22/11/23

Lecture - 2

(24/02/21)

optimization & Lagrange Multipliers

Continuing from Lecture - 1

→ Necessary and sufficient condition for positive semidefiniteness

$$\boxed{\forall f \quad \iint A(\alpha,\beta) f(\alpha) f(\beta)\, d\alpha\, d\beta \geq 0 \qquad A(\alpha,\beta) = A(\beta,\alpha) \iff \mathcal{F}[A](\omega) \geq 0}$$

Def:

$$\underbrace{Q(T) = \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} A(\alpha-\beta)\, e^{-i2\pi\xi(\alpha-\beta)}\, d\alpha\, d\beta}$$

a suitably defined function to prove our case

i.e we will prove that

$$\lim_{T\to\infty} T\,Q(T) = \int_{-\infty}^{\infty}\!\!\int A(\alpha-\beta)\underbrace{e^{-2\pi i \xi \alpha}}_{f(\alpha)}\underbrace{e^{+2\pi i \xi \beta}}_{f^*(\beta)}\,d\alpha\, d\beta \geq 0$$

This whole function represents our Quadratic form

i.e

$$\boxed{\int_{-\infty}^{\infty}\!\!\int A(\alpha-\beta)\, f(\alpha)\, f^*(\beta)\, d\alpha\, d\beta \geq 0}$$

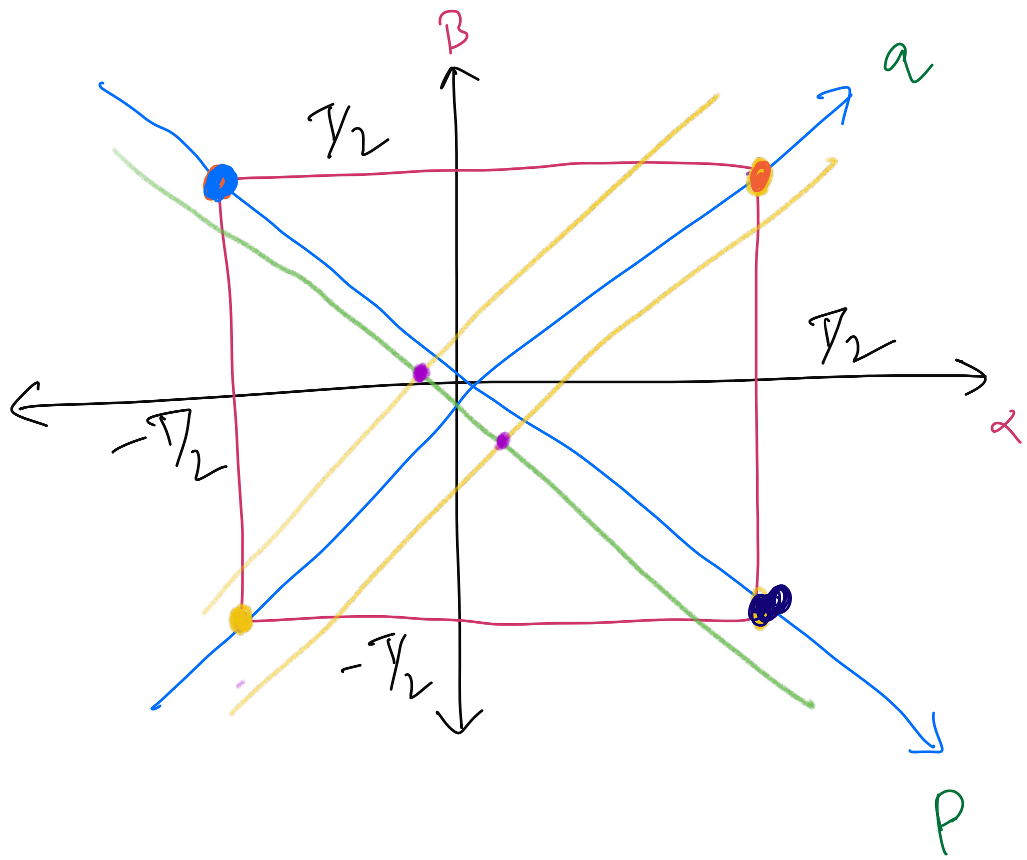
$$\lim_{T\to\infty} T\,Q(T) \geq 0 \Rightarrow \lim_{T\to\infty} Q(T) \geq 0$$

we have already proved This in Lecture-1

Now if want to compute $Q(T) \Leftrightarrow \mathcal{F}[A]\omega \geq 0$ and succeed in showing that in fact it is the Fourier Transform of 'A' Then The

Theorem is proved. (i.e Sufficient condition for PSD is proved)

$$\lim_{T\to\infty} Q(T) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} A(\alpha-\beta)\, e^{-2\pi i \zeta(\alpha-\beta)}\, d\alpha\, d\beta$$



$p = \alpha - \beta$ $\qquad$ $\alpha = \frac{p+q}{2}$

$q = \alpha + \beta$ $\qquad$ $\beta = \frac{-p+q}{2}$

α, β are the integration variables of in the integral they appear as difference $(\alpha - \beta)$

$\therefore$ it would be useful to transform them as above. (i.e we have a different reference system)

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \underbrace{\begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}}_{\text{det}} \begin{bmatrix} p \\ q \end{bmatrix}$$

$$\Downarrow$$

$$\frac{1}{2} \cdot \frac{1}{2} - \left(-\frac{1}{2}\right)\left(\frac{1}{2}\right) \Rightarrow 1/2$$

i.e Jacobian of this transformation is absolute value of it's determinant of the Matrix above.

we have to reparametrize the domain of Integration for 'P' $\rightarrow$ Blue to Black

$= \alpha - \beta$

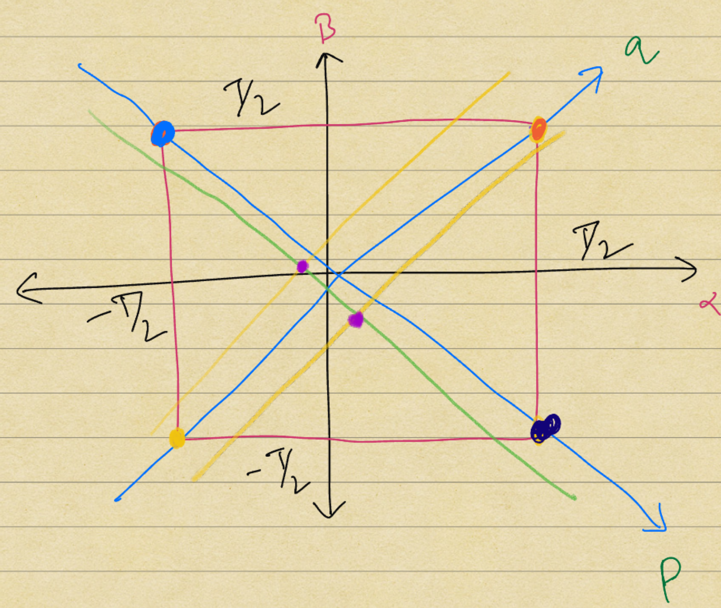
$\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ $\qquad$ $\left(\frac{\pi}{2}, -\frac{\pi}{2}\right)$

$\alpha - \beta = -\frac{\pi}{2} - \frac{\pi}{2}$ $\qquad$ $\alpha - \beta = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right)$

$= -\pi$ $\qquad$ $= \pi$

for 'Q' $\qquad$ Yellow to Red

If we look at the Purble which is the intersection of Green & yellow line



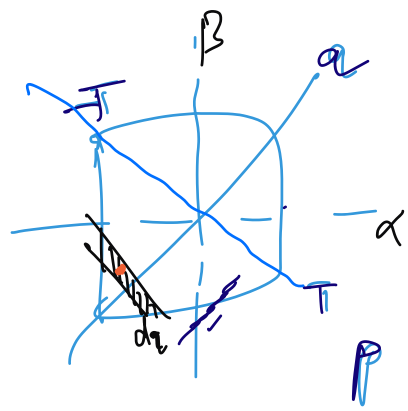
$p = \alpha - \beta$ $\qquad$ $\alpha = \frac{p+q}{2}$

$q = \alpha + \beta$ $\qquad$ $\beta = \frac{-p+q}{2}$

Q) what are the variables that identify with the yellow line that is depicted in the figure?

They are symmetric w.r.to zero.
(purple points)

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} \int_{|p|-T}^{T-|p|} A(p)\, e^{-2\pi i \xi p}\, \frac{1}{2}\, dp\, dq \geq 0$$



$\therefore$ $p = \alpha - \beta$

$q = \alpha + \beta$

dp limits $-T, T$, $q = \alpha + \beta$

$q = \alpha - \beta + 2\beta$

$q = p + 2\beta$

$$q = \cancel{p} + 2(p)$$

$$q = 0 + -T \quad = |p| - T$$

$$q = -p + 2(T/2) = T - |p|$$

Then

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A(p)\, e^{-2\pi i \zeta p} \int_{|p|-T}^{T-|p|} dq \; dp \geq 0$$

$$\Rightarrow \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A(p)\, e^{-2\pi i \zeta p} \left[2\left(T - |p|\right) \right] dp \geq 0$$

$$\Rightarrow \lim_{T\to\infty} \frac{1}{\cancel{2}T} \int_{-T}^{T} A(p)\, e^{-2\pi i \zeta p} \cdot \cancel{2}(T-|p|)\, dp \geq 0$$

$$\boxed{\Rightarrow \lim_{T\to\infty} \int_{-T}^{T} A(p)\, e^{-2\pi i q p} \geq 0}$$

$\Rightarrow$ This is the Fourier Transform of 'A'

i.e $\mathcal{F}[A](\omega) \geq 0$

$\rightarrow$ Now it is possible to prove that the if F.T of 'A' is positive then the quadratic form is positive

i.e sufficient part of the proof

I know $\mathcal{F}[A](q) \geq 0$

$$\forall f \quad \int_{-\infty}^{\infty}\int A(\alpha-\beta)\, f(\alpha) f(\beta)\, d\alpha\, d\beta$$

we will take inverse F.T of $\mathcal{F}(A)$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\int_{-\infty}^{\infty} \mathcal{F}[A](\xi)\, e^{+2\pi i \xi(\alpha-\beta)}\, d\xi}_{IFT = 'A'}\; f(\alpha) f(\beta)\, d\alpha\, d\beta$$

exchanging order of Integration

$$\Rightarrow \int_{-\infty}^{\infty} \mathcal{F}[A](\xi) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{+2\pi i \xi \alpha} f(\alpha) \cdot e^{-2\pi i \xi \beta} f(\beta)\, d\alpha\, d\beta\, d\xi$$

exchanging order of Integration is possible if the Integral doesn't diverge

$$\Rightarrow \int_{-\infty}^{\infty} \mathcal{F}[A](\xi) \left[\underbrace{\int_{-\infty}^{\infty} f(\alpha) e^{+2\pi i \xi \alpha}\, d\alpha}_{X} \; \underbrace{\int_{-\infty}^{\infty} f(\beta) e^{-2\pi i \beta \xi}\, d\beta}_{X^*} \right] d\xi$$

if ≥ 0

Seperated into Two one dimentional Integral

is ≥ 0

$\Rightarrow$ X, X^* is conjugate of X

Then their product is always positive

x^2

$\therefore$ Now if $\mathcal{F}[A](\xi) \geq 0$

Then The whole Integral will be positive

$\therefore$ Our Quadratic Form will be Positive ~~Semi Definite~~

if $\mathcal{F}[A](\xi) \geq 0$

$$= \int_{-\infty}^{\infty} \underline{\mathcal{F}[A](\xi)} \left| \int_{-\infty}^{\infty} e^{i 2\pi i \xi \alpha} f(\alpha)\, d\alpha \right|^2 d\xi \geq 0$$

F.T takes the role of

Eigenvalue

Note:

This is why F.T of a Function is called

a Spectrum

& Eigenvalues of a Matrix are

also called Spectrum

Optimization — Lagrange Multipliers

As an Engineer one is just expected to OPTIMIZE everything we work in Technology Domain!

ex: 1) Design an Analog circuit with Gain set to this value depending on constrained Resources.

under power constrain

environmental constrains etc

→ Usually if you Model your system well enough you will come up with a Function and given all the degrees of Freedom you may have in your design.

ex:
- Size of a Transistor in an Analog circuit
- Size of cache in a Digital system
- Time schedule of the software

Depending on constrains we can come up with Merit/Cost Function

It measures the Quality of the Design.

Merit/cost function $\boxed{f : \mathbb{R}^n \to \mathbb{R}}$

Obviously we want to Minimize the cost function & Maximize the Quality (Merit) function.

& ∴ we can't have all the possible values we wish to incorporate in the Design

ex: we can't make the size of Transistor as Big as we want just to let more current for more amplification.

∴ our Domain is Constrained i.e it is subset of $\mathbb{R}^n$

$$\max / \min \; f(x)$$

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

Usually the Minimum & Maximum of our constrained Domain is provided by the

Constrain function.

$$\boxed{c_j(x) \geq 0} \quad j = 0, \ldots, m-1$$

If both f & C are considered to be linear functions,

ex: ① x_0 size of some entity

$$c_0(x) = -x_0 + \theta$$

if $c_0(x) \geq 0 \implies -x_0 + \theta \geq 0$

$$x_0 \leq \theta$$

② $c_0(x) = -c_1(x)$

& $\left.\begin{array}{l} c_0(x) \geq 0 \\ c_1(x) \geq 0 \end{array}\right\} \implies \underline{c_0(x) = c_1(x) = 0}$

→ In this course we will deal with only one of the optimization method i.e Lagrange's Multiplier. which will deal with Quadratic forms.

Overlap:

Most problem we face in Statistical Signal processing would be

- Minimize the Error
- Maximize the Energy

→ $\min/\max \; f(x)$

$$\text{s.t} \quad E_J(x) = 0, \quad J = 0, \dots, M-1$$

Naive Algorithm:

1) pick 'x' at random
2) check if $E_J(x) = 0 \;\; \forall \; J = 0 \dots M-1$
3) compute $f(x)$ keep the max/min so far

Lagrange Multipliers approach

The idea is

* why do you pick something at random, pick something based on the equations,

  solve these equations and try to pick values that make sense of the equations and we treat solutions as candidates to the Function (f)

$$\begin{cases} \min/\max \ f(x) & f: \mathbb{R}^n \to \mathbb{R} \\ & E_j: \mathbb{R}^n \to \mathbb{R} \\ \text{s.t.} \quad E_j(x) = 0, \quad j = 0, 1, \ldots, m-1 & \end{cases}$$

'n' equations

$$\nabla_x f(x) - \sum_{j=0}^{m-1} \lambda_j \nabla_x E_j(x) = 0$$

Because $f: \mathbb{R}^n \to \mathbb{R}$

'n' unknowns

'm' equations

$$\left.\begin{aligned} &E_0(x) = 0 \\ &E_1(x) = 0 \\ &\vdots \\ &E_{m-1}(x) = 0 \end{aligned}\right\}$$

Satisfy these constrains

'x' is 'n' dimensional vector

Solving all the above equations gives us a set of considerable 'x'

$$x = \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix}$$

The unknowns we are interested in.

$$\boxed{\nabla_x f(x) - \sum_{j=0}^{m-1} \lambda_j \nabla_x E_j(x) = 0}$$

What we have up here is a Gradient equality! i.e vector equality

$\because$ 'x' is an 'n' dimensional vector

$\therefore$ The vector equation is a collection of 'n' equalities.

& we have 'm' equalities in the form of constrains

$$E_0(x) = 0$$
$$\vdots$$
$$E_{m-1}(x) = 0$$

* Therefore we have $n+m$ equation and

$n+m$ unknowns.

'x' 'λ' (lagrangian multiplier)

'n' dimensional vector

Gradient concept: A Gradient is a generalization of the derivative, the idea is that when we have a function from $f: \mathbb{R} \rightarrow R \longrightarrow f'$ derivative

But when we have a function $f: \mathbb{R}^n \rightarrow R$

i.e the function $f\left(\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix}\right)$

Here There is a possibility of 'n' derivatives i.e $\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots\ldots \frac{\partial f}{\partial x_{n-1}}$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_0} \\ \vdots \\ \frac{\partial f(x)}{\partial x_{n-1}} \end{bmatrix}$$

This is a Gradient vector.

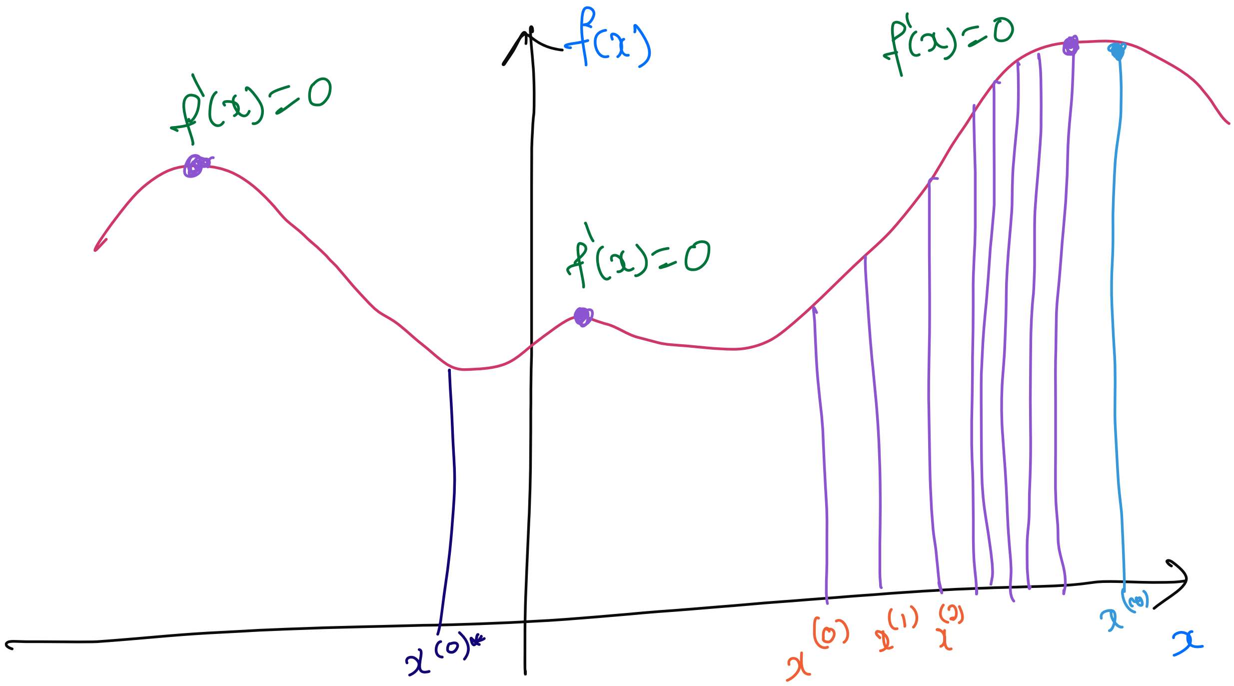
i.e vector of Gradients

→ Intuitive understanding of Lagrange's s/s of equations

why all the effort of defining equations with lagrangian multipliers and solving them along with constrains to get a set of candidates for our optimization problem.

lets look at an algorithm that is used in optimization i.e Gradient Algorithm for which Lagrange's Algorithm is a generalization

ex: simplest example $f: \mathbb{R} \to \mathbb{R}$ + no constant



If we start at $x^{(0)}$ and then calculate f(x) at $x^{(1)}$ and if the value is higher than $f(x^{(0)})$ then we move in the direction of increase of Gradient till we reach a maximum point

Q How do we know we reached a Maximum point ?

The point at $x^{(10)}$ where the Gradient is negative because it is after the maximum point where slope was zero

The above argument can be extended to Multiple dimensions.

But, if we choose a point $x^{(0)*}$ in the figure then we will be stuck with the local maximum.

* In the above example we know where maximums are calculate them and choose the best, but it is not possible if we don't know much about the function.

But this is exactly the philosophy of Lagrange Multipliers

In this course we learn to follow the analytic form of Lagrange Multipliers.

In Gradient Ascent/discent we have to go along the Gradient and now we will also have constrains (defines feasibility space)

** So we have to find points that have Gradients orthogonal to the feasibility space that is the point we should stop calculating Gradient.

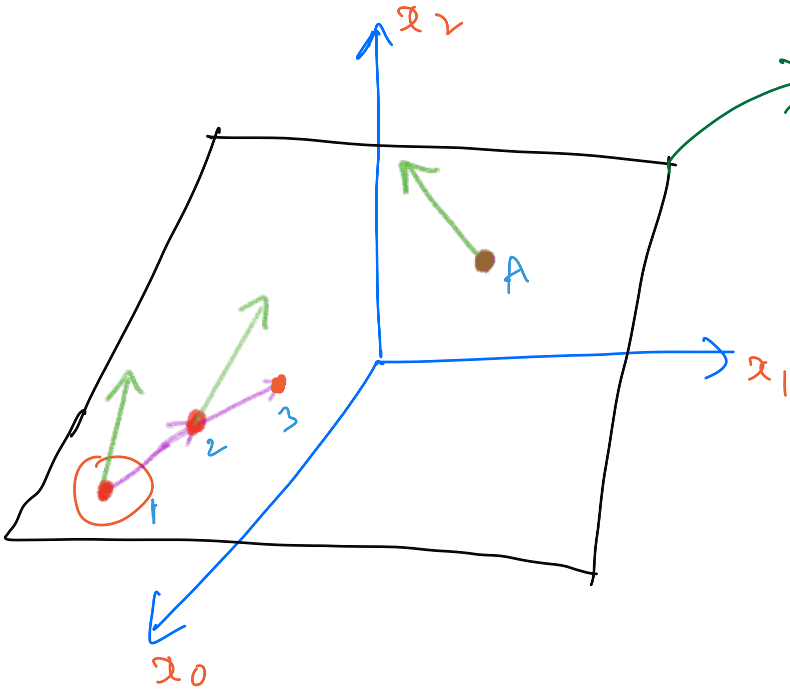
ex:

max $f(x)$

s.t a plane

A slightly more complecated example.

$f: \mathbb{R}^3 \rightarrow \mathbb{R}$



Feasibility Space is a plane in 3D

→ Since the points have to be on the Feasibility space we calculate the Gradient at a point and then take a next point which is the projection of Gradient on to the Feasibility Space and we proceed like this till we reach a point 'A' for which the Gradient is $\perp^{lar}$ to the feasibility space.

We know that the projection of Orthogonal vectors is Zero. What we really want is to calculate Gradient vector $\perp^{lar}$ to the Feasibility Space.

The condition that Gradient is $\perp^{lar}$ to the Feasibility space is given in the Lagrangian equation.

$$\boxed{\underbrace{\nabla_x f(x)}_{\text{Gradient of the Function}} = \underbrace{\sum_{j=0}^{m-1} \lambda_j \nabla_x E_j(x)}_{\text{Gradient of the Feasibility Space}}}$$

&

$$\underbrace{\nabla_x f(x)}_{\text{Vector}} = \underbrace{\sum_{j=0}^{m-1} \lambda_j \underbrace{\nabla_x E_j(x)}_{\text{vectors}}}_{\text{Linear Combination of vectors }\perp\text{ar to feasibility space}}$$

(Gradient of The Function)

Linear Combination of vectors $\perp$ ar to feasibility space

(There may be many vectors which are orthogonal to the feasibility space)

Note: (Tricky part) we know how to find orthogonal vectors for a plane. But for feasibility space which is not defined properly (on a curve (or a gibberish function.

25/11/23

It can be interpreted as the Gradient vector of the function is in the subspace of the linear combination of the vectors (set of the Gradients of the constrains)

Note: If all the vectors that are gradients of the constrains are ⊥lar to the feasibility space

$$\underbrace{\nabla_x f(x)}_{\text{vector}} = \sum_{j=0}^{m-1} \alpha_j \underbrace{\nabla_x E_j(x)}_{\text{vectors}}$$

orthogonal to the Feasibility Space.

⇒ Then the above equality requires that the "Gradient of our function" is the "Linear combination of vectors (constrain vectors) that are ⊥lar to the feasibility space".

We have not proved it yet, we will do it now

Imp: Earlier we wanted the Gradient of our function to be Orthogonal to the Feasibility Space.

** Therefore, now if we prove that the ** * Gradient of the function is the Linear combination of the Gradients of the constrain functions

(btw Gradient of function is a vector)

The equation of Lagrangian Equation we let us do this for the Non-smooth surfaces.

ex:

$$f : \mathbb{R}^2 \to \mathbb{R}$$

min|max $f(x)$

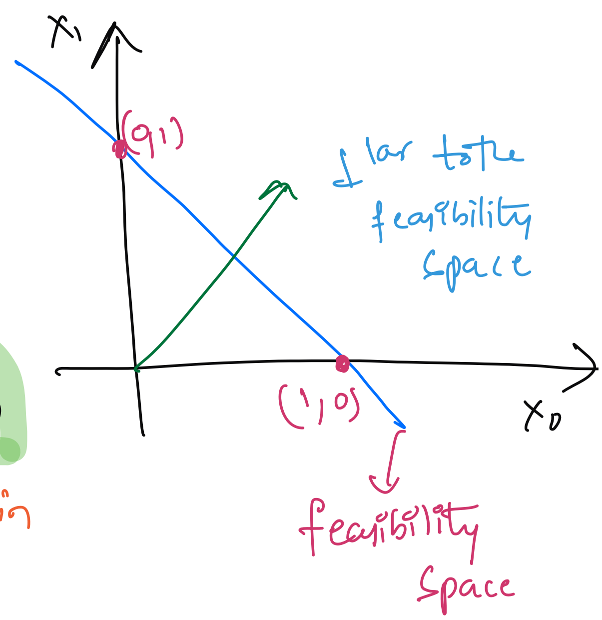
s.t $\boxed{x_0 + x_1 - 1 = 0}$

Single constrain

$f : \mathbb{R} \to \mathbb{R}$

$\Downarrow$

$$E_0[x] \doteq E_0\left(\begin{bmatrix} x_0 \\ x_1 \end{bmatrix}\right) = 0$$



'x' is a vector

$$\therefore \quad \nabla_x E_0\left(\begin{bmatrix} x_0 \\ x_1 \end{bmatrix}\right) = \begin{bmatrix} \dfrac{\partial E_0}{\partial x_0} \\[2ex] \dfrac{\partial E_0}{\partial x_1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\frac{\partial}{\partial x_0}(x_0 + x_1 - 1) \Rightarrow 1 + 0 + 0 = 1$$

$$\frac{\partial}{\partial x_1}(x_0 + x_1 - 1) = 0 + 1 + 0 \quad 1$$

→ let's go for a more complicated example

$$f: \mathbb{R}^3 \longrightarrow \mathbb{R}$$

(we can't plot this function, because it points from 3D to 1D, we need actually 4D plots to plot this)

But now if we have TWO constraints

$$E_0(x) = 0 \quad ?$$

$E_1(x) = 0$

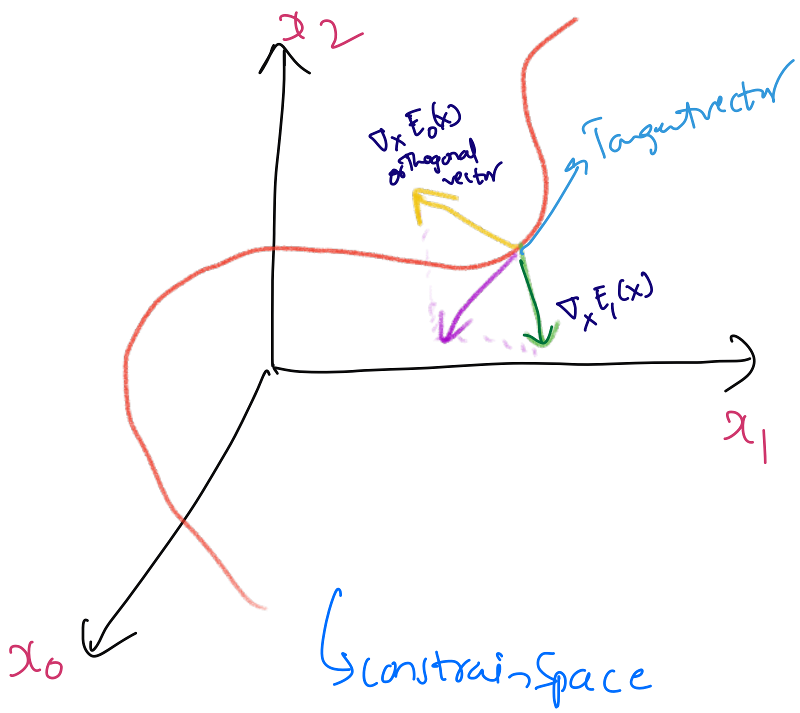
Then we can plot the Feasibility function in a 3D plot.

Note: The Function $f : \mathbb{R}^2 \to \mathbb{R}$ (requires 4D plot)

& constrains $f : \mathbb{R}^2 \to \mathbb{R}$ (requires 3D plot)

∵ (Two constrains)

If we assume feasibility space as a curved line in a 3D space as below.



∴ To draw a line in 3D space we need Two constrains

$$\therefore \left.\begin{array}{l} E_0(x) \\ E_1(x) \end{array}\right\} \text{Two constrains}$$

Now if we take Gradient of $E_0(x) = \nabla_x E_0(x)$

& Gradient of $E_1(x) = \nabla_x E_1(x)$

They would be two different vectors

$$\boxed{\nabla_x f(x) = \lambda_0 \nabla_x E_0(x) + \lambda_1 \nabla_x E_1(x)}$$

This how lagrangian multiplier equation works.

$|||^{ly}$ we can extend this concept to n-dimensional space

$$f : \mathbb{R}^n \to \mathbb{R}$$

$$E_j : \mathbb{R}^n \to \mathbb{R}$$ constraints

→ We can try and prove Lagrangian Equations by defining a superfunction

$$\boxed{\underbrace{L(\underbrace{x}_{\in \mathbb{R}^n}, \underbrace{\lambda}_{\in \mathbb{R}^m})}_{\text{Super Function}} = f(x) - \sum_{j=0}^{m-1} \lambda_j E_j(x)}$$

Superfunction: $\mathbb{R}^{n+m} \to \mathbb{R}$

combination degrees of freedom of function $f(x)$ i.e 'n' & lagrange multiplier 'λ'

$$\boxed{\nabla_{x,\lambda} L(x,\lambda) = 0}$$

Lagrange's system

The gradient of lagrange's s/s is zero!

let's have a look at the Equation of the system

$$\left[\begin{array}{l} \dfrac{\partial L}{\partial x_0} = \dfrac{\partial f}{\partial x_0} - \sum_{j=0}^{m-1} \lambda_j \dfrac{\partial E_j}{\partial x_0} = 0 \\ \dfrac{\partial L}{\partial x_1} = \dfrac{\partial f}{\partial x_1} - \sum_{j=0}^{m-1} \lambda_j \dfrac{\partial E_j}{\partial x_1} = 0 \\ \vdots \end{array}\right.$$

$$\vdots$$

$$\frac{\partial L}{\partial x_{n-1}} = \frac{\partial f}{\partial x_{n-1}} - \sum_{j=0}^{m-1} \alpha_j \frac{\partial E_j}{\partial x_{n-1}} = 0$$

$$L_x(x,\alpha)=0 \qquad \nabla_x f(x) - \sum_{j=0}^{m-1} \alpha_j \nabla_x E_j(x) = 0$$

This is the first part of Lagrangian s/s.

$$\frac{\partial L}{\partial \alpha_0} = \frac{\partial f(x)}{\partial \alpha_0} - \frac{\partial}{\partial \alpha_0}\left(\sum_{j=0}^{m-1} \alpha_j E_j(x)\right)$$

$$= 0 - E_0(x) = 0$$

||ly $$\frac{\partial L}{\partial \alpha_1} = 0 - E_1(x) = 0$$

$$\vdots$$

⋮

$$\frac{\partial L}{\partial \alpha_{m-1}} = 0 - E_{m-1}(x) = 0$$

equating all derivatives to zero

we have made above for our convenience the trick we used above

$$\therefore \nabla_{x,\alpha} L(x,\alpha) = 0$$

$$\therefore E_0, E_1, E_2 \ldots E_{m-1} = 0$$

$$\text{i.e } E_J = 0$$

Lecture-3

→ Extension of Lagrangian Formulation to **Complex Domain**.

→ Because sometimes the degrees of Freedom for 'x' in $f(x)$ are modelled by Complex

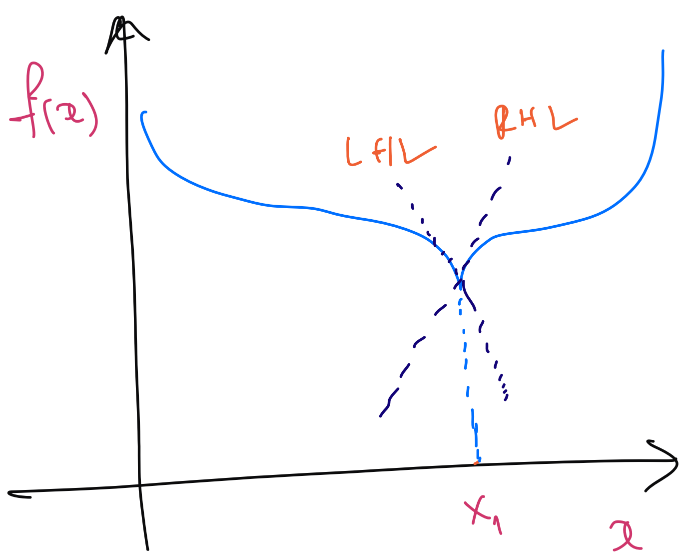
Numbers.

and Notion of Derivative concept in complex domain is not a straight forward concept.

** Our Aim is develop a way to go around the problem of Complex derivative so that we can apply "Lagrangian formulation" to Functions of Complex numbers with 'n' degrees of freedom

→ Review of Derivative

For $f: \mathbb{R} \to \mathbb{R}$ (Real domain)



f(x) cannot be derived

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

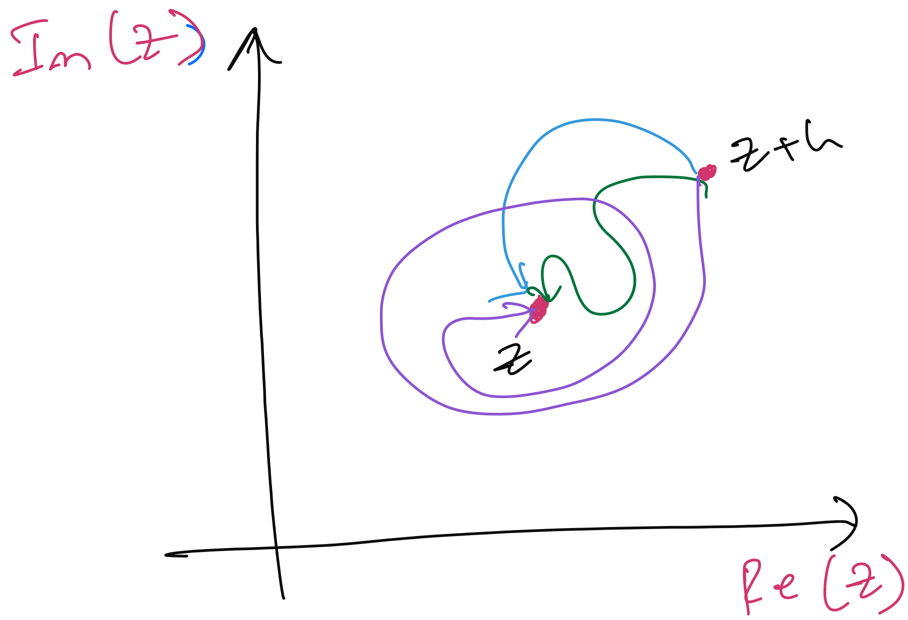
→ we can say the limit exists if the

$h^- \to 0$ & $h^+ \to 0$ The right hand limits are equal.

→ In the above figure the left hand derivative and right hand derivative are not same ∴ The derivative does not exist for $\boxed{x = x_1}$

→ When it comes to Complex Number, the notion of derivatives is completely different.

i.e $f : \mathbb{C} \to \mathbb{R}$ i.e $f(z) \to \mathbb{R}$



Now the whole Gauss plain is the Domain!
i.e the Complex Plain

$$f(z) \longrightarrow \mathbb{R}$$

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$ } The definition derivative remains the same

But we have different trajectories to approach 'z' from 'z+h' ve Examples have been drawn in the above figure.

* → Importantly, The point of existence of Derivative will boil down to what happens when we approach 'z' from Real direction & Imaginary direction Seperately.

→ Regrettably, there are quite common operations that we define in the Complex Domain that prevent the $\lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$ irrespective of the direction approaching 'z'

4 one of operation is Conjugate of 'z' i.e z^*

Because, Complex conjugate is implicitly involved in Division. z^* or $\bar{z}$ Cannot be Derived i.e $\frac{z_1}{z_2}$ are two complex Numbers

z_2 divides z_1

$$\Rightarrow \frac{z_1}{z_2} \times \frac{\bar{z_2}}{\bar{z_2}} \Rightarrow \frac{z_1 \times \bar{z_2}}{|z_2|^2}$$

This thing involves multiplication by conjugation and division by square magnitude of ~~z_2~~

i.e $$\frac{1}{h} = \frac{h^*}{|h|^2}$$

→ Conjugate form is such a problem that if we want to build the most elementary function i.e Quadratic form in complex domain.

$$z^{\dagger} A z$$

Complex Transpose & Conjugate

∴ z^{+} can not be Derived in Complex Domain.

→ Therefore Lagrangian method of optimization Cannot be directly applied to Complex Domain.

→

But, There is a work around to all this!

Brandwood method

Brandwood was an Engineer, who was an Antenna Designer.

He was designing an "Antenna Array" & the problem was to Linearly combine signal coming from many different Antennas and for him the degrees of Freedom w.r.t. which he wanted to minimize things was an Array of coefficients.

i.e. vector of coefficients

$$\min / \max \quad f(z)$$

s.t $E_j(z) = 0 \quad j = 0, \dots, m-1$

$f : \mathbb{C}^n \to \mathbb{R}^+$ often ≥ 0

$E_j : \mathbb{C}^n \to \mathbb{C}$

$$\boxed{\nabla_{z,\lambda} L(z,\lambda) = 0}$$

Derivative is impossible due to z^* because $f(z) = z^+ A z$

There is a method that goes through a special kind of derivative which is not a usual derivative. It is called either psuedo or seperate Derivative.

$\forall \; F : \mathbb{C} \to \mathbb{R} \longrightarrow F(z) \quad z = x + iy$

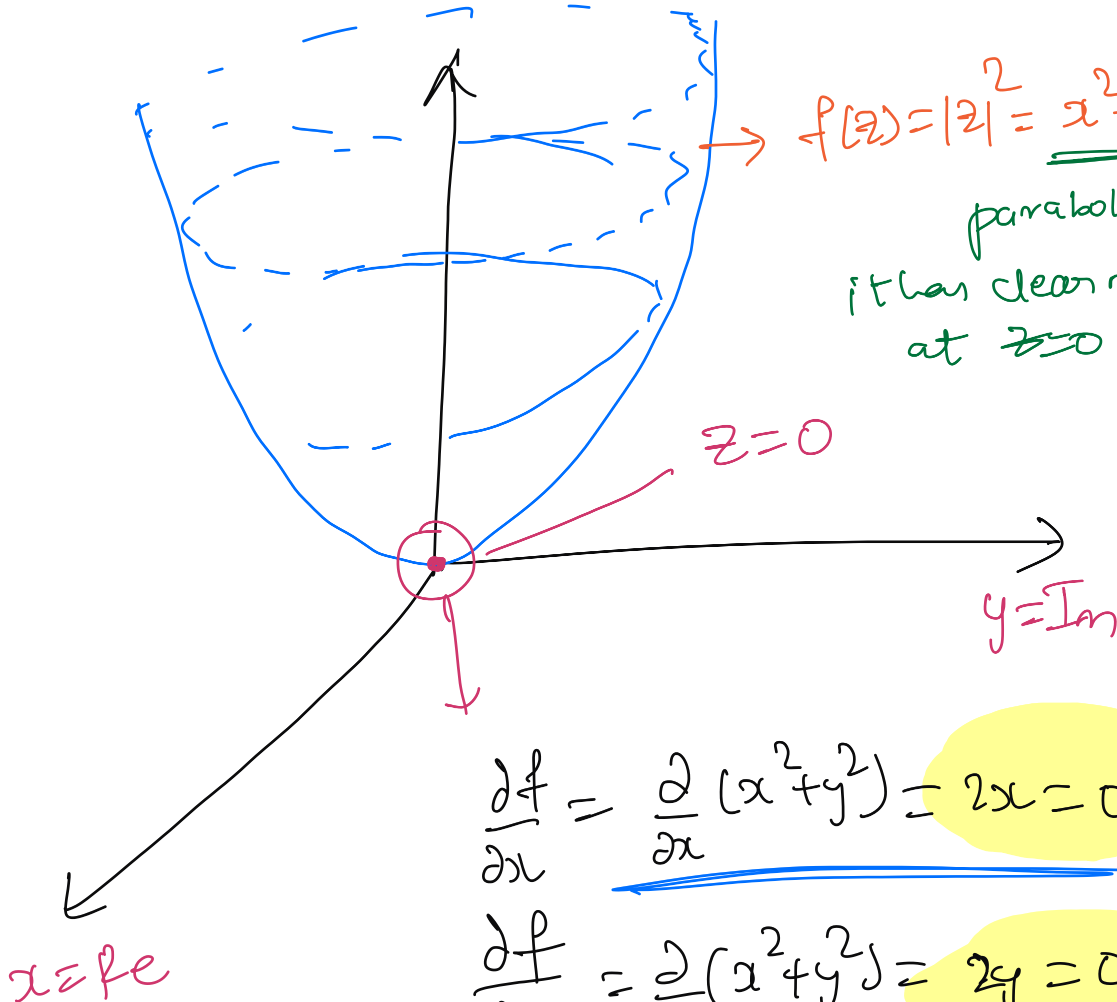
$\exists \; H : \mathbb{R}^2 \to \mathbb{R} \longrightarrow |z|^2 = x^2 + y^2$

There exists

$||$

$\exists\ G: \mathbb{C}^2 \to \mathbb{R} \to z^* z$

as indicated if $f(z) = |z|^2$



$f(z) = |z|^2 = \underline{x^2+y^2}$

paraboloid

it has clear minimum at ~~z=0~~

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(x^2+y^2) = 2x = 0$$

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(x^2+y^2) = 2y = 0$$

If we forget about complex Domain and look at $f(z)$ as a Function of x, y

i.e $f(z) = x^2+y^2$

then it is $f : \mathbb{R}^2 \to \mathbb{R}$

$$\therefore \frac{\partial f}{\partial x} = 2x = 0, \quad \frac{\partial f}{\partial y} = 2y = 0$$

i.e $x = y = 0$

This perfectly identifies with the complex number

$$\boxed{z = 0 \text{ i.e } x+iy}$$

Using the above observation

Even when the Derivative for complex function is not there, it is not there because the definition

$$f(z) = |z|^2 = z^* z$$

of $f(z)$ suffers from the existence of z^*

we can calculate the Derivative by splitting the Complex variable 'z' into its Real & Imaginary part.

And write the Gradient w.r.t to Real & Imaginary part and set that Gradient to Zero.

ex: If we have 20 CV's we will split them into 20 Real & 20 Imaginary parts. Both are real valued variables

So we have 40 Real variables

& Now, if I want to look for something i.e minima i.e derivative = 0
maxima

'Now, we have look for the derivative = 0 with '40' dimensional vector in Real Domain from '20' dimensional vector in Complex Domain.

This is perfectively feasible, but examples might seem to be complicated

ex: Our Quadratic form in Complex Domain

$z^{+} A z$

$$z^\dagger A z = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} A_{jk} z_j^* z_k$$

For this example to split every set into a Real & Imaginary part and to make the computation for gradient would be a complicated task and prone to error.

→ So, we will try to solve these kind of cases without leaving Complex Domain and it is feasible in some cases.

$z = x + iy$

$\forall \quad F : \mathbb{C} \longrightarrow \mathbb{R} = F(z)$

$\exists \quad H : \mathbb{R}^2 \longrightarrow \mathbb{R} = |z|^2$

$\exists \quad G : \mathbb{C}^2 \longrightarrow \mathbb{R} = z^* z$

$$H(z) = H(\mathrm{Re}(z), \mathrm{Im}(z))$$

$\|$ $\qquad\qquad$ $\|$

$$F(z) = G(z, z^*)$$

* Illustration of how H, F are defined

$F(z) = |z|^2 = z^* z$

1. $H(x,y) = x^2 + y^2$

- This function will give the same result of $f(z)$

$H(x,y)$ should be defined as above

2. $G(a,b) = ab$

$G(z^*, z) = z^* z$

which is the square module of 'z' which exactly as $f(z)$

Here, we are defining two function $H(x,y)$ & $G(a,b)$

So, whenever we have a function $f: \mathbb{C} \longrightarrow \mathbb{R}$

i.e $f(z) = |z|^2$

we can define both functions $H(x,y)$ & $G(a,b)$

a function using Real & imaginary part seperately

A further Auxillary function that helps distinguishing the Role of z & z^* into our expression.

$\therefore$ we can see that both functions

$$H(x, y) = G(z, z^*)$$

let's derive both sides w.r.to $\partial/\partial x$ & $\partial/\partial y$

$$\frac{\partial}{\partial x} H(x,y) = \frac{\partial G}{\partial a}\frac{\partial z}{\partial x} + \frac{\partial G}{\partial b}\frac{\partial z^*}{\partial x}$$

$$\frac{\partial}{\partial y} H(x,y) = \frac{\partial G}{\partial a}\frac{\partial z}{\partial y} + \frac{\partial G}{\partial b}\frac{\partial z^*}{\partial y}$$

using chain rule

we know $a = z, \; b = z^*$

& $z = x + iy$

$$\frac{\partial z}{\partial x} = 1, \quad \frac{\partial z}{\partial y} = i$$

$\therefore$ above partial diff equation will transform to

$$\frac{\partial}{\partial x} H(x,y) = \frac{\partial G}{\partial a}(1) + \frac{\partial G}{\partial b}(1)$$

$$\frac{\partial}{\partial y} H(x,y) = \frac{\partial G}{\partial a}(i) + \frac{\partial G}{\partial b}(-i)$$

$$\Rightarrow \boxed{\begin{aligned} \frac{\partial}{\partial x} H(x,y) &= \frac{\partial g}{\partial a} + \frac{\partial g}{\partial b} \\ \frac{\partial}{\partial y} H(x,y) &= i\left(\frac{\partial g}{\partial a} - \frac{\partial g}{\partial b}\right) \end{aligned}}$$

Solving for $\frac{\partial g}{\partial a}$ and $\frac{\partial g}{\partial b}$

$$\Rightarrow \begin{bmatrix} \frac{\partial}{\partial x} H(x,y) \\ \frac{\partial}{\partial y} H(x,y) \end{bmatrix} = \frac{\partial g}{\partial a} \begin{bmatrix} 1 \\ i \end{bmatrix} + \frac{\partial g}{\partial b} \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{\partial}{\partial x} H \\ \frac{\partial}{\partial y} H \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} \frac{\partial g}{\partial a} \\ \frac{\partial g}{\partial b} \end{bmatrix}$$

Multiplying by A^{-1} on both sides

$$\Rightarrow \frac{1}{-2i} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial x} \\ \frac{\partial H}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial a} \\ \frac{\partial g}{\partial b} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \dfrac{\partial G}{\partial a} \\[2ex] \dfrac{\partial G}{\partial b} \end{bmatrix} = \begin{pmatrix} -\dfrac{1}{2i}(-i) & -\dfrac{1}{2i}(-1) \\[2ex] -\dfrac{1}{2i}(-i) & -\dfrac{1}{2i}(1) \end{pmatrix} \begin{bmatrix} \dfrac{\partial H}{\partial x} \\[2ex] \dfrac{\partial H}{\partial y} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \dfrac{\partial G}{\partial a} \\[2ex] \dfrac{\partial G}{\partial b} \end{bmatrix} = \begin{bmatrix} +\dfrac{1}{2} & +\dfrac{1}{2i} \\[2ex] +\dfrac{1}{2} & -\dfrac{1}{2i} \end{bmatrix} \begin{bmatrix} \dfrac{\partial H}{\partial x} \\[2ex] \dfrac{\partial H}{\partial y} \end{bmatrix}$$

$$\Rightarrow \quad \frac{\partial G}{\partial a} = \frac{1}{2}\left[\frac{\partial H}{\partial x} + \frac{i}{(i)\,i}\frac{\partial H}{\partial y}\right]$$

$$\frac{\partial G}{\partial b} = \frac{1}{2}\left[\frac{\partial H}{\partial x} - \frac{1\,(i)}{i\,(i)}\frac{\partial H}{\partial y}\right]$$

$$\boxed{\begin{aligned} \frac{\partial G}{\partial a} &\Rightarrow \frac{1}{2}\left[\frac{\partial H}{\partial x} - i\,\frac{\partial H}{\partial y}\right] \\ \frac{\partial G}{\partial b} &= \frac{1}{2}\left[\frac{\partial H}{\partial x} + i\,\frac{\partial H}{\partial y}\right] \end{aligned}}$$

→ For stationary point i.e Maxima (or) Minima

$$\frac{\partial G}{\partial a} = \frac{\partial G}{\partial b} = 0$$

which would imply that $\frac{\partial H}{\partial x} = \frac{\partial H}{\partial y} = 0$

* So requiring the Gradient w.r.to Real & Imaginary (x) (y) part is zero

is the same as $\frac{\partial G}{\partial b} = \frac{\partial G}{\partial a} = 0$

$$G(a,b) = ab$$

$$\frac{\partial G}{\partial a}(a,b) = \frac{\partial}{\partial a}(ab) = b = z^* = 0$$

$$\frac{\partial G}{\partial a}(a,b) = \frac{\partial}{\partial b}(ab) = a = z = 0$$

$z = 0$

This is for Stationary point

If $z = 0$ & $z^* = 0$ ⇒ $z = 0$

Now we figured How to split get the derivative if we want to calculate the derivative of the complex function at $z=0$ eventhough there is no derivative in the complex domain.

Back to Lagrange's Multipliers.

Back to Lagrange's multipliers

$$L(z,\lambda) = f(z) - \sum_{j=0}^{m-1} \lambda_j^{Re} \, Re\{E_j(z)\} + \lambda_j^{Im} \, Im\{E_j(z)\}$$

$z \in \mathbb{C}^n$, $\lambda \in \mathbb{R}^{2m}$

$$E_j(z)=0 \iff \begin{matrix} Re\{E_j(z)\}=0 \\ Im\{E_j(z)\}=0 \end{matrix}$$

$$E_j(z)=0 \iff \mathrm{Re}\{E_j(z)\}=0$$

$$\mathrm{Im}\{E_j(z)\}=0$$

$$\lambda_j = \frac{1}{2}\left(\lambda_j^{\mathrm{Re}} + i\lambda_j^{\mathrm{Im}}\right)$$

$$L(z,\lambda) = f(z) - \sum_{j=0}^{m-1} \lambda_j^* E_j(z) + \lambda_j E_j^*(z)$$

$\in \mathbb{C}^n$ $\in \mathbb{C}^m$

$$2\,\mathrm{Re}\left\{\lambda_j^* E_j(z)\right\} = 2\left(\mathrm{Re}\{\lambda_j^*\}\,\mathrm{Re}\{E_j(z)\} - \mathrm{Im}\{\lambda_j^*\}\,\mathrm{Im}\{E_j(z)\}\right)$$

$$= \lambda_j^{\mathrm{Re}}\,\mathrm{Re}\{E_j(z)\} + \lambda_j^{\mathrm{Im}}\,\mathrm{Im}\{E_j(z)\}$$

$$\tilde{\nabla}_{z^* \lambda^*} L(z,\lambda) = 0$$