Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

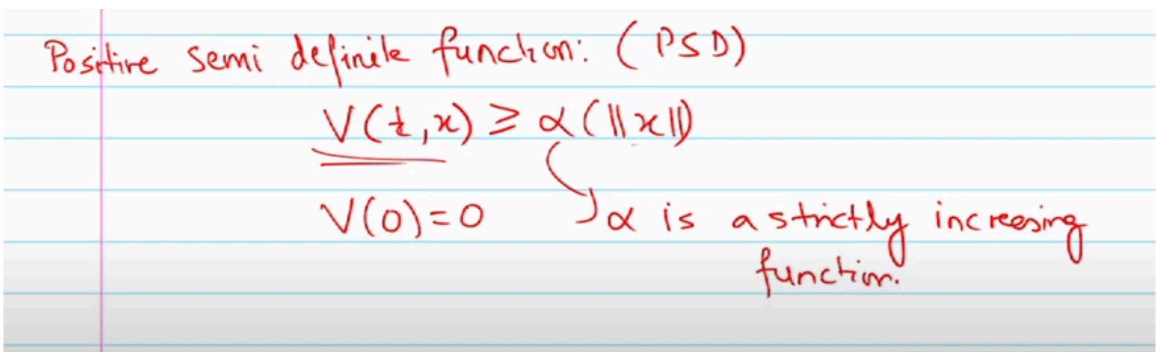
Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-4

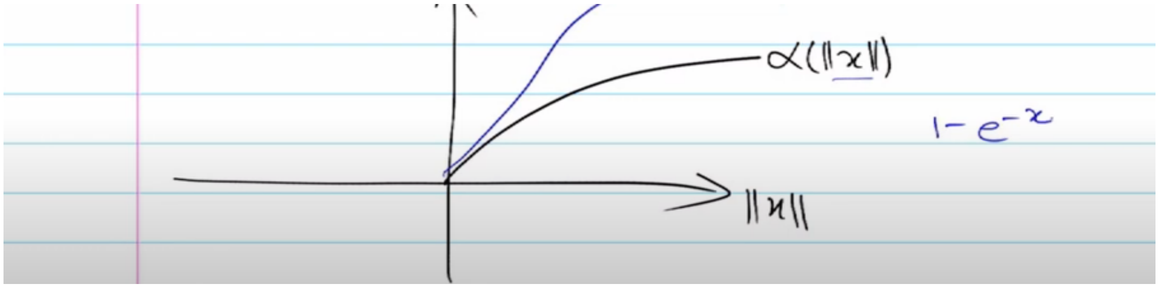
Lecture-15: Connections to optimization problems

Thank you very much. So, what is a positive definite function, a positive semi-definite function? So, we say a function V and can be function of t and x, this is positive semi-definite. If this holds true where alpha is any alpha is a strictly increasing function. So, meaning if I can lower bound this function by another strictly increasing function which increases with the norm of x. So, then this function is positive semi-definite. We also need that V of 0 is 0.



So, at the equilibrium this function is 0. So, the I mean, so we are trying to construct something called Lyapunov function. So, energy should be 0 at the equilibrium. So, V of 0 is 0 and this holds true right.

So, that means alpha as norm of x is equal to 0, because origin is the equilibrium and this function also evaluates to 0 right, because it is lower bounded and strictly. So, it is a positive function and it looks something like this. So, an alpha your alpha function can look something like this right. So, norm of x this is alpha times this is alpha of norm of x this function and your positive this particular function v can be any function on above it right. So, this function as norm of x goes to infinity this function need not go to infinity.



So, this like for instance 1 minus e to the negative x something like this right. So, this function need not go to infinity ok, but if x goes to infinity implies that this function also goes to infinity then we call it positive definite then V is positive definite. So, because this V is lower bounded by this particular strictly increasing function and as norm of x goes to infinity V is I mean V is always up like basically upper bounds this function. So, V has to go to infinity right. So, this imparts something called radial unboundedness to your function.

If $\|x\| \to \infty \Rightarrow \alpha(\|x\|) \to \infty$, then V is Positive Definite (Radial unboundedness)

It is a radial unboundedness property. So, as you move away from the equilibrium or the origin. you grow larger and larger and keep growing larger. So, it is basically it starts growing unbounded and an example would be V of x is if I choose it to be half x square, this is a positive definite function. Because it is x norm of x goes to infinity whether it is plus infinity minus infinity V of x also goes to infinity right and V of 0 is 0 ok.

$$V(x) = \frac{1}{2}x^2$$

So, this is an example of positive definite function. alpha is a function, it is not a constant as just like f of x you write it and alpha is a function, it is a strictly increasing function. So, V is negative definite if minus V is positive definite. So, why do we care about the negative definiteness? Positive definiteness I mean you understand that we want this like as the system is far like as let us say the x is farther away from the equilibrium your I mean the system is likely to have more and more energy right. So, that is the positive definiteness part.

Why do we care about the negative definiteness? Yeah. So, the time derivative of that particular function we want that to be act like a negative definite function right. or at least closer to negative definite function. So, that we can say that that is going to decrease and eventually go to going eventually go to 0 ok. So, using these ideas, so this kind of function V is called a Lyapunov function and using this idea we can characterize whether a system is stable or whether an equilibrium sorry whether an equilibrium is stable asymptotically stable exponentially stable and so on.

So, it is stable in the sense of Lyapunov if V of x is greater than equal to 0 for every x or greater than equal to 0 is fine or you can say it is V of x is greater than yeah greater than 0 for every x ok, V of 0 is 0. and V dot x is less than equal to 0 ok. So, if that is the case then we say that the equilibrium is stable why because whatever energy you have we know that at least we are not putting in more energy right. So, V dot is negative definite or negative semi definite rather. So, it may we may still be at the same energy level, but we are not going to be increasing the energy of this system.

* <u>Stable</u> If $V(x)>0 \quad \forall x \setminus \{0\}$

$V(0)=0$

$\dot{V}(x) \leq 0$

So, we are not going to go unboundedness unbounded and since stability, as I said intuitively is intuitively related to boundedness. So, this is going to be stable the equilibrium is going to be stable this is clear. So, that would be that would be one way to characterize a stability of an equilibrium whether it is stable. What about asymptotic stability? if we can find a function Lyapunov function V for asymptotically stable equilibria, if first of all V is positive definite. So, meaning that V of x is greater than 0, V 0 is 0 and what about V dot? Strictly less than 0 for all x So, then that means we are always decreasing the energy of the system and this would be something called asymptotically stable or locally asymptotically stable, but we can provide globally global guarantees if V is radially unbounded.

* Asymptotically <u>Stable</u>: If $V(x)>0 \ \forall x \setminus \{0\}$

$V(0)=0$

$\dot{V}<0 \quad \forall x \setminus \{0\}$

Locally AS

So, the moment it becomes radially unbounded then we can provide global guarantees, but this is how you sort of characterize the asymptotic stability of the equilibrium point ok. What about exponential? So, again we want V to be positive, so this thing is again there. What about V dot? Let us say I get V dot is nothing but some alpha V. So, would this imply the equilibrium to be exponentially stable? Not, even if it is class k function, would this imply exponentially stable? So, yeah, first of all alpha is greater than 0, let us just assume this, sure.

* Exponentially Stable (ES):

$$V(x) > 0 \quad \forall\, x \setminus \{0\}$$
$$V(0) = 0$$
$$\dot{V} \leq -\alpha V \;, \quad \alpha > 0$$

No. So, it would imply that V is exponentially stable or V converges exponentially fast. It does not say anything about x. V is a function of x, right? So, all it says is that V converges exponentially fast or decays exponentially fast, but it does not say anything about what our x is going to look like.

Second V. same V. So, I mean had this been directly in terms of x then you would have said that x converges x decays exponentially fast or x converges to equilibrium exponentially fast. This I mean all we can say about this is that V can V as a function of x converges to V 0 exponentially fast, but it does not directly translate to x converging to equilibrium exponentially fast right. I mean it can be an arbitrary function of x and x may be x may just be oscillating a lot even though it is converging. So, all you can say that x is like I mean the equilibrium is asymptotically stable and not even as. So, if this is the condition you can only say that it is just with these you cannot guarantee that it is exponentially stable all you can say is it is still asymptotically stable.

So, there are more things that we need to show or in order to guarantee that an exponential convergence in V would also imply an exponential convergence in x and for that these conditions are. So, this does not imply that x t converges exponentially first. the guarantee it only guarantees exponential convergence of V right. So, then what do we need to show in order to guarantee that this would like I mean a condition in V would also imply like an exponential convergence in V would also imply an exponential convergence in x and for that, we first have to show that V of x is sandwiched between they using alpha 1 and alpha 2, let me use those. So, it is sandwiched between, so again alpha 1 and alpha 2 greater than 0, v dot that is less than equal to some alpha 3 x square.

$$\alpha_1\|x\|^2 \le V(x) \le \alpha_2\|x\|^2 \qquad \alpha_1, \alpha_2 > 0$$

$$\dot{V} \le -\alpha_3\|x\|^2$$

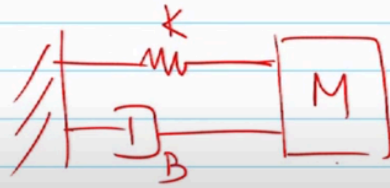
$$\|x(t)\| \le \underline{m}\, e^{-\underline{\alpha}(t-t_0)} \qquad m \le \left(\frac{\alpha_2}{\alpha_1}\right)^{1/2} \qquad \alpha \ge \frac{\alpha_3}{2\alpha_2}$$

if these two are satisfied and obviously V 0 is I mean since x is equal to 0 and it is sandwiched between alpha 1 and alpha 2 like this. So, V 0 is going to be 0. So, if these conditions are satisfied then you can say that trajectories they converge exponentially fast or x converges exponentially fast. So, one of the conditions or the definition for exponential stability of the equilibrium was x t was m e to the negative alpha t minus t naught right and how does this m and alpha related to these coefficients alpha 1 alpha 2 and alpha 3. So, you can show that m is going to be less than equal to alpha 2 over alpha 1 power half and alpha is going to be greater than equal to alpha 3 over 2 alpha 2, but then only if V satisfies these conditions that is when you can guarantee that x also converges exponentially fast.

So, as I said finite and fixed time stability we are going to look at it in due course of time. So, because those are sort of relatively recent, but then in the most of the background on stability and sort of marrying stability of dynamic of equilibrium of dynamical systems to convergence behavior in optimization algorithms. So, that would be more or less clear from what we have looked at so far. So, let us look at a few examples. So, a spring mass damper system I believe everyone is.

So, you have spring of like spring constant k, we have a damper with coefficient v and a mass with m. And let us say we denote by q the position of this mass and q dot is the velocity right. So, what could be a good choice or a candidate for Lyapunov function here? So, in some sense energy of the system and that can be half k q square plus half M q dot square right? And what is the dynamical system underlying dynamical system here? So, m q double dot plus B q dot plus k q equal to 0 ok. So, that is the dynamical system that we have right. So, how do we bring this first of all how do we bring this dynamical system in a form x dot is equal to g x.

Ex:



$$\left.\begin{matrix} q \\ \dot{q} \end{matrix}\right\} \quad V = \frac{1}{2}Kq^2 + \frac{1}{2}M\dot{q}^2$$

$$M\ddot{q} + B\dot{q} + Kq = 0$$

$$\dot{x} = g(x)$$

So, what is going to be x here? q and q dot. q and q dot right. So, that would be q and q dot. So, d by dt of x or x dot would be So, you get q dot here, you get q double dot here, and q double dot you can write it as minus K over m times q minus B over m times q dot. So, this is x dot this whole thing and this is g of x.

$$x = \begin{bmatrix} q \\ \dot{q} \end{bmatrix} \qquad \underbrace{\frac{d}{dt}\begin{bmatrix} q \\ \dot{q} \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} \dot{q} \\ -\frac{K}{M}q - \frac{B}{M}\dot{q} \end{bmatrix}}_{g(x)}$$

So, what is the equilibrium of this particular dynamical system? 0 comma 0 right, why? Because when you said g of x equal to 0 that means q dot is equal to 0 and if q dot is equal to 0 this term is 0 and because this whole term is also 0 that means q is also 0 right. So, x equal to 0 0 is the equilibrium, so that is the equilibrium, so origin is the equilibrium.

$$g(x) = 0 \Rightarrow x_e = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Now let us try and see what kind of stability guarantees we can provide for this particular equilibrium. Is it stable, is it asymptotically stable, is it exponentially stable and so on. So, what do we need to do in order to characterize the stability of this equilibrium? We have to look at how V dot behaves right.

First of all is it a positive definite function right because I can write V as half q q dot you have k here right and this matrix is positive definite matrix because k and m are greater than 0. So, this is positive definite matrix. So, this is a positive definite function.

In fact also radially unbounded right as q dot norm of q dot goes to 0 and norm of q goes to 0 this grows radially unbounded. So, this is radially unbounded as well.

$$V = \frac{1}{2} K q^2 + \frac{1}{2} M \dot{q}^2$$

So, we have all the nicer properties and V of 0 is also 0. So, we have all the nicer properties. So, V is a positive definite function. Let us see what how V dot looks like. So, again V is half k q square plus half M q dot square.

V dot would be and this particular term M q double dot we can write it in terms of the. So, this would be q dot times. So, what is m q double dot? That is this particular term over here right that becomes and this basically gives us minus B q dot square which is negative semi-definite right. So, this is you cannot say this is less than 0 strictly less than 0 for this is in fact less than equal to 0, why less than equal to 0 and not strictly less than 0 for every let us say for every x.

$$\begin{aligned}\dot{V} &= K q \dot{q} + \underline{M \dot{q} \ddot{q}} \\ &= K q \dot{q} + \dot{q}(-B\dot{q} - Kq) \\ &= -B\dot{q}^2 \underline{\leq 0} \qquad \forall\, x \setminus \{x_e\}\end{aligned}$$

Yeah. So, it is only a quadratic function of only q dot right. So, that means no matter what your q is you need not be the equilibrium point. So, there are points of the form q comma 0 such that this particular term is still equal to 0. This is negative semi-definite and not negative definite. So, all we can say is it is less than equal to 0.

So, V is positive definite, V dot is less than equal to 0. So, therefore the equilibrium is as of now we can just say that the equilibrium is stable. So, V is basically greater than 0 for every x not equal to x equilibrium, V of 0 is 0 and V dot is less than equal to 0 for every x. So, with these three, we can only guarantee stability. From our observation, we know that this system is, the equilibrium is not just stable, but it is going to be at least asymptotically stable, right? So, we know that it just does not stay bounded, but it also eventually converges to its equilibrium point, right.

So, how can we argue something like that from here? So, the only way if I try to find points such that V dot is identically equal to 0. So, for that to happen this term is identically equal to 0. So, if q dot is identically equal to 0 right, if q dot is identically equal to 0. So, for all time t. So, the only way that, so it may happen that V dot at one particle.

So, the whole reason, the reason that V dot is less than equal to 0 with V dot less than equal to 0 we can only claim stability is because I mean it may happen that. So, momentarily it hits 0, but if it is again starts becoming negative then we are fine right because it would in the next time instant it would still be decreasing. the only issue would come up if it becomes identically equal to 0 that it hits a particular point and on that point it becomes identically equal to 0 which is not the equilibrium point, right. So, if V dot, if Q dot becomes identically equal to 0, so that means for this to happen if Q dot is identically equal to 0, so like the x dot is equal to 0 only if Q is equal to 0, right.

because q dot is identically equal to 0. So, q is equal to 0. So, that means even so using invariance principle you can guarantee that this is going to be asymptotically because even for points when it is. So, even for points in like this which for which which are not equilibrium, but this thing is momentarily equal to 0. It basically soon escapes that particular point and it is again starts becoming negative. So, from there you can you can argue that it I mean the equilibrium is asymptotically stable.

So this is one particular choice of Lyapunov function that we worked with. So let me choose a different, so same example, same problem, but now we choose a slightly different Lyapunov function and let us see if we can guarantee something more. So now this time we are going to be choosing V of x to be half q q dot So, some small epsilon m, epsilon times m. So, depending on the value of epsilon, I can always make this function positive definite, right. I can find a value of epsilon such that this function is positive definite.

Ex: $V(x) = \frac{1}{2}\begin{bmatrix} q & \dot{q} \end{bmatrix}\begin{bmatrix} K & \epsilon M \\ \epsilon M & M \end{bmatrix}\begin{bmatrix} q \\ \dot{q} \end{bmatrix}$

$\rightarrow$ V is PD

So, V is going to be positive definite, radially unbounded. So, all the nicer properties are there. Now, let us now take the, consider V dot. And if you take V dot here, you can show that this Vdot turns out to be and again for a suitable choice of epsilon, you can make this function negative definite V dot to be negative definite. So now you see that by choosing a different Lyapunov function altogether, the same equilibrium which at best we could guarantee to be asymptotically stable, we can actually now in fact guarantee it.

$$\dot{V} = -\begin{bmatrix} q & \dot{q} \end{bmatrix} \begin{bmatrix} \epsilon K & \frac{1}{2}\epsilon B \\ \frac{1}{2}\epsilon B & B-\epsilon M \end{bmatrix} \begin{bmatrix} q \\ \dot{q} \end{bmatrix}$$

Why? Because first of all, if I look at the maximum eigenvalue of this particular matrix here, it is positive definite matrix. So V of x is going to be bounded by alpha 3 which is going to be lambda max of this let us let us call this matrix A ok. And v dot and if I call this matrix let us say delta. So, v dot is going to be less than equal to.

$$\lambda_{min} \|x\|^2 \leq V(x) \leq \lambda_{max}(A) \|x\|^2 \longrightarrow \text{E.S.}$$

$$\dot{V} \leq -\lambda_{min}(\Delta) \|x\|^2 \qquad \text{Exponentially Stable}$$

So, minus yeah. So, the minimum if I like it is minus lambda. So, lambda of this So, it has this kind of thing right and now you can we can guarantee that it is in fact it is going to be equilibrium is going to be as exponentially stable ok. So, the same equilibrium for which we could by choosing a different Lyapunov function we could says more about it right. So, now we can say that not only it converges asymptotically fast in fact it converges exponentially fast.

So there is a stronger convergence rate. And in order to obtain that, we had to choose this Lyapunov function. And as I said, Lyapunov, for his PhD work, he was trying to come up with a way through which you can construct these Lyapunov functions. But I mean, it's still an open problem, constructing this Lyapunov function in for a general system where you can always guarantee, like even if you know, let's say it's exponentially stable equilibrium. I mean, to be able to find a Lyapunov function through which you can show that. that is something very take very tricky in general and it comes from experience yeah.

So, yeah we constructed in such a way that so that v dot turns out to be negative definite. Yeah, there is no physical significance yeah, there is no physical significance. So, that is why I mean it is difficult to come up with a suitable choice of Lyapunov functions because I mean it is I mean there is some physical significance in the sense that if you look at the half k q square and half m q dot square those are still going to be there, but there is going to be some cross terms as well between q and q dot right. and it does not come like there is no physical intuition behind it. Maybe there is there will be a physical intuition if you sort of transform these points in another like using some kind of transformation you transform them into another space where and construct the synergy function in that space then it would it may make sense, but yeah in general no.

There is no I mean in this case in this example at least there is no direct physical significance. So, this basically also explains the difficulty of choosing how to like choose a good suitable Lyapunov function right because it is like as you can see that there is no intuition behind I mean. So, one would like in this case I believe people would have arrived at this particular Lyapunov function knowing that it is exponentially I mean the equilibrium is exponentially stable and then trying to construct using I mean sort of basically a converse way to construct just trying out different combinations and it is so worked out.

for this particular case. So, that is another challenge. So, I mean at least to your in the context of optimization it would not be too difficult to choose Lyapunov functions because there are not many choices honestly for convex optimization. So, what could be a good proxy? So, coming back to optimization and to start with let us just focus on simple gradient flow. So, to start with we are going to focus on gradient flows which are of the form x dot So what could be a good choice for Lyapunov function? So let us also assume some structure. So let us say that well f is mu strongly convex. So what is one particular characterization of first order condition for convexity that the gradient of f should be 0, right? So a good choice of Lyapunov function can be half right.

And now if I take v dot, if I consider v dot, so that becomes gradient of f transpose h n f times x dot, ok. Why? Because I can write this as half gradient f transpose gradient f and the derivative of the gradient f is h n. So, this is what it would be, right. So, the function is mu strongly convex, we know that. So, since f is mu strongly convex this implies Hessian of f mu times identity right.

Gradient Flows: $\dot{x} = -\nabla f(x) \longrightarrow f$ is μ-SC

$$V = \frac{1}{2}\|\nabla f\|^2 = \frac{1}{2}(\nabla f)^T(\nabla f)$$

$$\dot{V} = (\nabla f)^T(\nabla^2 f)\dot{x} = -(\nabla f)^T(\nabla^2 f)(\nabla f)$$

Since f is μ-SC $\Rightarrow \nabla^2 f \succeq \mu I$

$$\dot{V} \le -\mu \underbrace{\|\nabla f\|^2}$$

So, this would mean, so we will come to this part later, but using this dynamical system what is x dot equal to negative of gradient of f. So, I can write this as minus gradient f transpose Hessian f times gradient f. And if I use this property now, so this is nothing but v dot is less than equal to minus mu times right and what is this term equal to? Twice of v right, twice of the Lyapunov function. So, what can we argue? So, v dot is less than equal to minus 2 mu v. So at least we know that the Lyapunov function converges exponentially fast, that is something that we know.

$$\dot{V} \le -2\mu V$$

What about the convergence of x, exponential convergence of x to z like let us say x to x star, how can we guarantee that? So, lower bound is like from here. So it is not that straightforward. So unless let us say you assume that f is also L-smooth. So you get bound one way.

If you assume f is mu is strongly convex, you get bound another way. And then from there you can potentially guarantee that the convergence to x star is also going to be exponentially fast. But without assuming L-smoothness, as of now I mean we cannot guarantee at least from this analysis, that this is going to be exponential like convergence of x to x star is going to be exponentially fast, but convergence of gradient f to 0 that is going to be exponentially fast. So, this is one particular candidate of Lyapunov function. What else can we think of? Let us say now I assume. this time we assume that f is not strongly convex, but f satisfies PL Inequality.

* Assume f satisfies PL-inequality.

$$\frac{1}{2\mu}\|\nabla f\|^2 \geq \mu\left(f(x)-f^*\right)$$

So, what is PL Inequality? Now I am not saying it would not converge to the equilibrium, x would converge to the equilibrium or the optimal solution, but does x converge exponentially fast to x star that we cannot say. From this exercise we can say that V converges exponentially fast to 0, right. And that would mean that gradient of f converges exponentially fast to 0, but it does not say anything about x converging exponentially fast to 0. So this was the if I look at, if I assume the setting where f is satisfies PL inequality with coefficient mu.

$$V = f(x) - f^*$$

$$\dot{V} = \nabla f^T \dot{x}$$

$$= (\nabla f)^T(-\nabla f) = -\|\nabla f\|^2$$

$$= -\frac{1}{2\mu}(2\mu)\|\nabla f\|^2$$

$$\leq -2\mu\left(f(x) - f_*\right)$$

$$\dot{V} \leq -2\mu V$$

So, this turns out to be mu f of x minus f star right. So, now if I look at this particular term, is this term greater than equal to 0? It is right. At the optimal value it is going to be 0, but other than optimal it is going to be greater than equal to 0. That can be another suitable choice of Lyapunov function right and in this case because we know that f satisfies peer inequality it may actually make sense to choose this kind of Lyapunov function.

So let us do that. So let us simply choose this as your Lyapunov function. It is equal to 0 at x star. It is greater than 0 like I mean assuming that the PL-inequality is there with unique minimizer. It is strictly greater than 0 outside of that right. So what is V dot then? gradient of f transpose x dot, which is gradient of f transpose negative of gradient of f, which is basically saying this particular term.

Now, I can write this as minus 1 over 2 mu and then I also multiplied by 2 mu. and this is going to be less than equal to minus 2 mu times this particular term, which by definition is or PL inequality this is what it is right. And I get V dot is less than equal to minus 2 mu V, the same condition right. So, we know that V converges exponentially fast ok. So, this pretty much gives you an idea depending on the kind of assumptions that you make on the function, this basically gives you an idea as to what kind of Lyapunov function that you can potentially work with. So, in this case and obviously I mean we want this property that V should be equal to 0 at the optimal solution.

So, there are not many ways through which you can characterize optimality, one is either through this or something like this. So, in most cases in fact you are going to see that these are the Lyapunov functions that we are going to be working with when designing or analyzing a particular optimization algorithm. Is this clear? So, with this I would like to end today's lecture unless there are any other questions. Sorry, yeah. So, in order to guarantee exponential convergence of x to x star, we would need to assume that, we would need to assume L-smoothness.

Yeah, yeah. . Thank you. Now, wait why mu square? So, this is the sorry PL inequality that is my bad. I had already included mu here. So, this is your PL inequality. So, yeah this is the PL inequality here.

because f of x is always going to be greater than equal to f star right. Only at the optimal value it is going to achieve minimum everywhere else it is going to be greater than equal to minimum right. So, let us say f of x is x square or let us say f of x is x minus 1 square whole square. So, at x equal to 1 it is going to be 0 everywhere else it is going to be something greater than 0 right. So, f of x is always going to exceed f star.

So, then it is positive semi-definite Yeah f star is the optimal. So, that is the definition of PL inequality we consider the optimal f's. But then you it is PL inequality would still imply that you have a it is non-convex that is fine, but you have this kind of a invex function right. So, you take either you will have unique minimizer or you will have it to be a constant function if you look at this particular form. So, either it will be a unique minimizer like this or this a simple function like this would also a constant function is also satisfies PL inequality, you would not have too many cases. If it is PL inequality plus convex then it basically becomes the previous case which is strongly convex.

anything else? So, even for non-convex cases the characterization of local minima is through like gradient of f or being 0 or f of x, f of x minus f star being 0 and so on, right. Yeah, the same mu of strong convexity. Because every strongly convex function also satisfies PL inequality, right. With the same mu.

either single minima or even a constant function also such as I mean is a, but then here.
Thank you