Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

Computer Science & Engineering, Electrical Engineering, Mathematics

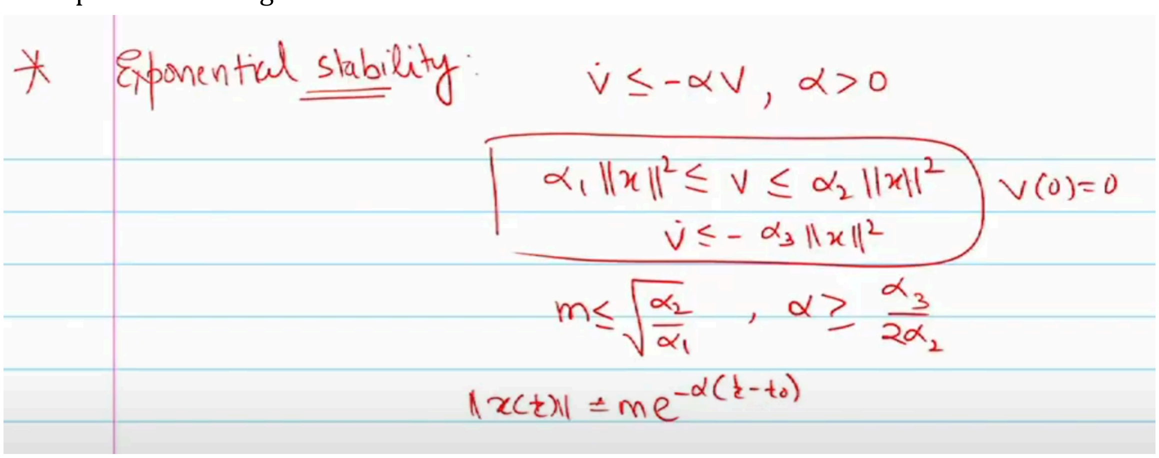
Indian Institute of Technology Bombay

Week-5

Lecture - 16: Exponential Stability

Alright, so before we start any questions on the last class on different forms of stability. Okay, so just to briefly recap. So, for exponential stability So again, when we talk about stability, we talk about stability of equilibrium points, right. So, for exponential stability v dot less than equal to some alpha v alpha greater than 0. Does that imply exponential stability of the equilibrium point? No, right. So, we needed more conditions and the conditions were of the form, let us say something like this.

or alpha 3. Then with exponents m which is going to be square root alpha 2 over alpha 1 and alpha which is going to be greater than alpha I think alpha 3 over 2 alpha 2. So, with this you can show if the Lyapunov function and the time derivative of it satisfies these inequalities. obviously assuming that origin is the equilibrium here, so v0 is 0, then xt essentially or the norm of it is given by something like this, where m is this particular term and alpha is this thing.



So, that is about exponential convergence of xt to the origin or the equilibrium. So, in the last class we looked at strongly convex function and also one particular case where we also looked at the function satisfying peer inequality and we showed that the Lyapunov function at least was exponentially convergent we could not tell more about the exponential convergence of the of x to the optimal solution. So, let us briefly revisit that example as well. So, we assume f is let us say mu strongly convex ok. So, with mu strong convexity what do we know? We know that the Hessian of f this is going to be lower bounded by mu times identity right.

So, if I consider a Lyapunov function f which is half square and why is this a valid Lyapunov function? Because only at x equal to x star your gradient vanishes and everywhere else it is greater than 0. So, v is greater than equal to 0 and v of x star is equal to 0. So, that means v of x star is equal to 0 and v of x is greater than 0 for every x. Is this clear? So, if I take the time derivative of this particular Lyapunov function that turns out gradient f transpose h in of f times x dot and x dot if I consider the gradient flow to be my underlying dynamical system which is this is your gradient this gradient flow. So, this particular v dot turns out to be minus gradient of f which is by strong convexity which is less than equal to mu times this thing whole square right or another way to write this is v dot is.

$*$ Assume: f is μ-SC $\quad | \quad \nabla^2 f \succeq \mu I$

$\boxed{\dot{x} = -\nabla f(x)}$ $\hookrightarrow$ GF

$V = \frac{1}{2}\|\nabla f\|^2$ $\quad$ only at $x = x^*$, $\nabla f(x^*) = 0$

$V(x^*) = 0$

$V(x) > 0 \quad \forall x \setminus \{x^*\}$

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

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

$$\leq -\mu \|\nabla f\|^2$$

$\dot{V} \leq -2\mu V \implies V$ converges exponentially fast

What about x?

So, if you multiply and divided this by 2. So, minus 2 mu times v because v by definition is this term. So, this is v dot is less than equal to minus 2 mu times v. So, with this we can at least say that v converges exponentially fast. what about x? So what about x? How can we claim or sort of maybe one way or the other that x also converges exponentially fast or does not converge exponentially fast? So if I look at the Lyapunov's condition for exponential stability, I somehow need to provide some kind of bounds on my Lyapunov function v, right.

v dot I think it is, I mean anyway in the process you are going to be getting these kind of bounds for v dot, but we also need to provide some bounds for v. And if v is simply defined in terms of f, so always think of strongly, strong convexity as a, in this case for instance that means a gradient in some sense they are at lower bounded by mu in like intuitively right. So, you get bound you can possibly get a bound one going one I mean one way right. So, if you want to get an upper kind of like upper bound you would have to assume more on the function right and that that means we would also need l smoothness of the function because otherwise we cannot like we need somehow need to use this particular inequality or at least get to this particular inequality with the strong convexity you can only possibly claim about this particular thing right. So, let us see in order like what else needs to be assumed in order to guarantee exponential convergence of x to x star ok.

Any questions on this alright. So, let us begin the lecture and let us try to argue that with L smoothness you can arrive at the exponential stability of simple gradient flow. So, just to reiterate we are working with this particular dynamical system with x star being the optimal or optimizer. So, that means gradient of f of x star is 0. So, we assume that f is mu strongly convex and l smooth.

So, what does these two conditions tell us? So, if f is mu strongly convex by definition of strong convexity, so f of y is greater than equal to f of x plus gradient f of x transpose y minus x plus mu over 2. So, this is the definition of or the first order condition for strong convexity right. Now, if I choose let us say choose y equal to x and x equal to x star. what is gradient of f at x star 0 right. So, that means we are left with this what we are left with it is f of x star which is also f star plus mu over 2 x minus x star whole square right.

So, that is first condition. So that is that basically comes from since f is mu strongly convex. So this is simply using the first order condition for strong convexity. Now we also know that f is L smooth. So if f is L smooth, so we know that f of y is less than equal to f of x.

GF: • $\dot{x} = -\nabla f(x)$ with x^* being the optimizer

$\nabla f(x^*) = 0$

• Assume that f is μ-SC and L-smooth.

Since f is μ-SC:

$$f(y) \ge f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2}\|y - x\|^2$$

Choose $y = x$

and $x = x^*$

$$\boxed{f(x) \ge f^* + \frac{\mu}{2}\|x - x^*\|^2} \quad — (1)$$

combing

plus gradient x transpose y minus x plus l y 2. That is something that we had already looked at when we were deriving the convergence or the order of convergence for gradient descent and heavy ball and other algorithms. So, again if I choose y equal to x and x equal to x star that means f of x is less than equal to f of x star. gradient of f at x star is going to be 0. So, this is going to be L by 2 x minus x star whole square.

f is L-smooth:

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2$$

$$\boxed{f(x) \leq f(x^*) + \frac{L}{2}\|x-x^*\|^2} \quad \text{--- } (2)$$

So, this comes from L smoothness. So, now if I combine these two, so let me if I combine, combining 1 and 2. So, what do we get f of x minus f star that is less than equal to L by 2 x minus x star whole square and greater than equal to mu by 2 norm x minus x star whole square right. And this is say if I end up choosing v to be like this. So, I have an upper and lower bound on v.

that is what we wanted to derive right. So, remember this is what we wanted to derive an upper and lower bound in this case because origin was the equilibrium. So, you did not have any x star here, x star was actually equal to 0. In this case, I mean x star is the equilibrium or the optimal solution. So, you have a non-zero x star possibly.

So, you have this kind of inequality, right. So, we are halfway there. So, we also need to derive an equivalent condition on v dot, right. So, let us see. So, if I choose v to be f of x minus f star, what is v dot? Gradient of f of x transpose x dot and x dot by definition is negative of gradient of f.

So, this becomes And we know that every strongly convex function satisfies PL inequality. So, that means, if I write this as 2 mu 1 over 2 mu, this is going to be less than equal to minus 2 mu times f of x minus f star. And this by definition is your Lyapunov function V. So, what do we get? minus 2 mu times v and we know that v is greater than equal to this particular term, because what do we want to derive? We want to derive v dot in terms of something less than equal to x minus x star whole norm square. So, for that we know that this thing is less than equal to, so your v is essentially greater than equal to.

v is greater than equal to this particular term. So, is this clear? Because this particular v which is f of x minus f star. So, that is greater than equal to this particular term. So, minus v is less than equal to this particular term minus of this term and that is what we are going to be replacing this with. So, what do we get? We get v dot is less than equal to minus mu square x minus x star whole square.

Combining ① and ②:

$$\boxed{\frac{\mu}{2}\|x-x^*\|^2 \le \underbrace{f(x)-f^*}_{V} \le \frac{L}{2}\|x-x^*\|^2} \quad \text{—} \quad ⓐ$$

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

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

$$= -\|\nabla f\|^2 = -2\mu \frac{1}{2\mu}\|\nabla f\|^2 \le -2\mu\big(\underbrace{f(x)-f^*}_{V}\big)$$

$$\dot{V} \le -2\mu V \le -2\mu\left(\frac{\mu}{2}\right)\|x-x^*\|^2$$

$$\boxed{\dot{V} \le -\mu^2\|x-x^*\|^2} \quad \text{—} \quad ⓑ$$

So, this is let us call this equation A and this is your B. So, we have derived the sufficient condition in order to show that x converges to x star exponentially fast right. We just need to find out the rates and again if I look at the rates. So, m was upper bounded by square root of alpha 2 over alpha 1. So, what is alpha 2 here? So, alpha 1 is mu over 2, alpha 2 is L over 2 and alpha 3 is mu square right.

Is everyone following this? So, m is less than equal to alpha 2 square root alpha 2 over alpha 1 and again you see that L over mu sort of plays a role here right. So, which is essentially equal to square root L over mu. And if you now look at your homework problem which is problem number 5 or 6, you have terms of the form square root L square root mu and you can see where it is actually coming from. what about rate alpha that should be greater than equal to alpha 3 over 2 alpha 2. So, alpha 3 is mu square 2 alpha 2 is L.

So, this basically is essentially mu square over L and from A and B you can show that x t minus x star is less than equal to m or rather square root L over mu e to the minus mu square over L t, let us say t naught is equal to 0. So, this is the exponential convergence on the optimizer. So, earlier we only had exponential convergence on the function value, but now we also show exponential convergence on x. So, not just we do not only show that f converges to f star exponentially fast, we also show that x converges to x star exponentially fast. But in order to show that we needed an additional assumption which is that function is also L smooth.

So, without L smoothness, it is difficult to claim that f converges, it is difficult to claim that x converges to x star exponentially fast. Would you mind summarizing the workflow one more time? Is the definition of exponential stability clear? So again to reiterate, so if you want to show that, so if you have v dot less than equal to negative alpha v, with this you cannot show that x basically is exponential, like x star is basically exponentially stable, right. So this only says that v converges exponentially fast. But if you want to show that x is exponentially stable equilibrium, then you have to show these set of inequalities. So, v has to be has to have an upper and lower bound in terms of x minus x star norm square and v dot also has to have some kind of upper bound alright.

So, in the previous lecture when we looked at f to be mu strongly convex, we could show this particular result. which says that v converges exponentially fast, but it does not tell us anything about convergence of x to x star exponentially fast right. So, in order to show that we probably need to assume more on the function and that is where we also require the function to have l smoothness. So, now our first thing is to make sure that these inequalities are satisfied for the Lyapunov function that we are going to be working with right. So, if f is mu strongly convex then you we have this particular definition of first order condition for mu strong convexity and that basically gives us this equation 1 right.

If f is l smooth then we get this particular inequality and this gives us equation 2. So, from 1 and 2 we can you can see that v is or like f of x minus f star it is sandwiched between these two terms. And that also tells us that you can possibly choose this as a Lyapunov function. Why? Because first of all, this v is greater than equal to 0. If there is a unique minimizer, it is exactly equal to 0 at x star.

Everywhere else, it is strictly greater than 0. So, and it is upper and lower bounded by norm x square or in this case, because x star is not necessarily 0. So, it is bounded by x minus x star norm square. So, this basically tells us or gives us the first part of this equation.

this inequality. Now we need to derive the second part of the inequality which is the condition on v dot in terms of norm of x square right. Now if I look at the derivation for v dot, so v dot essentially is this particular term which using PL inequality, so this is from PL inequality you can replace this norm f square with f of x minus f star right. So, this is just using Peer Inequality because every strongly convex function also satisfies Peer Inequality. So, that means what we get is v dot is less than equal to negative 2 mu times v and now we read now we basically need to write v in terms of x because we want to show this particular inequality. So, we want to be able to write v in terms of norm x square right.

And since this is already true that v is greater than equal to this, so that means minus of v is less than equal to minus of this particular term and from here we can actually replace v upper bound negative or upper bound v with this particular term right. So that means v dot is less than equal to negative mu square times x minus x star. So now we have all the ingredients to show exponential stability of the equilibrium x star. So with alpha 1, alpha 2 and alpha 3 given like this and that is what we show.

Is this clear now? Alright. Again like in order to show exponential convergence it usually requires you to assume both l smoothness and mu strong convexity. Without smoothness you do not get the upper bound, without strong convexity you do not get the lower bound. So, you kind of need both. Yeah. not a minimax problem, but let us say you can view this as a large control gain problem or essentially if I so like if so the question well I think we will possibly look at a similar result in the later lectures.

$$\alpha_1 = \frac{\mu}{2}; \quad \alpha_2 = \frac{L}{2}; \quad \alpha_3 = \mu^2$$

$$m \le \sqrt{\frac{\alpha_2}{\alpha_1}} = \sqrt{\frac{L}{\mu}} \qquad \alpha \ge \frac{\alpha_3}{2\alpha_2} = \frac{\mu^2}{L}$$

From (a) and (b): $$\boxed{\|x(t) - x^*\| \le \sqrt{\frac{L}{\mu}}\, e^{-\mu^2/L\, t}}$$

but if I work with a dynamical system of the form x dot is negative gradient of f right. Now, if I choose a large control gain let us say minus c times negative like minus c times gradient of f and if the norm like basically c kind of subsumes the norm on the disturbance then you can kind of then you can still guarantee convergence. So, it is not a minimax problem, but it is more of a I mean right now we are just viewing optimization algorithms as dynamical systems and we are just talking about the stability of the equilibrium. at least in this treatment we are not treating this as an optimization problem altogether even though there is one to one correspondence, but we do not treat this as an optimization problem.

Any other questions? Yeah. though in continuous time we are all I mean all the results that we are showing are in continuous time, but this also holds true for discrete time as well. You can show that for even for discretized setting where x k plus 1 is x k the suitable steps is step size you can converge exponentially fast. Thank you.