STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

Indian Institute of Science

Week 4

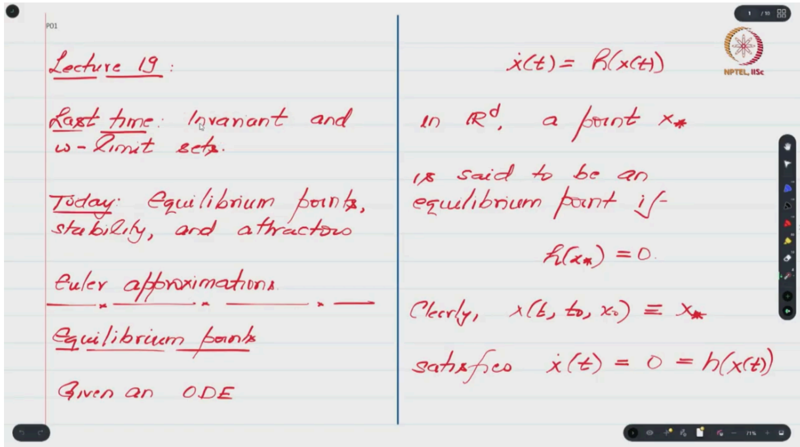
Lecture 19

Stability of ODEs and Lyapunov Methods

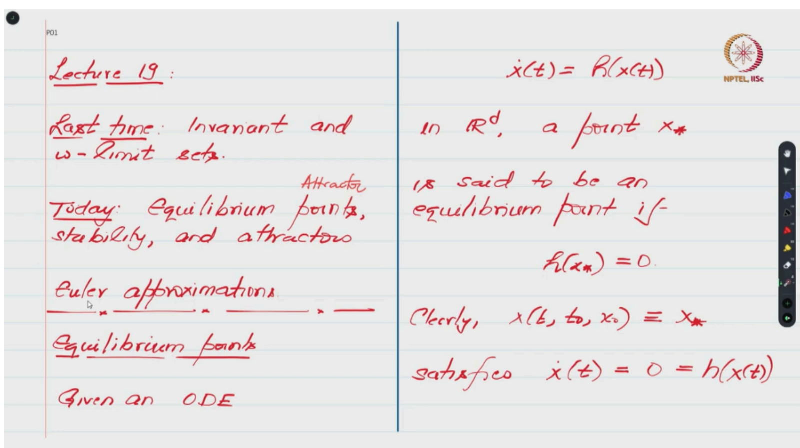
Hello and Namaste, everyone. Welcome to Lecture 19 of this NPTEL course on Stochastic Approximation. Let us do a quick recap of what we have been doing this week. We have been going over some basics of ordinary differential equations. We are doing a quick recap of some foundational concepts in ordinary differential equations.

In the previous class, in particular, we looked at some notions of positive invariance, negative invariance, invariant sets, omega limits, and alpha limits. In particular, one of the last things we showed in the previous class was that for any initial condition X0, its omega limit set is actually invariant. So, in today's class, we will introduce a few more definitions and understand a few more concepts. We will also try to understand an approximate way of looking at solution trajectories of an ODE, which we will refer to as the Euler scheme. These things will lay the foundation for how we are going to analyze stochastic approximation algorithms in the subsequent weeks. So, let us begin today's discussion.

So, if you recall, in the previous class, we looked at these concepts of invariant sets and omega limit sets. And in today's class, we will look at some definitions again. So, in particular, we will look at this notion of equilibrium points. We will also look at this concept called an attractor for an ODE, and we will discuss the stability of both equilibrium points and attractors. Finally, we will look at a concept called Euler approximation.



Which, at a very high level, provides us with a numerical approximation technique for finding solution trajectories of an ODE. So, let us begin by understanding the definition of an equilibrium point. So, given an ODE, X dot of T equals H of XT in RD. So, when I say an ODE is in RD, I mean that this function H is a map from RD to RD. So, with respect to such an ODE, a point X star is said to be an equilibrium point.



If H of X star is 0, so whatever vector X star is such that H of X star is 0, such a point we will refer to as an equilibrium point. So, why is this point special? This point is special because if you define a function of time T to be equal to X star for all T, then this constant trajectory will actually be a solution of this ODE. This is because If you take the derivative of this with respect to T, from the fact that this is a constant trajectory, you will get that this is 0.

On the other hand, since X star satisfies the property that H of X star is 0, we can see that H of X of T, which is basically a shorthand for this expression, which is equal to X star, then one can see that H of this expression is 0, and hence X dot of T equals H of XT for this constant trajectory.

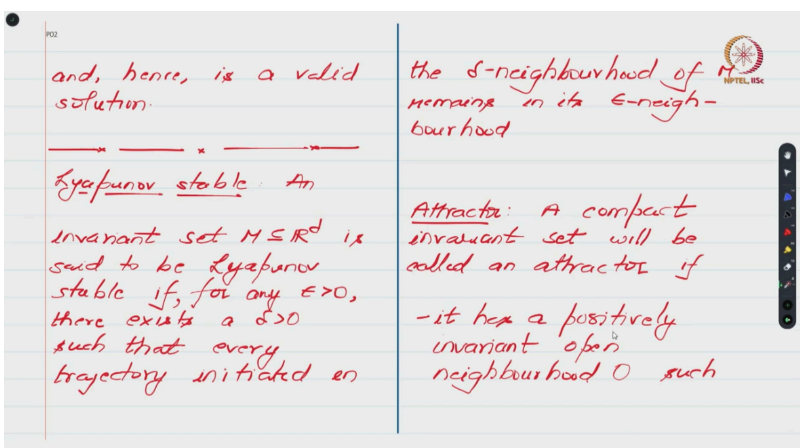
$$\dot{x}(t) = h(x(t))$$

$$h\left(x_*\right) = 0$$

$$x\left(t, t_0, x_0\right) = x_*$$

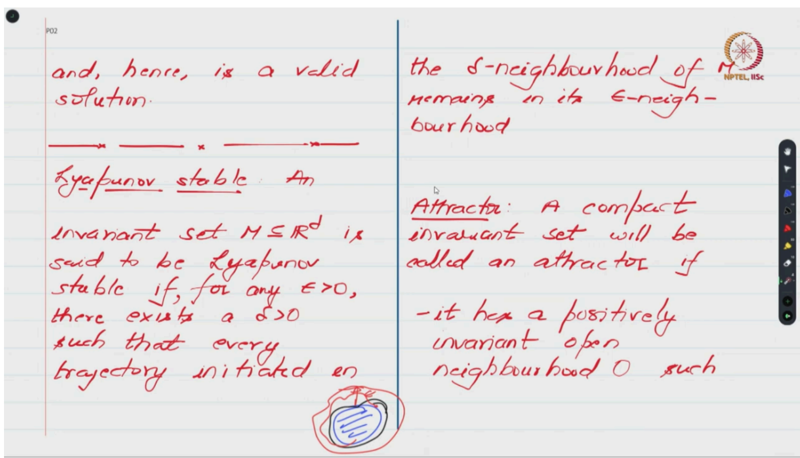
$$\dot{x}(t) = 0 = h(x(t))$$

So, in that sense, an equilibrium point is special. Okay, I mean, to put it in different words, if you start a solution trajectory at an equilibrium point, both in positive time and negative time, the solution trajectory will stay there. Okay. So, in that sense, equilibrium points are stable. So, now the question that we would like to ask is: if you had an equilibrium point and you started a solution trajectory in the neighborhood of an equilibrium point, would you go to that equilibrium point or not? So, towards answering such questions, we will first introduce the concept of Lyapunov stability.

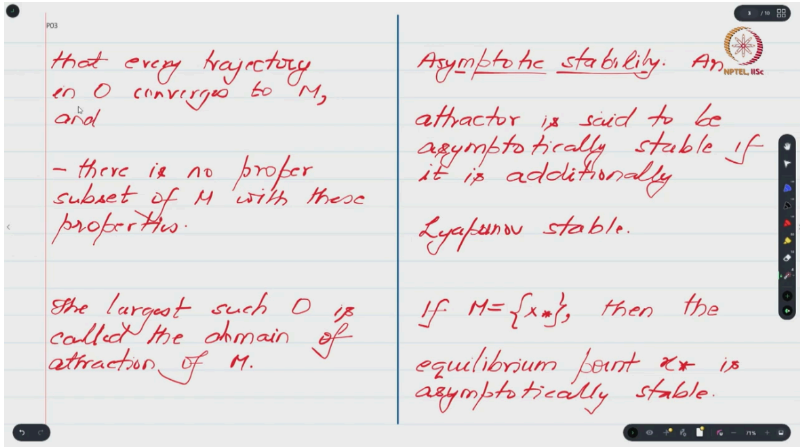


So, we will say an invariant set M is Lyapunov stable. If for any epsilon greater than 0, there exists a delta greater than 0 such that every trajectory initiated in the delta neighborhood of M remains in its epsilon neighborhood, right? So, what it means is, let us say we have a set M like this, okay? So, this M would be said to be Lyapunov stable if for every epsilon neighborhood.

So, for every epsilon neighborhood of this set M, there exists a delta neighborhood such that if you start the solution trajectory here, the solution trajectory remains within this epsilon neighborhood for all t greater than or equal to 0. So, if you start here, then from that point onwards, the solution trajectory should remain within the epsilon neighborhood, and this should hold true for any epsilon. So, for any epsilon, I should be able to find a delta that ensures this condition, and if we are able to do that, one can conclude that this invariant set is Lyapunov stable. So, this is one definition. The next definition that we need to understand is that of an attractor.



So, a compact set, a compact invariant set, will be called an attractor if it satisfies two conditions. First is that it has a positively invariant open neighborhood O. Positive invariance, recall, means that if you start a solution trajectory within O, the solution trajectory will remain in O forever—I mean, at least in positive time. So, we need a positively invariant open neighborhood O such that for every trajectory initiated in O, it eventually converges to M. Furthermore, M has no proper subset with the above property. Is this okay?

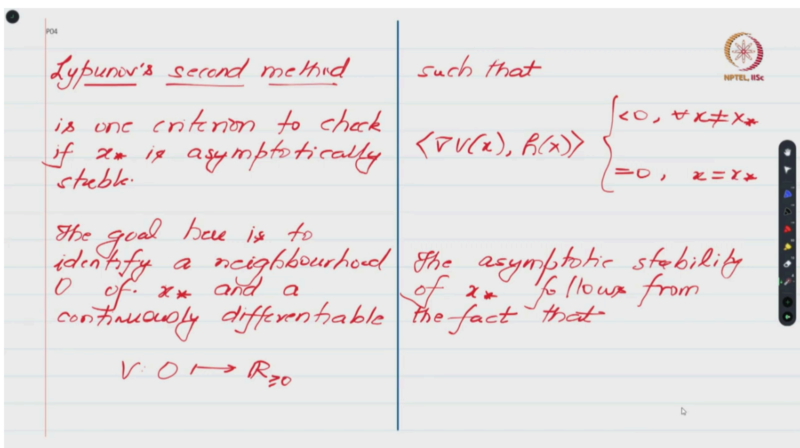


So, as long as M does not have a proper subset with the above property, we will call M an attractor. And the largest such open neighborhood for which this fact holds with regard to M will be called the domain of attraction. So, the domain of attraction makes sense here because since M is an attractor and O is the neighborhood under this definition, it implies that if you start a solution trajectory within O, then the solution trajectory will actually go to M. So that is why this is called the domain of attraction. Wherever you start within O, you will go to M. That is why this is referred to as the domain of attraction.

The next concept that we need to understand is that of asymptotic stability. In other words, when would we say an attractor is asymptotically stable? An attractor is said to be asymptotically stable if it is additionally Lyapunov stable. So, an attractor means that it has a positively invariant neighborhood from where if you start a solution trajectory, you go to this attractor M. We not only require that, but we also require this Lyapunov stability.

By Lyapunov stability, recall it means that for every epsilon, we need a delta neighborhood of M such that if you start within the delta neighborhood, the solution trajectory remains within this epsilon neighborhood. So, when you additionally ensure this Lyapunov stability, we will say that this attractor is actually asymptotically stable. And when this attractor M consists of only a singleton X star, then instead of saying that the set is asymptotically stable, we will simply say that the point X star itself is asymptotically stable. So, let us quickly summarize what we have done so far. We have understood the concept of asymptotic stability.

We have understood the concept of attractors, Lyapunov stability, and we have also understood the concept of equilibrium points. Now, one can ask—if you remember the question I had asked a few minutes back—if you start in the neighborhood of an equilibrium point (or now we can also talk about attractors), what is the guarantee that you will go towards that equilibrium point or this invariant set M? So, one way to check whether the trajectory will move towards either this equilibrium point X star or this invariant set M is via what is called the Lyapunov method. So, the idea in the Lyapunov method is to show that or is to first identify a neighborhood O of X star and a continuously differentiable function V, right, which is defined over O and maps elements of O (a subset of Rd) to non-negative real numbers.



$$V: 0 \mapsto R \geq 0$$

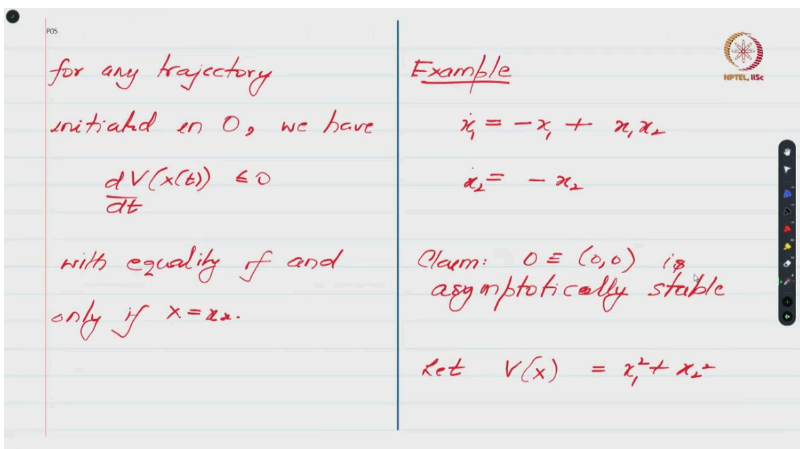
So, let us again understand what we are trying to do. We are right now focusing on showing whether X star is asymptotically stable or not, and towards showing that, we are now going to construct what is called a Lyapunov function. So, what is a Lyapunov function? For a Lyapunov function, first you need to identify a neighborhood O and a function V, which is defined on this neighborhood O and maps elements of O to the set of non-negative real numbers.

And keep in mind that O is a subset of Rd containing X star. And what we require—or when would we say such a V is a Lyapunov function—is if it satisfies a condition like this. That is, the inner product between grad of V of X and H of X. This is a vector in Rd; this is a vector in Rd. So, you can take their inner product. So, this inner product

Should be strictly less than 0 for all x not equal to x star and should equal 0 if and only if x equals x star, right?

$$< \nabla V(x), h(x) > \{< 0, \quad \forall\, x \neq x_* \quad = 0, \quad x = X_*$$

And one can ask, okay, if you have a Lyapunov function, so what? So, one can show that whenever you have a Lyapunov function, then your x star will actually be asymptotically stable, and it is asymptotically stable because one can then show that along any solution trajectory initiated in O, the value of V actually decreases. It decreases because the derivative of V with respect to time is less than or equal to 0, with equality if and only if X equals X star.

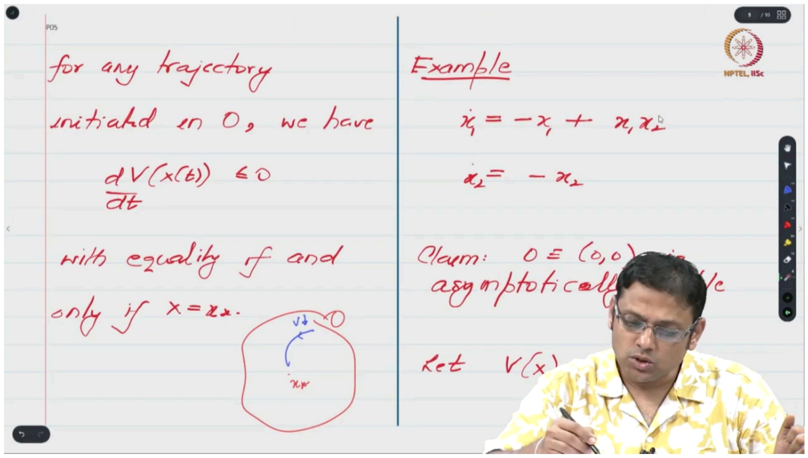


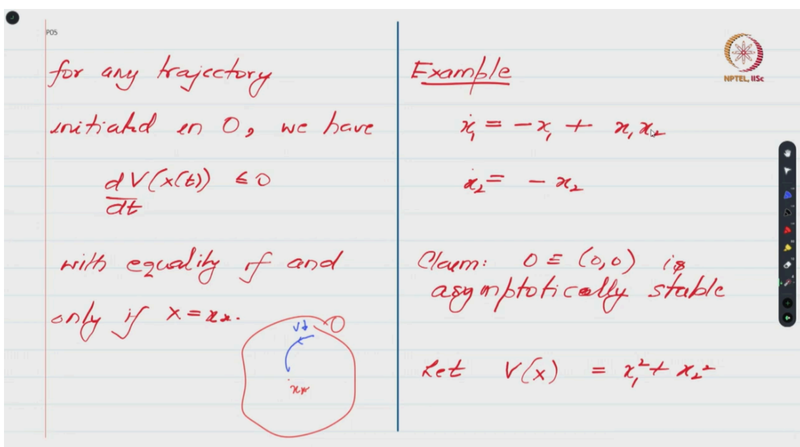
$$\frac{dV}{dt}(x(t)) \leq 0$$

So, what it means is that, let us say here you have your X star, Here, let us say, is your neighborhood O, and let us say you start over here. So, along this solution trajectory, the value of V keeps decreasing, and since it keeps decreasing, it has to go to a place where V takes, in some sense, the smallest value, and X star will be a point where V of X star takes the smallest value. Is this okay? And hence, it has to go over there because, for any other point other than X star, right, this strict negativity ensures that the value of V will continue to decrease.

Is this okay? So, let us look at an example of a Lyapunov function. So, towards that, let us define this two-dimensional ODE. In this case, your X1 dot is equal to minus X1 plus

X1 X2, and X2 dot equals minus X2. And we are now going to show that the origin is actually an asymptotically stable equilibrium point with regard to this, which basically means that if you start sufficiently close to the origin, then the solution trajectory of this ODE will actually go towards the origin.





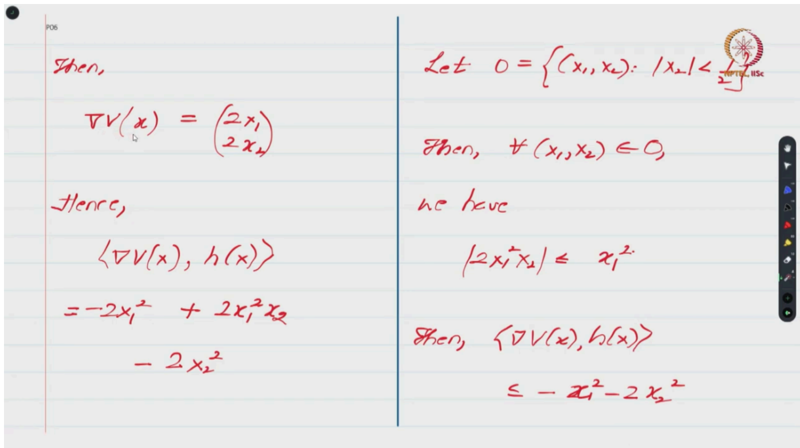
So, in order to show that, we are going to construct a Lyapunov function, and you know, in this case, we will look at the Lyapunov function which is given by x1 square plus x2 square. So, now let us see why this satisfies a property like this. So, why does this satisfy this? So, let us take the derivative of V with respect to x1 and x2.

$$\dot{x}_1 = -x_1 + x_1 x_2$$

$$\dot{x}_2 = -x_2$$

$$V(x) = x_1^2 + x_2^2$$

By the way, this x over here is basically a shorthand for this vector x1, x2.



So, the gradient of V with respect to x consists of 2x1 and 2x2. So, now, if you take the inner product of grad V and H, then one can trivially see that it would result in an expression involving minus 2x1 square minus 2x2 square along with this middle term, which is 2x1 square x2. Is this okay? So, this is very easy to check.

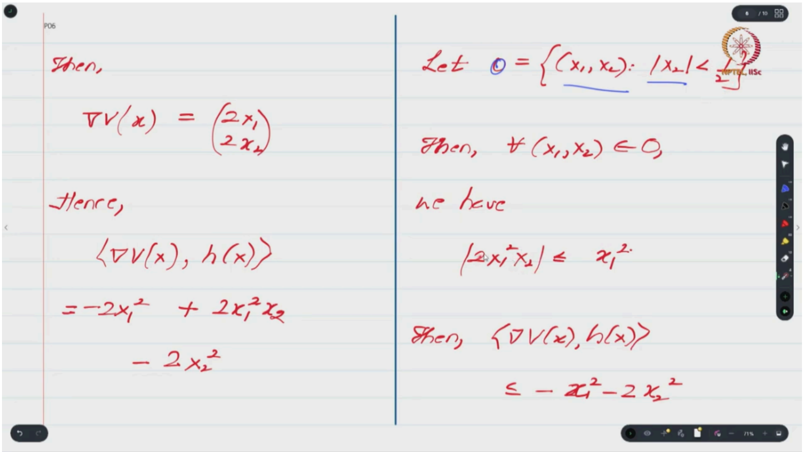
You can do it on your own.

$$\nabla V(x) = \left(2x_1 \; 2x_2\right)$$

$$< \nabla V(x), h(x) \geq -2x_1^2 + 2x_1^2 x_2 - 2x_2^2$$

Now, if this term was not there, then this whole term would have been negative. However, since this term is there, sometimes this term can become positive, which will violate the condition that is required under the definition of a Lyapunov function. However, if you define, you know, this neighborhood O—so let me put a different symbol over here—the neighborhood O to consist of all those x1, x2 where the absolute value of x2 is strictly less than half. So, if you focus the definition of V in this neighborhood, then one can see that for all x1, x2 within this neighborhood, the value of 2x1 square x2, in particular the absolute value, is actually less than x1 square, right?

So, your x2 is less than half. So, 2 times the absolute value of x2 will be strictly less than 1. Hence, this expression will actually be strictly less than x1 squared, right? And from this, one can conclude that for any x which lies in this neighborhood O, this inner product is actually minus x1 squared minus 2 x1 squared. So, recall that you had a minus 2 over here, and since this expression is strictly less than x1 squared, their sum will result in a minus x1 squared, and this minus 2 x2 squared is what you had over here.



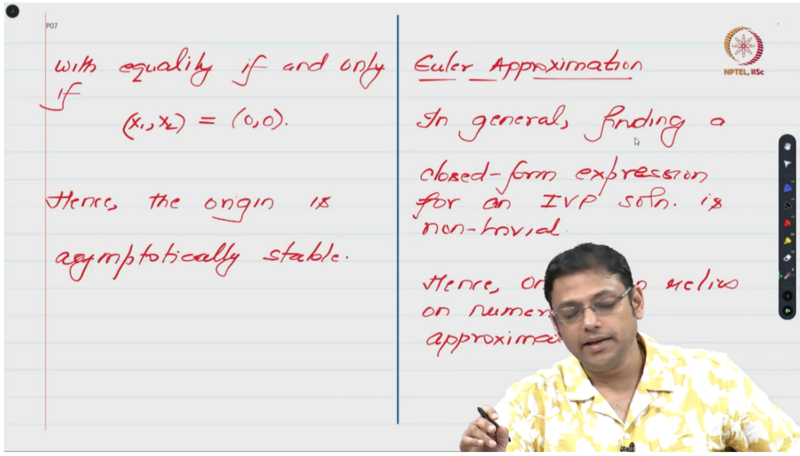
$$0 = \{(x_1, x_2): |x_2| < 1\}$$

$$\forall (x_1, x_2) \in 0$$

$$|2x_1^2 x_2| \leq x_1^2$$

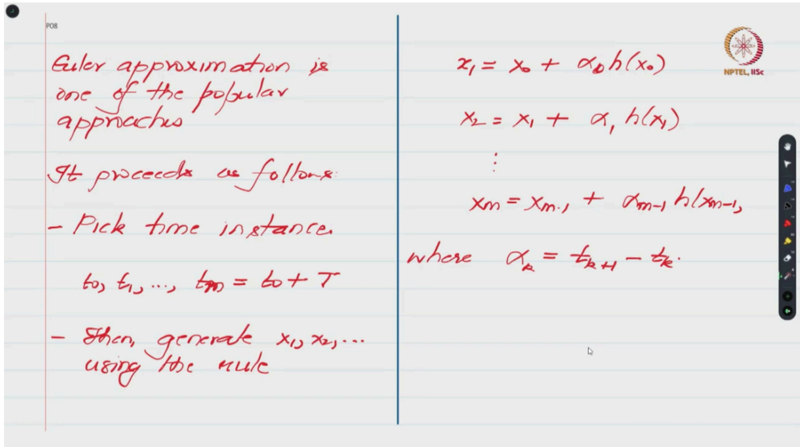
$$< \nabla V(x), h(x) > \leq -x_1^2 - 2x_2^2$$

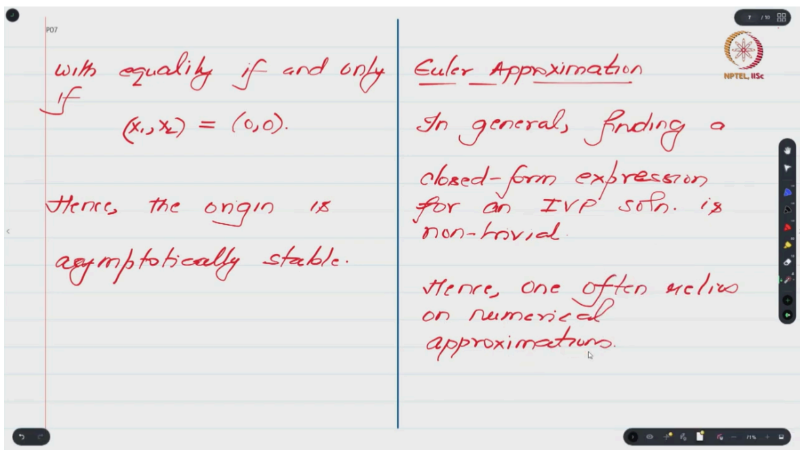
So, what we can now conclude is that the inner product between grad V and H of X is actually upper bounded by minus X1 squared and minus 2 X2 squared sum. And because X1 squared and X2 squared are positive, the negative of this sum will indeed be negative. And the equality over here will hold if and only if x1 and x2 are both 0. And because of that reason, one can conclude that this V actually is a Lyapunov function defined on this neighborhood O. And because we can construct a Lyapunov function, one can conclude that the origin is actually asymptotically stable for this ODE. So, now we will talk about this Euler approximation to approximate solutions of an ODE.



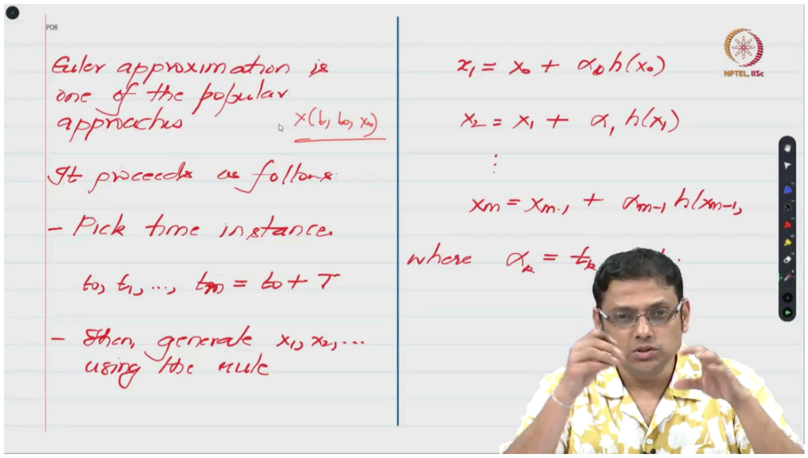
So, we will discuss this carefully, and I request all the readers to keep this Euler approximation in mind because when we start the analysis of stochastic approximation algorithms, you will see a very close connection to this Euler approximation, and then you will be able to correlate. Why we are using ordinary differential equations to analyze stochastic approximation algorithms, okay? So, we will discuss all this in subsequent classes. For today, let us look at this scheme called Euler approximation, where the goal is to find approximate solutions of ODEs. First, let us ask why do we need to find approximate solutions of ODEs? Well, the answer is that in general, if you are given an initial value problem, then finding a closed-form expression for the solution to that IVP is often non-trivial.

So, it is very hard to find, in general, the expressions for the solution trajectory. Hence, what one often does is rely on numerical approximations. So, that is what one does. And one of the popular approaches for numerical approximation is what is called the Euler approximation. And the idea in Euler approximation is as follows, right?





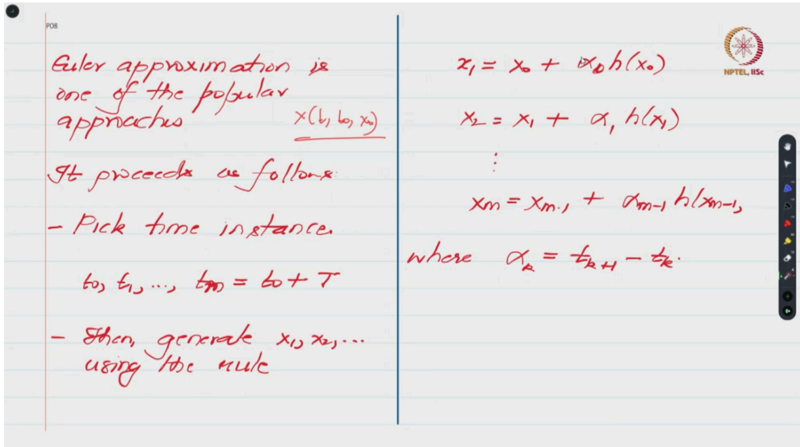
So, your goal, recall, is to approximate a solution trajectory. Let us say the solution trajectory is X, T, So, this is your solution trajectory, and the goal is to approximate this—that is, you want to find some loose approximation to the behavior of this trajectory over here. Recall that we do not have an explicit expression for this; that is the situation we are in. So, how does the Euler approximation work?



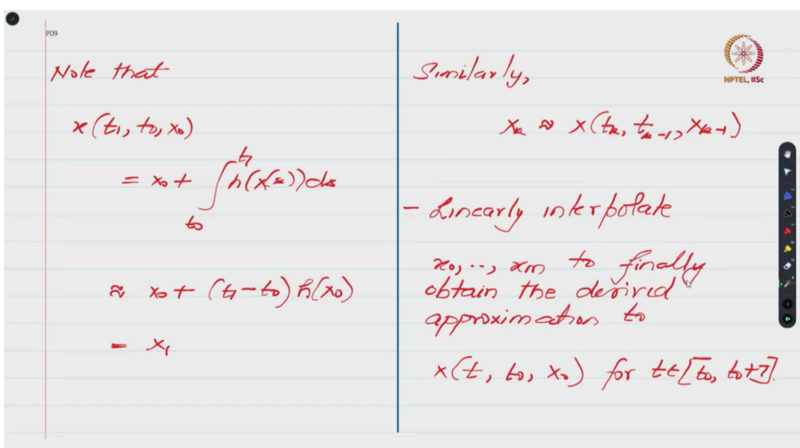
Well, the first thing it does is pick some time instances, starting from t0 all the way up to tm. So, it picks some t0, t1, t2, all the way up to tm, where tm is actually capital T time instance away from t0, right? And this capital T, you will soon see, is the window or the length of the window in which we try to approximate this solution trajectory. Then, the next step in this Euler approximation idea is to generate x1, x2, x3, and so on, using the following formula. So, what is this formula?

Well, you know X0. So, what you do is you take X0 to it, add alpha 0 times H of X0. Now, H is something we will presume is known to us. So, if you know X0, you can compute H of X0, which will be a vector. You take that vector, multiply it with alpha 0.

Where alpha 0 is basically t1 minus t0, right? So recall that you have t0 here, t1 over here. So you take the difference between them and call it alpha 0, and then you look at this sum and whatever is the result, call it X1. Now, again, you repeat the same process: you start at X1, add alpha 1 times H of X1, and whatever is the resulting value, you call it X2. And you continue this way of defining X3, X4, X5, and so on until Xm, which is given by the expression Xm minus 1 plus alpha n minus 1 times H of Xn minus 1, right? Where this alpha k is defined to be tk plus 1 minus tk.

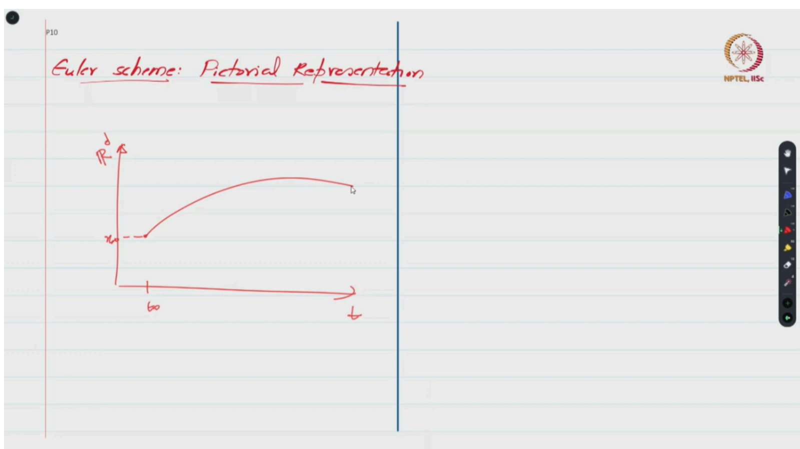


Is this okay? So, now one can ask why we are saying that these X1, X2, X3 somehow are approximations to your Euler solution. So, towards that, see that your value of x of t1, t0, X0 from the fact that this is a solution to the initial value problem will equal X0 plus this integral t0 to t1 h of x of s ds. So, this just follows from the fact that this value corresponds to a solution trajectory of this IVP. And then, if you recall your Riemann approximation idea and so on, one can see that this integral is actually approximately equal to X0 plus T1 minus T0 times H of X0.

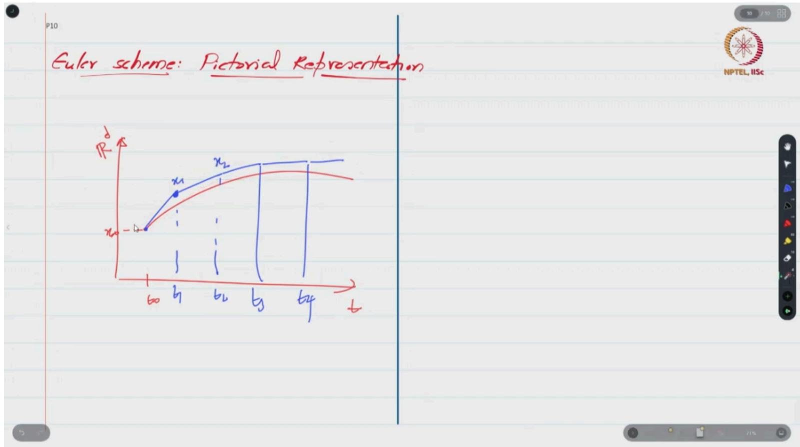


And this expression over here is actually what we have defined to be X1. So, in this sense, you can see that the value of X1 approximates the value of the solution trajectory at time instance T1. And in a similar sense, one can see that Xk approximates the value of the solution trajectory. which at time tk minus 1 is started at xk minus 1, and whatever the value of the solution trajectory at time tk is, that in some sense one can view as being approximated as xk, right? And this last step in the Euler approximation is to linearly interpolate this X0 to Xm, and whatever is the resultant trajectory that we will perceive as an approximation to the true solution trajectory for any T between T0 and T0 plus T. So, now let us look at a pictorial representation.

So, what we will do is we will have time on the x-axis, and, for simplicity, let us presume that your Rd is over here, and let us say this is T0, and somewhere over here is your x0, right? And let us say the true solution trajectory is something like this. Right. And now, what the Euler approximation idea is, you start at x0, right, and in some sense compute something like this. So, this will give you your x1, and this is where your t1 would be, and similarly, compute x2 by looking at the value of h at x1 and computing some vector like this. So, this will be t2, and this will be x2, and so on and so forth, okay?



So, this will be t3, this will be t4, and so on. So, you can see that this linearly interpolated trajectory will become closer to this red trajectory as we increase the number of points between t0 and t0 plus t. So, this just follows from the Riemann approximation idea, and this blue trajectory over here is what we will refer to as an approximation to this red trajectory, which is the true solution. So, with this, let me stop. Please keep this Euler approximation idea clear. Known very well to you, right.



We will repeatedly appeal to this when we look at the analysis of stochastic approximation algorithms in the subsequent weeks. So, with that, let me say thank you. Hope to see you again. Bye.