absolute value of t
- Triangle inequality.
  $$\|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|.$$

If any function has the above properties, it can be set to be a Norm. In particular, we'll use a very interesting set throughout, which is <u>Unit Norm Ball</u> (B).

$$B = \underbrace{\{x \mid \|x\| \leq 1\}}$$

Set of all x such that $\|x\| \leq 1$

Let us look at several examples of Norm.

① <u>vector norm.</u>

$\|\underline{x}\|_2$, which is l_2 - Norm.

(i) $\sqrt{\langle x, x \rangle} \Rightarrow \sqrt{x^T x}$.

② <u>l_1 - norm</u>

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

③ <u>Infinity Norm.</u>

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

④ l_p - Norm : It generalizes the whole class of Norms.

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \quad , \quad p \geq 1 .$$

Matrix Norms

① Frobenius Norm

$$\|X\|_F$$

② Sum - Absolute value (SAV) Norm

$$\|X\|_{SAV} = \sum_{i=1}^{n} |X_{ij}| \quad , \quad \text{if } X \text{ is a } n \times n \text{ Matrix,}$$

Equivalence of norms

For any x or X, its norm is upper and lower bounded by some constant times another Norm.

$$\alpha \|x\|_a \leq \|x\|_b \leq \beta \|x\|_a \quad , \quad \forall \, a, b \geq 1$$

$\alpha, \beta \rightarrow$ do not depend on X, but depend on x itself (or) a, b.

When $\|x\|_b$ is small $\Rightarrow$ $\|x\|_F$ is also small.

When $\|x\|_b$ is large $\Rightarrow$ $\|x\|_F$ is also large.

Weighted Norm

It is a special type of norm.

$$\|v\|_a = \sum_{i=1}^{n} \overset{\text{Weights}}{a_i} \, v_i^2 \quad , \quad \text{where } v, a \in \mathbb{R}^n$$

Is this a valid norm? (or)

For what kind of weights, this is a valid Norm?

The weight a_i should be strictly > 0 in order to be a valid Norm. (ie) $a_i > 0$.

Counter-example : Suppose that $a_1 = 0$; $a_2, \ldots, a_n > 0$

We can find v such that $\|v\|_a = 0$, while $v \neq 0$. (Definiteness property is violated).

Eg. $v_1 = 1$, $v_2, \ldots, v_n = 0$.

$\Rightarrow a_i > 0 \;\; \forall i$ for valid norm.

Alternatively,

$$\|v\|_A^2 = v^T A v = [v_1 \ v_2 \ \dots \ v_n] \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & a_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

$$= \sum a_i v_i^2$$

where A is a diagonal Matrix. (ie) $[A]_{ii} > 0$.

In general case, where A is not diagonal, but Real Symmetric $n \times n$ matrix, $A \in S^n$, motivates the idea of EVD.

Symmetric Eigen Value Decomposition (EVD).

For simplicity, we use Symmetric EVD through out.
For Symmetric EVD, the Eigen values are always REAL.
For Asymmetric EVD, the Eigen values are COMPLEX.

General Symmetric Matrix (A) should be able to written as product of 3 matrices. (ie)

$$A = Q \Lambda Q^T$$

where, $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ is a diagonal Matrix.

Eigen values.

$$Q = \underbrace{\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}}$$

Eigen vectors.

$$Q \in \mathbb{R}^{n \times n}$$

$$a_i \in \mathbb{R}^n$$

Alternate way to represent EVD.

$$A = Q \Lambda Q^T = \sum_{i=1}^{n} \lambda_i \underbrace{q_i \, q_i^T}$$

Sum of Rank-1 Matrices.

Note that q_i is a column vector, q_i^T is a row vector.

$$\underbrace{q_i \cdot q_i^T}_{\text{Multiplication}} \Rightarrow n \times n \text{ Matrix}$$

$$\Rightarrow \text{Rank}\,(q_i\, q_i^T) = 1.$$

Also, Q is orthogonal.

(i) $QQ^T = Q^TQ = I$ and

$Q^{-1} = Q^T$.

Also, $\langle q_i, q_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

$$A q_i = \left(\sum_{j=1}^{n} \lambda_j q_j q_j^T \right) q_i$$

$$= \sum_{j=1}^{n} \lambda_j q_j q_j^T q_i$$

out of these n terms, only one of the term will be non-zero (i) when $i = j$

$$= \lambda_i q_i$$

Properties

Recall: $Tr(AB) = Tr(BA)$

$Tr(ABC) = Tr(CAB) = Tr(BCA)$

① $\Rightarrow Tr(A) = Tr(Q \Lambda Q^T) = Tr(\underbrace{Q^T Q}\, \Lambda)$

$\because Q$ is orthogonal, $Q^TQ = I$.

$= \underbrace{Tr(\Lambda)}$

Trace of a Matrix is Sum of the diagonal entries

$= \sum \lambda_i$

$$\det(A) = \underbrace{\prod_{i=1}^{n} \lambda_i}$$

Product of the Eigen values

where, $\det(A) > 0$ only when $\lambda_i > 0 \ \forall\, i = 1, \dots, n$.

② $\Rightarrow \det(A - \lambda I) = \underbrace{(\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)}$,

Scalar

Characteristic polynomial, which depends on the Eigen values $(\lambda_1, \lambda_2, \dots, \lambda_n)$ of A.

Example: $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, $A - \lambda I = \begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}$

$$\det(A - \lambda I) = |A - \lambda I| = \underbrace{(3-\lambda)^2 - 1 = 0}_{\text{Characteristic Equation}}$$

The roots of this characteristic equation are

$$\underbrace{\lambda = 2, 4}$$ Eigen values of A

Let's find the Eigen vectors

$$A q_i = \lambda_i q_i = 2 q_i$$

$$\Rightarrow \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} q_{11} \\ q_{12} \end{bmatrix} = \begin{bmatrix} 2 q_{11} \\ 2 q_{12} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 3q_{11} + q_{12} \\ q_{11} + 3q_{12} \end{bmatrix} = \begin{bmatrix} 2q_{11} \\ 2q_{12} \end{bmatrix}$$

$$\Rightarrow 3q_{11} + q_{12} = 2q_{11}$$

$$\Rightarrow q_{11} + q_{12} = 0$$

Let $q_{11} = \alpha$, $q_{12} = -\alpha$

$$\Rightarrow q_1 = \begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} \text{ for some } \alpha.$$

$$q_1^T q_1 = \|q_1\| = 1.$$

So; $\alpha^2 + \alpha^2 = 1 \Rightarrow \alpha = \frac{1}{\sqrt{2}}$.

So, $q_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

q_2' which is orthogonal to q_1. (ie) $q_1^T q_2 = \langle q_1, q_2 \rangle = 0$

Therefore, $q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

PROOF OF THE CAUCHY SCHWARZ INEQUALITY

we have,

$$\langle u, v \rangle \le \|u\|_2 \|v\|_2$$

l_2 - Norms

$u, v \in \mathbb{R}^n$

$u, v \to$ Vectors

$t \to$ Scalar

$u^T v = v^T u$

If $\|u + tv\|_2 \ge 0$,

$\|u + tv\|_2^2 \ge 0$ as well.

$$\Rightarrow (u + tv)^T (u + tv) \ge 0$$

$$\Rightarrow (u^T + tv^T)(u + tv) \ge 0$$

$$\Rightarrow u^T u + tv^T u + u^T tv + t^2 v^T v \ge 0$$

$$\Rightarrow \underbrace{\boxed{u^T u + 2t(u^T v) + (v^T v)t^2 \ge 0.}}_{\substack{\text{This LHS quantity should also} \\ \text{be } \ge 0 \text{ for minimum possible} \\ \text{value of } t \text{ that we can keep in the LHS.}}} \quad \forall t.$$

$$\Rightarrow \min_t \underbrace{\left(u^T u + 2t(u^T v) + (v^T v)t^2\right)}_{\text{This is a simple quadratic function.}} \ge 0$$

This can be minimized by differentiating w.r.t t and set equal to 0.

$$\Rightarrow \frac{d}{dt}\left(u^T u + 2t(u^T v) + (v^T v)t^2\right) = 0.$$

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

$$\frac{d}{dx}(\text{const}) = 0.$$

$$\Rightarrow 0 + 2(u^T v) + 2t(v^T v) = 0$$

$$\Rightarrow t = -\frac{u^T v}{\|v\|_2^2}$$

This is the value of t which minimizes LHS. Substituting this value of t in inequality equation will minimize LHS.

$$\Rightarrow u^T u + 2\left(\frac{-u^T v}{\|v\|_2^2}\right)(u^T v) + (v^T v)\left(\frac{-u^T v}{\|v\|_2^2}\right)^2 \ge 0$$

$$\Rightarrow \|u\|^2 - \frac{2}{\|v\|_2^2}(u^T v)^2 + \|v\|^2 \frac{(u^T v)^2}{\|v\|^4} \ge 0$$

$$\Rightarrow \|u\|^2 - (u^T v)^2\left(\frac{2}{\|v\|^2} - \frac{1}{\|v\|^2}\right) \ge 0$$

$$\Rightarrow \frac{\|u\|^2 \|v\|^2 - (u^T v)^2}{\|v\|^2} \ge 0$$

$$\Rightarrow \|u\|^2\|v\|^2 - (u^T v)^2 \geq 0$$

$$\Rightarrow \|u\|^2\|v\|^2 \geq (u^T v)^2$$

$$\Rightarrow \boxed{(u^T v)^2 \leq \|u\|^2\|v\|^2}$$

Taking square root on both sides

$$\Rightarrow u^T v \leq \|u\|\,\|v\|$$

$$\Rightarrow \boxed{\langle u, v \rangle \leq \|u\|\,\|v\|}$$

This is the Cauchy Schwarz Inequality.

For Matrices also, it can be proved in the same way.

(i) $Tr(A^T B) \leq \|A\|_F \|B\|_F$

<u>Positive Semidefinite Matrices</u>. (PSD).

Recall, the weighted Norm $\|v\|_P^2 = v^T P v$, where $\underbrace{P \in S^n}_{P \text{ is Symmetric}}$

What are the properties that P should have, for this to be a valid Norm?

<u>Answer</u>: If P is diagonal, then all the diagonal entries has to be +ve.

If P is a general Matrix, then P has to be +ve Definite.

$v^T P v > 0$ whenever $v \neq 0$.

This particular definition of norm is sufficient to define the positive definiteness property of a general Symmetric Matrix B, so, a symmetric matrix P with defined a valid norm, if $v^T P v > 0$, whenever v is non-zero.

Let's take EVD of P.

$$\Rightarrow P = Q \Lambda Q^T$$

$$v^T P v = \underbrace{v^T Q}_{z^T} \Lambda \underbrace{Q^T v}_{z} = z^T \Lambda z, \text{ where } z = Q^T v$$

$$= \sum \lambda_i z_i^2$$

$\sqrt{v^T P v}$ is a valid norm when all Eigen values are positive $(\lambda_i > 0)$

For diagonal matrix to define a valid norm, it is required that Diagonal elements to be strictly positive.

For general matrix to define a valid norm, it is required that Eigen values to be strictly positive.

Thus, every result we have for diagonal matrix can be applied to general matrix with Eign values, instead of diagonal elements.

In summary, $\sqrt{v^T P v}$ is a valid norm when $\lambda_i > 0,\ i = 1, 2, \dots, n$ w.r.t z. As $\underbrace{z = Q^T v \Rightarrow v = Qz}_{\text{one to one mapping of } z \text{ \& } v}$, it is a valid norm w.r.t v.

Recap

① $v^T P v > 0 \quad \forall\ v \neq 0$

② $\lambda_i(P) > 0 \quad \forall\ i$

→ Equivalent.

when P is positive definite

Notation : $P > 0$

$P > 0 \not\Rightarrow$ Entries of P are positive

$\Rightarrow$ Eigen values of P are positive.

③ $P = L L^T$, where L is Lower Triangular & full Rank.

$L \rightarrow$ 

This is called Cholesky decomposition.

$L^T \rightarrow$ 

It is a very quick way of checking numerically whether a Matrix is Positive Definite or not.

Instead of finding the Eigen values and checking whether they are positive definite or not, we can find cholesky decomposition of P. If the cholesky decomposition is successful, then it is positive definite. If it is unsuccessful, then MATLAB throws an error.

Basically, cholesky decomposition is found through a series of steps. If those steps fail, then the matrix is not positive definite.

Note: $v^T P v = v^T L L^T v = (L^T v)^T (L^T v) = \|L^T v\|_2^2 \geq 0 \; \forall \; v \neq 0.$

$L^T v \neq 0$ when $v \neq 0$.

This holds when L is full rank.

This is just faster way to check numerically whether P is Positive definite ($P > 0$) or not.

There is a related notion of positive semi definite matrices. PSD matrices do not define a norm by-the-way, but PSD is a very useful notion.

Let P be positive semidefinite Matrix, $P \in S^n$ ← Symmetric Matrix

(a) $P \geq 0$ (Positive and non-negative) (Not entrywise)

(b) $\lambda_i(P) \geq 0$ (Eigen values are non negative)

(c) $v^T P v \geq 0 \;\; \forall \; v \in \mathbb{R}^n$ (Here it is possible that v is non zero, but $v^T P v$ is zero)

(d) $P = AA^T$ for any $A \in \mathbb{R}^{n \times m}$, $m \leq n$.

Here, we cannot have Cholesky decomposition.

Suppose, $P = AA^T$

$$v^T P v = v^T A A^T v = (A^T v)^T (A^T v) = \|A^T v\|_2^2 \geq 0 \;\; \forall \; v$$

We cannot say that A is full Rank, and we dont necessarily require that. We only say that $\|A^T v\|_2^2 \geq 0$. We cant say that $\|A^T v\|_2^2$ is non-zero.

$$\Rightarrow P \geq 0.$$

Note: We define Positive Definite (PD) and Positive semidefinite (PSD) matrices only for Symmetric matrices. The reason being, for non-symmetric case, $\lambda_i(P)$ may be complex, where there wont be the notion of positivity and negativity.

$\|v\|_P$ is a Norm for $P > 0$. (Positive Definite)

not a Norm for $P \geq 0$ (Positive semidefinite)

Let us discuss the Singular Value Decomposition (SVD) and the Fundamental spaces associated with a Matrix.

Let A be a $m \times n$ rectangular Matrix. (i) $A \in \mathbb{R}^{m \times n}$

The fundamental spaces associated with this Matrix are as follows.

⊙ Range Space

Range space of A, $R(A) = \{\underline{v} \in \mathbb{R}^m \mid \underline{v} = A\underline{u}\}$,

for any $u \in \mathbb{R}^n$.

Another way of saying this is, any vector in the Range space of A is expressible as a linear combination of the n columns of A.

$$\text{(i)}\ v = u_1 \begin{vmatrix} \\ a_1 \\ \\ \end{vmatrix} + u_2 \begin{vmatrix} \\ a_2 \\ \\ \end{vmatrix} + \dots + u_n \begin{vmatrix} \\ a_n \\ \\ \end{vmatrix}$$

This gives an element of the range space of A.

It is also called as Column Space.

$R(A)$ is a subspace.

(i) if $v_1, v_2 \in R(A)$, then there exists u_1, u_2 such that

$V_1 = Au_1$,

$V_2 = Au_2$.

Then, $\alpha_1 \underline{v_1} + \alpha_2 \underline{v_2} = A(\alpha_1 \underline{u_1} + \alpha_2 \underline{u_2})$

$(\alpha_1 v_1 + \alpha_2 v_2) \in R(A)$, coz it satisfies the property "Expressible as a linear combination of columns of A".

Dimension of the subspace,

$$\begin{aligned} \dim(R(A)) &= \text{No. of linearly independent columns of } A \\ &\leq n \\ &= \text{Column Rank of } A \end{aligned}$$

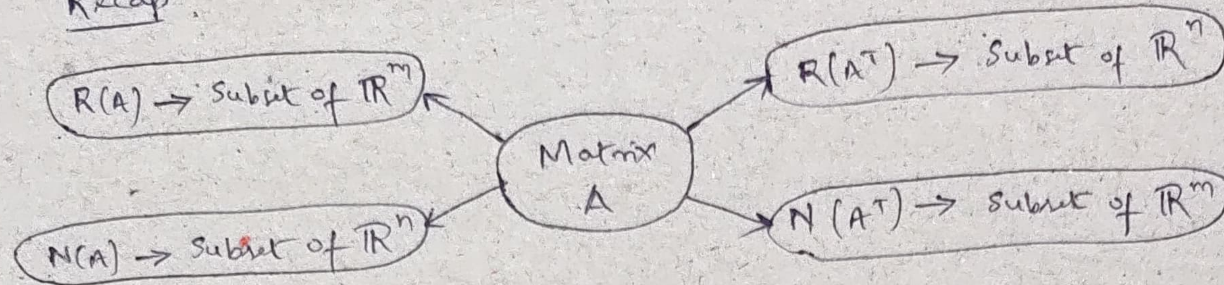
⊙ Null Space

$$N(A) = \{u \in \mathbb{R}^n \mid Au = 0\}.$$

$$R(A^T) = \{u \in \mathbb{R}^n \mid u = A^T v,\ v \in \mathbb{R}^m\}$$

$$N(A^T) = \{v \in \mathbb{R}^m \mid A^T v = 0\}.$$

Recap.



These are the 4 fundamental spaces associated with the Matrix A, which are actually related.

Let's find out what is the subspace that is orthogonal to the $R(A)$. The vectors that are orthogonal to all the vectors within $R(A)$, is nothing but the subspace that is fully orthogonal to $R(A)$.

Let us say that w is such a vector, which is orthogonal to $R(A)$. What properties does w satisfy?

$$w \in R(A)^\perp \iff w^T(\underbrace{Au}) = 0 \quad \forall u$$

Any vector in $R(A)$ of the form Au
$v = Au$

$$\Rightarrow \underbrace{(A^T w)^T u} = 0 \quad \forall u$$

This is inner product of two vectors, which is always 0, regardless of what u is..
This is only possible if both the vectors are 0.
(or) $\underbrace{A^T w \perp u \quad \forall u \in \mathbb{R}^n}$
This is only possible when $A^T w = 0$ and $w \in N(A^T)$.

So, $R(A)^\perp = N(A^T)$.

Similarly, we can prove that $N(A)^\perp = R(A^T)$.

Another property that we have is

$$\text{rank}(A) = \dim(R(A)) = \dim(R(A^T)) = r \leq m, n.$$

This is where SVD comes in.

Let's rewrite what we have. The Basis vectors of 4 fundamental spaces are

- Basis vectors of $R(A)$
  $v_1, v_2, \ldots, v_r \in R(A)$
- Basis vectors of $N(A^T)$, which is orthogonal to $R(A)$
  $v_{r+1}, v_{r+2}, \ldots, v_m \in N(A^T)$.

$\Big\}\ v_i \in \mathbb{R}^m$.

⊙ Basis vectors of $R(A^T)$

$u_1, u_2, \ldots, u_r \in R(A^T)$

⊙ Basis vectors of $N(A)$, which is orthogonal to $R(A^T)$

$u_{r+1}, u_{r+2}, \ldots, u_n \in N(A)$

$\Big\}\ u_i \in \mathbb{R}^n$

Singular Value Decomposition (SVD)

SVD is essentially writing A in terms of these vectors.

$$A = \underbrace{\begin{bmatrix} \underbrace{v_1 \cdots v_r}_{R(A)} & \underbrace{v_{r+1} \cdots v_m}_{N(A^T)} \end{bmatrix}_{m\times m}}_{V} \underbrace{\begin{bmatrix} \sigma_1 & \cdots & 0 & 0 \\ & \ddots & & \\ 0 & \cdots & \sigma_r & 0 \\ 0 & \cdots & & 0 \\ \vdots & & & \\ 0 & \cdots & & 0 \end{bmatrix}_{m\times n}}_{\Sigma} \underbrace{\begin{bmatrix} u_1^T \\ \vdots \\ u_r^T \\ u_{r+1}^T \\ \vdots \\ u_n^T \end{bmatrix}_{n\times n}}_{U}$$

Singular values

$R(A^T)$, $N(A)$

$$= \sum_{i=1}^{r} \sigma_i v_i u_i^T \quad \leftarrow \text{Sum of Rank 1 matrices}$$

$$= V \Sigma U^T \quad \leftarrow \text{does not depend on } v_{r+1} \text{ (or) } u_{r+1}.$$

SVD vs EVD

We have, $A = V \Sigma U^T$, where V and U are orthogonal

(i) $V^T V = V V^T = I$

$U^T U = U U^T = I$

$$A^T A = (V\Sigma U^T)^T (V \Sigma U^T)$$
$$= U \Sigma^T \underbrace{V^T V}_{I} \Sigma U^T$$
$$= U \Sigma^T \Sigma U^T$$
$$= U \Lambda U^T$$

Eigen values, $\Lambda = \begin{bmatrix} \sigma_1^2 & \cdots & 0 & 0 \\ & \ddots & & \\ 0 & \cdots & \sigma_r^2 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}_{n\times n}$

$\sigma_i^2(A) = \lambda_i(A^T A)$, $i = 1, 2, \ldots, r$

Similarly,

$\sigma_i^2(A) = \lambda_i(A A^T)$
