STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

Indian Institute of Science

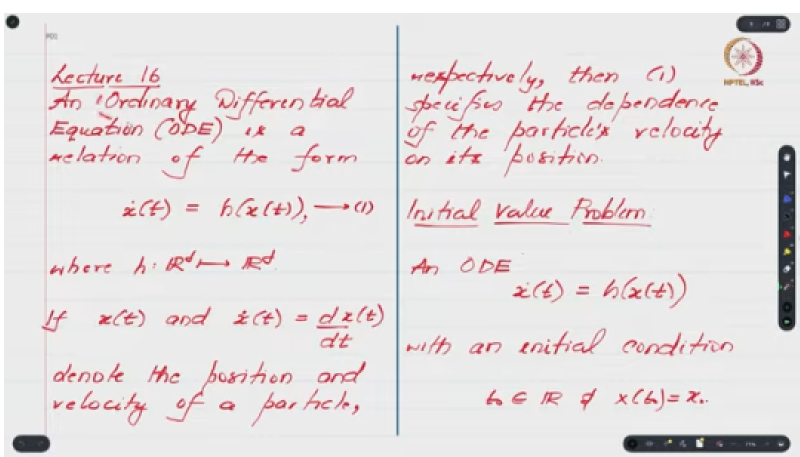
Week 4

Lecture 16

Existence and Uniqueness of Solutions to Ordinary Differential Equations

Hello and Namaste, everyone. Welcome to Week 4 of this NPTEL course on Stochastic Approximation. So, let us do a quick recap of what we have done over these three weeks. In the first week, we went over some basics of Stochastic Approximation. In the second week, we covered some fundamentals of probability theory.

In particular, we looked at the concept called conditional probability. And in the third week, we used these basic concepts to discuss martingales, martingale differences, and also talked about sufficient conditions under which martingale sequences converge. So, that was, in some sense, one half of the background we need to understand stochastic approximation. The other half we need is that of differential equations, particularly ordinary differential equations, whose basics we will cover during this fourth week. So, in Lecture 16, we will give an overview of ordinary differential equations, which we will abbreviate as ODE.



And, in subsequent weeks, we will see how the combination of ideas from ordinary differential equations and martingale differences can be combined to understand the behavior of stochastic approximation algorithms. So, what is an ODE? An ODE is an equation of this form. On the left-hand side, you have the derivative of a time-dependent function X of t, and on the right-hand side, you have H of X of t. Throughout this discussion, we will presume that H is a function which takes input from the Rd space and outputs something again in the Rd space.

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

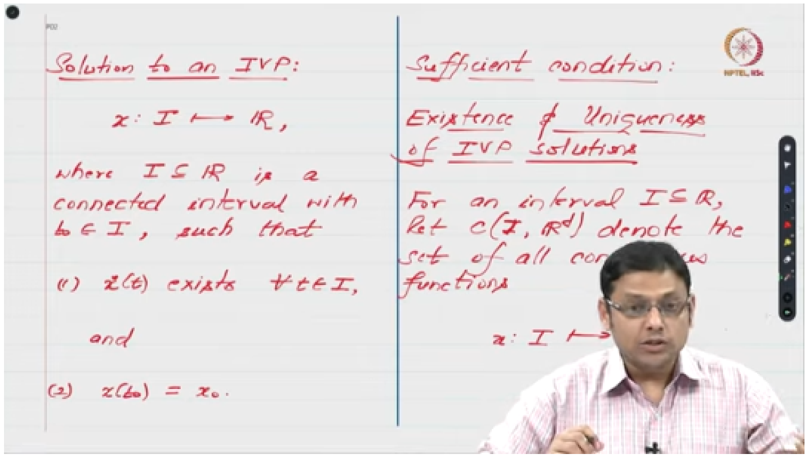
$$h: R^d \mapsto R^d$$

What is the intuitive interpretation of an equation like this?

Well, if you think of x of t as denoting the position of some particle and x dot of t as denoting the velocity of this particle, then this ordinary differential equation specifies how the velocity of this particle depends on its position. Is that okay? So, that is what an ODE is all about. And one can then ask, you know, if you have an equation like x squared equals 4, what is its solution? In the same way, for an equation of this form, right, we can ask, you know, what is the solution to this equation, right?

More specifically, we can define something called an initial value problem, right? In this initial value problem, we have this ODE x dot of t equals h of x of t, right? And in addition to that, we have an initial condition, right? such as the following form: we have some T0 in R, that is the real line, and we have been given this additional condition that X of T0 equals X0. So, the initial value problem is basically an ordinary differential equation along with an initial condition of the form which includes a time and the value that your solution trajectory should satisfy at that time instance.

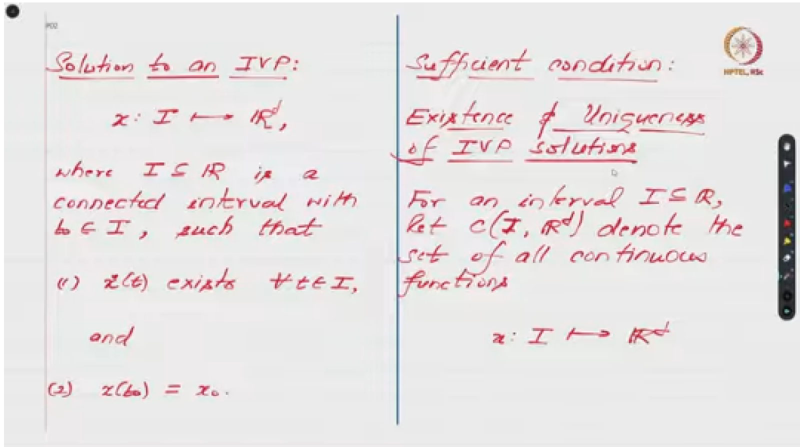
So, now one can ask, what is the solution to an initial value problem? Well, a solution to an initial value problem, as I said, is a trajectory, meaning it is a function whose domain is a subset of the real line, and in some sense, this domain represents the time instances—you can think of that—and the output of this trajectory is actually an element in Rd. Sorry, there is a typo here, so this should be Rd. So, what is a solution to an IVP or initial value problem? It's a trajectory which takes as input something in the real line and spits out something in Rd, right?



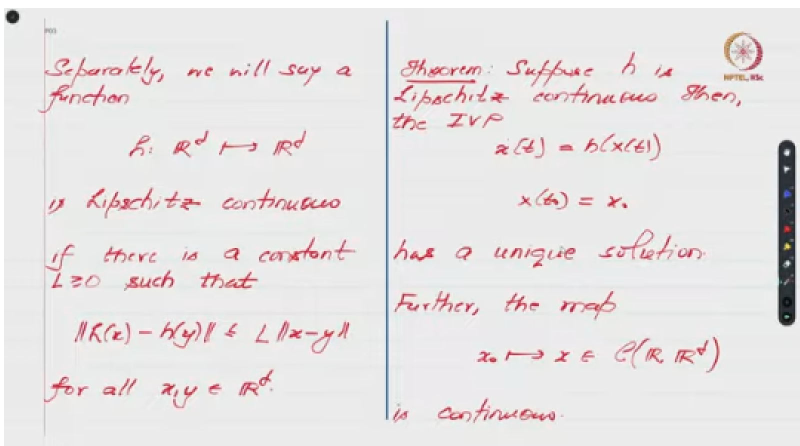
$$x: I \mapsto R^d$$

And typically, we will assume that this I is a connected interval. You know, which is a subset of R, and also it includes this, you know, time instance t0 that has been given to us as part of this initialization condition that has been specified, right? So, in other words, we have to, you know, if you go back to that position-velocity relation, The solution to an IVP is basically a trajectory followed by a particle which satisfies that relation for every T in I. And in addition, it satisfies this initial condition that, at the specific time instance T0, the solution trajectory should indeed pass through X0. So, you know, whenever you have seen an equation of the form x squared equals 4 or something like that, one can ask when such an equation would have a solution, and furthermore, we often ask, you know, does this equation also have a unique solution whenever the solution exists?

So, in the same way, one can ask what conditions guarantee the existence and uniqueness of solutions to an IVP. So, we will now discuss one sufficient condition for the existence and uniqueness of solutions to this initial value problem. And for that, we need to introduce a few notations. The first among these is this set. So, for any interval I which is a subset of R, we will use this notation, which is CI, RD, to denote the set of all continuous functions.



which go from I to RD. So, notice that this I is from here, and this RD is over here. So, basically, this set consists of all continuous functions whose domain is I and whose range is RD. So, this is what this notation means. Separately, we will say a function H which goes from RD to RD.



is Lipschitz continuous. So, we will say such a function is Lipschitz continuous if there exists a non-negative constant L such that the following condition holds. So, what does this condition say? It says that the distance between the output at x and y under this function h—so h is your given function—you look at h of x and you look at h of y, right? And you look at the distance between them. At this point, I have not specified what metric is, but you can choose whatever metric you like, for example, the Euclidean metric, right? So, you look at the distance between them.

And we will say this function H is Lipschitz continuous if this distance is upper bounded by L times the distance between X and Y, right? And this relationship should be true for all XY in RD.

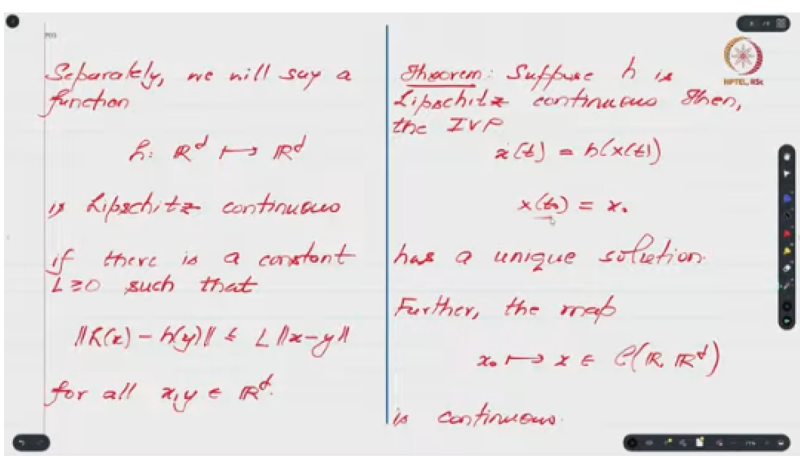
$$h: R^d \mapsto R^d$$

$$||h(x) - h(y)|| \leq L||x - y||$$

$$forall\ x, y \in R^d$$

So, let me again summarize: a function H is Lipschitz continuous if there is a constant L greater than or equal to 0 such that the distance between the outputs is upper bounded by L times the distance between the inputs, and this should hold for all x, y in RD. This L should be independent of the choice of x and y; in that sense, it is a universal constant. So, with these notations in place, let us now discuss a sufficient condition for existence and uniqueness of solutions to an IVP, and that is stated as part of this theorem. So, the theorem reads as follows.

Suppose the function H, which is the driving function in your IVP, is Lipschitz continuous. Then the claim is that the initial value problem, which is given over here, So, recall that the initial value problem has two parts. It first has the ODE relation, and the second is that it has some initialization condition. It requires that the solution trajectory that you identify at time t0 should pass through x0.



So, this theorem says that if H is Lipschitz continuous, then the initial value problem indeed has a unique solution. So, notice what this theorem is saying. On the one hand, it says that there is a solution, and on the other hand, it is insisting that this solution is unique. So, as I told you, recall what a solution means. A solution means a function which specifies, at a given time instance, where this particle is supposed to be located.

So, this is the mapping from time to a value in Rd, and this theorem is insisting that if your H is Lipschitz continuous, you will be able to find a unique map which satisfies these two conditions simultaneously. The second half of this theorem says that the map between X0 to X is continuous. So, let us try to understand what this result is trying to say. So, X0 over here is the initial condition. So, for every initial condition,

the first half of the result says that there exists a unique solution. So, you take that solution, which satisfies the initial condition X0, and come up with this mapping. So, every initial condition will be mapped to some trajectory, right? And, you know, this first half of the theorem says that this trajectory actually is an element of CR, RD, right?

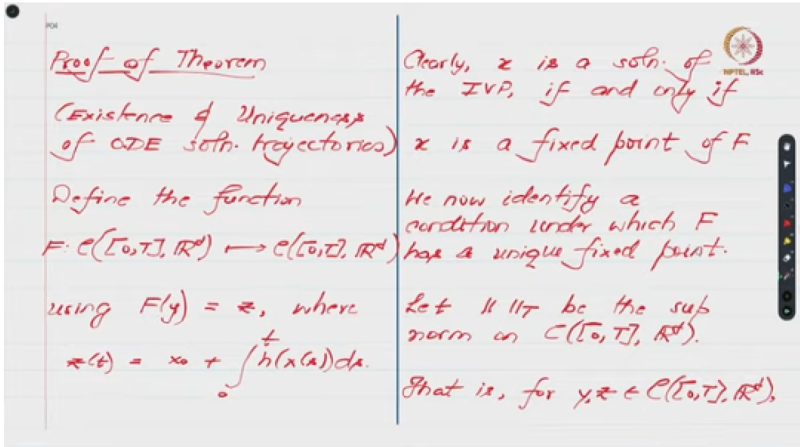
$$\dot{x}(t) = h(x(t))$$

$$X\left(t_0\right) = x_0$$

$$x_0 \mapsto x \in C\left(R, R^d\right)$$

So, if you recall the notation from the previous slide,

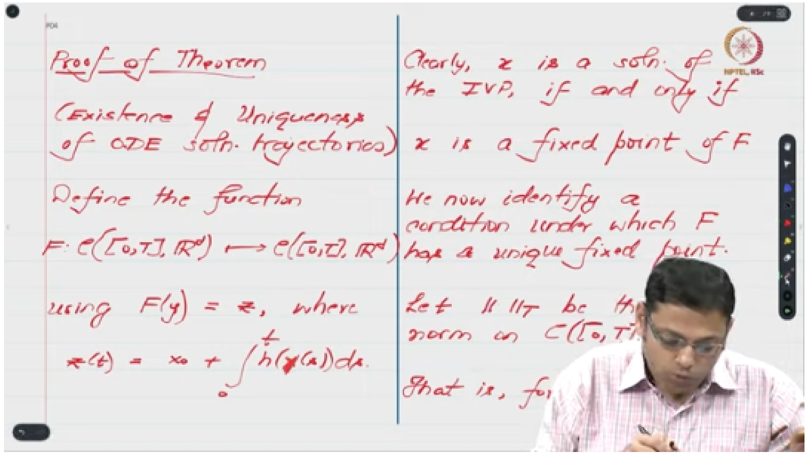
page or previous slide, here we had I and the theorem basically says that you have a solution trajectory which exists for every time in the real line. So, the second half of the result says that the mapping between X0 and this solution trajectory is continuous. So, we will keep what continuous means and so on and so forth for discussion in the next class. In today's class, we will focus on proving the first half of this theorem. So, let us see how does this theorem go about.

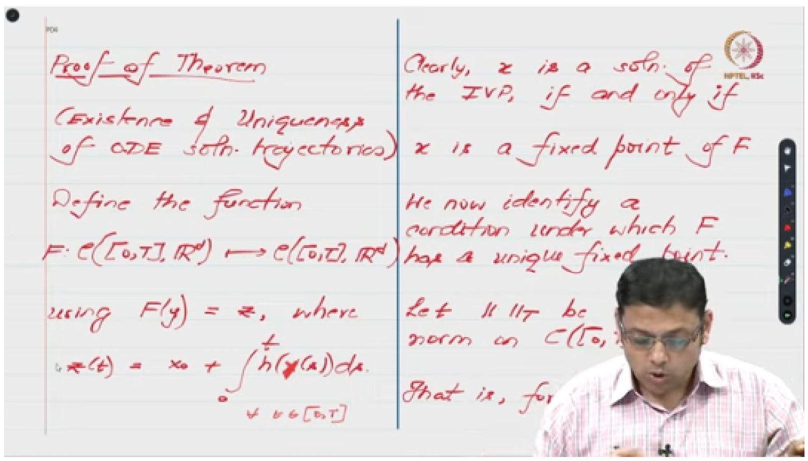


So, to begin with we are going to define a function f and we are going to first prefix or fix a value of t and let us focus on this set. So, this set is the set of continuous functions which maps time instance which lies between 0 and capital T to a value in Rd. So, this is the collection of all such functions. This is similarly the same collection of functions which map a time instance in 0 to t to rd. And now we are going to define a function which takes a trajectory here to a trajectory here.

So this is this function and the way this function is defined is given over here. So you give me some y from this space, f of y will give you another trajectory which lies over here. where at the t-th instance, z of t satisfies this relation over here. Okay, so there is a typo. Let me correct this.

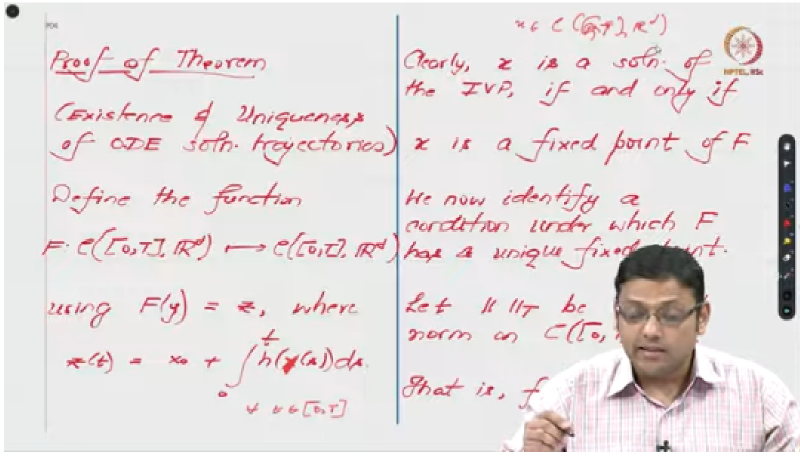
This actually should be y of s here, okay, not x of s. Okay, so whatever is the input trajectory, right? You put it over here, take H of that, integrate it from 0 to T. This T and this T are the same, okay? Added to X naught, which is the initialization condition that has been given to you, and whatever is the output, you assign the value Z of T, you know, to that. So, in this way, you define Z of T for all T and which is there between 0 and capital T, right? So, in this way, you have this Z of T trajectory defined. So, I hope you can see that if I give you any continuous trajectory that lies in this set, okay, you can use this F to construct an output trajectory which lies over here in the following sense, right?



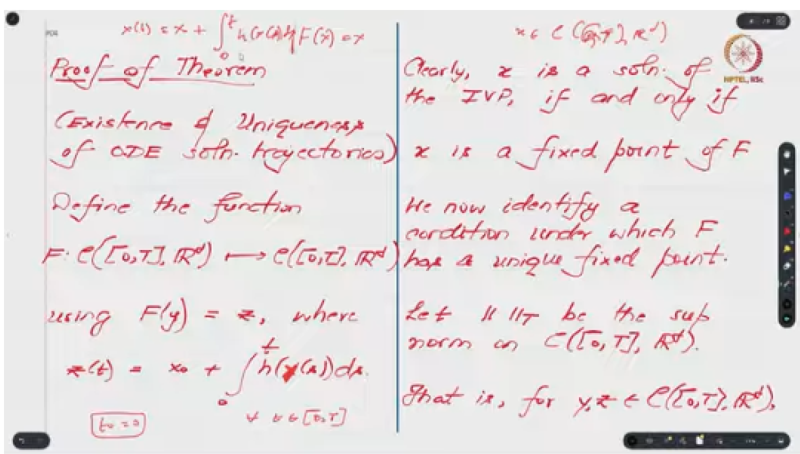


And also observe that you know, from the—because this is continuous, right?—from the fundamental theorem of calculus, one can immediately see that not only is Z of t a continuous function of time, it is also differentiable. So, now let us see how this relation or this definition of F is related to the question that we had, which is that of the solution to an IVP. So, one can see that a solution X or X, which lies in this map C0T, RD. So, this is a solution of the initial value problem in this interval.

So, this interval—let me write it cleanly—it is 0 to T. So, in this interval, your x is a solution trajectory if and only if x is a fixed point of F. So, what do I mean by that? You give x as input over here, and the output should also be x. So, a fixed point basically means that your F of x should be x. In other words, your x of t should equal x naught plus integral from t naught—sorry, integral to T, H of X of S, DS, okay? And I forgot to mention over here, right? For simplicity, we will consider your t0 to be 0.

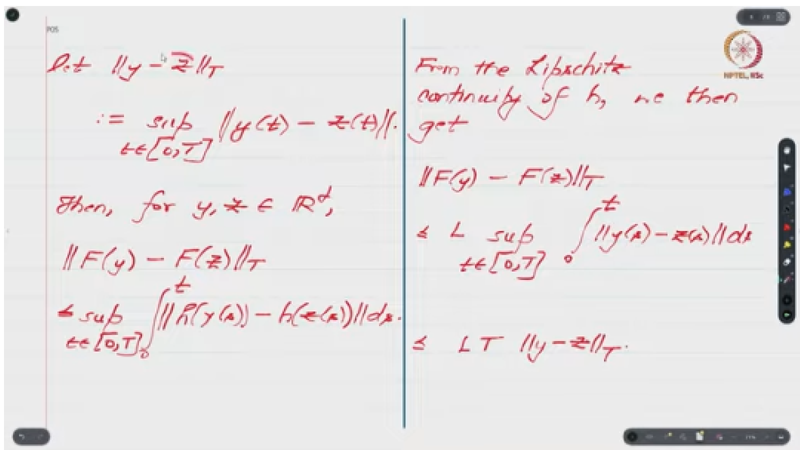


So, this is just for simplicity, but the same ideas will go through even if you had a general x0. So, in this sense, you can see that you have, you know, x on the left-hand side and x on the right-hand side. And from this perspective, you can see that your h actually satisfies this fixed-point relation. Hence, if you can somehow show that f has a fixed point, then one can conclude that indeed the initial value problem that we had, you know, that is given to us, has a solution on the interval 0 to t, right?



So if you recall in that problem statement, the conclusion was that we will have a solution trajectory over the entire real line, right. So, we will at this point focus only on having a solution between the interval 0 to t, and later on, we will see how to extend this solution trajectory beyond this interval 0 to t. Is this okay? So, what is the summary so far? We have defined this function f over here, and the conclusion is that the initial value problem has a solution if and only if f has a unique fixed point. Let me repeat: the initial value problem has a unique solution if and only if f has a unique fixed point.

So, now what we are going to do is we are going to identify a condition under which this function f has a unique fixed point. Towards that, what we will do is we will first define this norm over here. So, what is this norm? This norm takes two elements y and z in this set and says that the distance between these two trajectories is the supremum of the distance between y and z.



So recall that y and z take a time instance and spit out an element in Rd. So, y of t is an element in Rd, and z of t is an element in Rd. So, y of t minus z of t are two d-dimensional vectors. We compute their difference, and this norm is your norm of choice in the d-dimensional space.

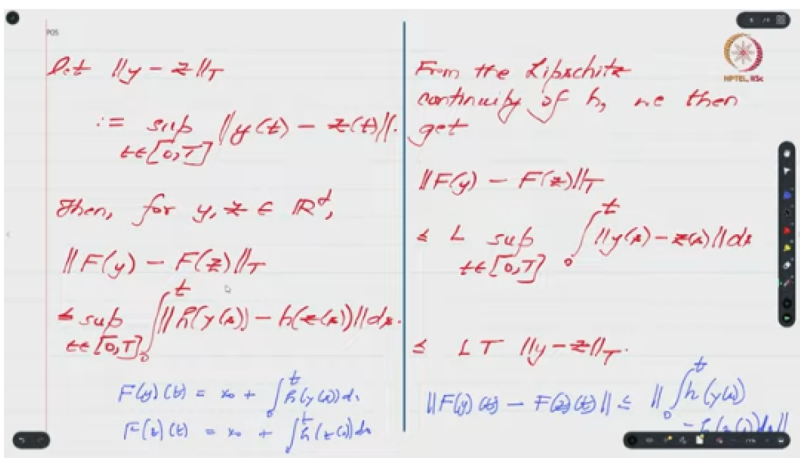
For example, this could be the Euclidean norm. So, you take this distance and take the supremum over little t in [0, capital T]. So, whatever the supremum is, this will be some real number, and that number you assign to the value of this metric. So, the idea is that you take Y and Z, which are two solutions—sorry, I shouldn't say solutions—two trajectories, right? And you look at the

difference between their values at time t and look at the norm of that vector, right? And you take the supremum of that norm; whatever the answer is, that is the definition of this norm, right? So, now let us use this definition to see how far f of (y) is from f of (z), right? So, f of (y) minus f of (z) is—you know, there is some x_0 plus something, and f of

(z) will also have some x0 plus something. Because x_0 is common to each other, they will cancel out, and we will be left with some expression like this, okay? So, let us just do it carefully.

So, what is your f of y? f of y is in particular... f of y at time instance t is basically x naught plus integral 0 to t h of y of s ds. Similarly, your f z of t is actually equal to x naught plus integral 0 to t h of z of s ds. So, if you take their difference, I hope you agree that you will end up with f of y of t minus f of z at t, right?

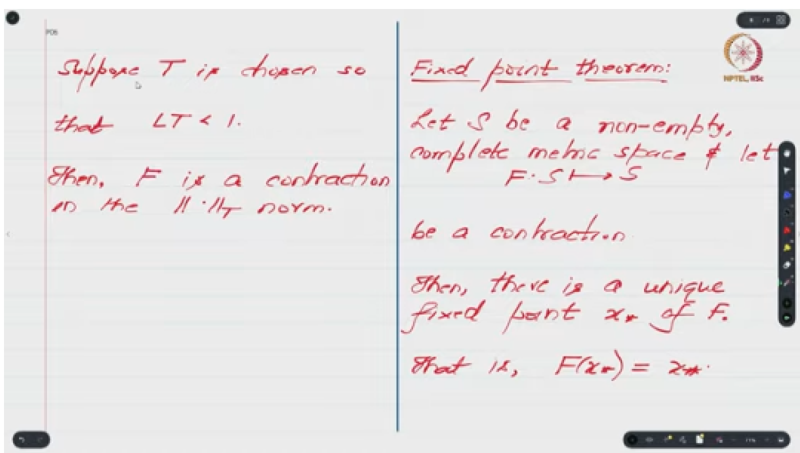
This norm is basically less than, you know, norm of integral 0 to t H of Y of S minus H of Z of S DS and then I would have this norm over here. Now, observe what I have done over here on the left hand side, we have this metric with regards to this time instance capital T which is defined in this sense. So, you will have a supremum taken over all these values and by properties of this metric and integrals one can take this inside via what is called as the triangle inequality.



So, one can take this inside and this is what you end up with and then because of this supremum you will end up with this supremum over here and this is how you get this relation. So, in other words, f of y minus f of z, these are two trajectories in C0TRD. So, the distance between them in this sense is upper bounded by this expression. Now, we know that this function h is actually Lipschitz continuous. Because it is Lipschitz continuous, we can pull out this L over here and we would be left with the distance between y of s and z of s, right?

Now, this expression is upper bounded by, you know, this expression is upper bounded by y minus z subscript t, right? Because this expression is the supremum. So, if you look at the distance at any s, this must be upper bounded by this. So, we can take this expression now this does not depend on S. So, we can pull it outside the integral and once you take the supremum you will end up with this capital T and this is the upper bound that you will end up with.

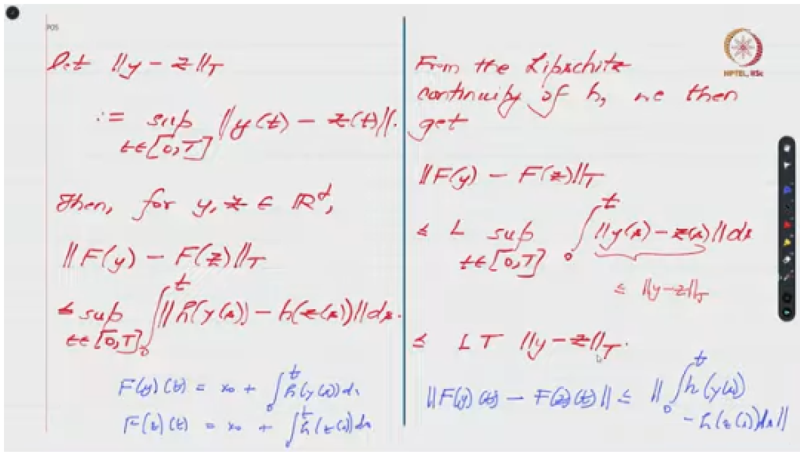
So, now what do we have? You give me two trajectories, then the distance between the outputs is upper bounded by LT times the distance between the inputs. That is the summary of this calculation. So, now what we will do is we will pick a value of t such that lt is strictly less than 1. So, because LT is strictly less than 1, one can conclude that F is a contraction.

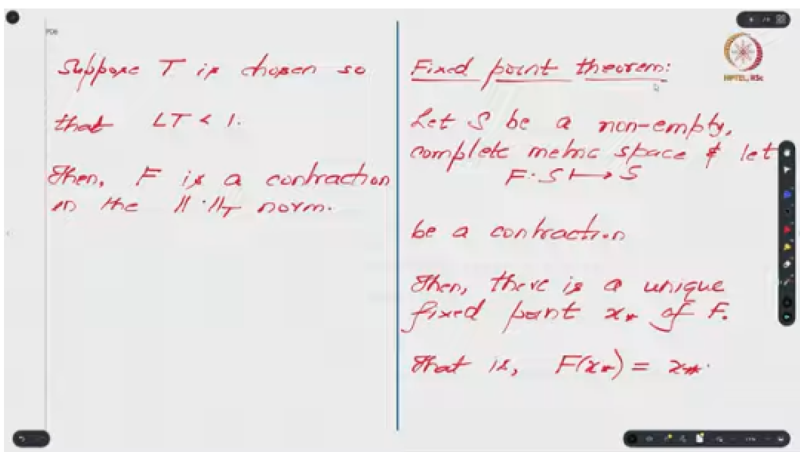


So, what does contraction mean? The contraction means that the distance between outputs is strictly less than the distance between the inputs. In particular, because this is less than 1, the distance is reduced by a factor which is strictly less than 1. So, this is what a contractive mapping means and this constant over here is again not dependent on the choice of y and z right. This is independent of the choice of y and z. So, what we are saying is that you know whatever input trajectories you give me y and z the distance between the outputs will be strictly less than.

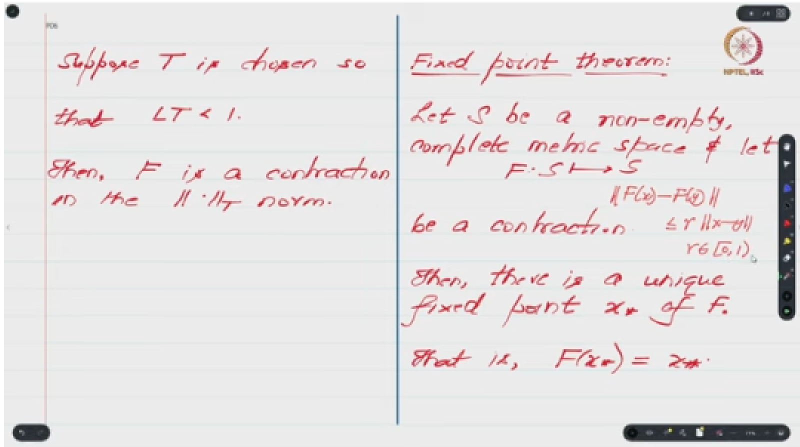
LT times the distance between Y and Z, that is, the distance between the outputs actually shrinks, you know, relative to the distance between the inputs. So, this is what we mean by F being a contraction. Now, there is a very interesting result about contractive maps.

In particular, we have something called a fixed point theorem, which goes as follows. Suppose S is a non-empty complete metric space, and F is some function which takes as input



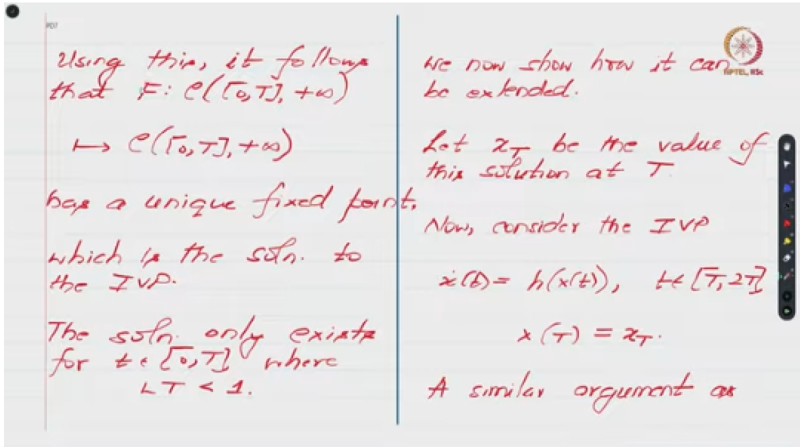


something from this metric space and spits out something in the metric space. And let us say F is a contraction, which means that F of X minus F of Y is less than gamma times X minus Y for some gamma which is strictly less than 1. Is this okay? So, we have this condition.



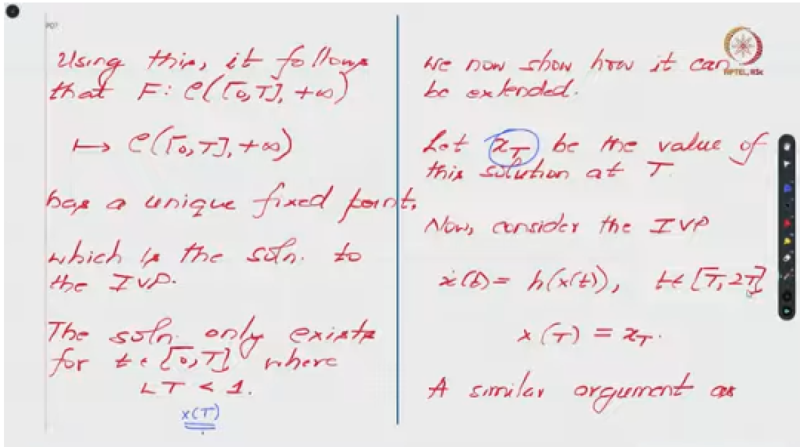
So, if you have this condition, then we will say that we have a contraction. Then the fixed point theorem, in particular, this is known as the Banach fixed point theorem. It says that whenever you have a contraction, there exists a unique fixed point X star which, you know, satisfies this condition. In other words, you have a fixed point of F, that is, F of X star equals X star. Fixed point means that whatever you give as input, that is what comes out as output.

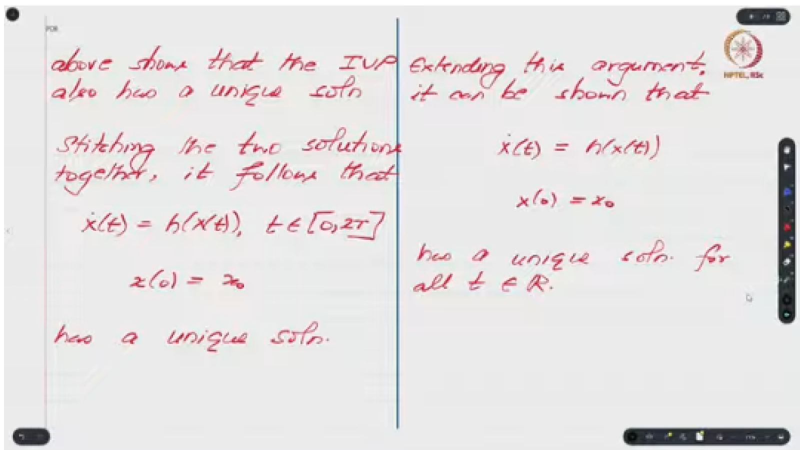
And the fixed-point theorem says that as long as you have a contraction, there is a unique fixed point. So, we can now use this property in the context of the F that we have. So, recall the F that we had defined, which goes from this set of continuous trajectories to this set of continuous trajectories, and we had shown that by the choice of capital T, this F is actually a contraction. Hence, from the Banach fixed-point theorem, one can conclude that this F has a unique fixed point. Now, because it is a unique fixed point, that automatically implies that we have a unique solution to the initial value problem, which restricts itself to the time interval 0 to T. So, let us go back to that initial value problem.



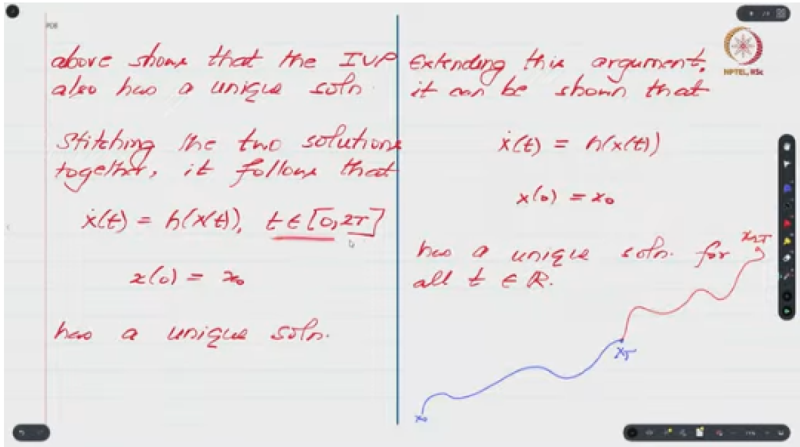
So, the initial value problem that we have been given is over here, and what we have managed to show is that if we restrict our time to lie between 0 to capital T, then this initial value problem has a unique solution, thanks to this fixed-point theorem. So, now the question is, can we extend this solution that we have between 0 to T to beyond the 0 to T interval? So, the idea is very simple. So, whatever is the value of the solution trajectory—the unique solution trajectory at time T—let us denote that as xT.

So, the initial value problem has this solution x, and whatever is the value at time T, you assign that or call that as x subscript T. And now you look at a different initial value problem, which has the same ODE, but now the initialization is different, and you look at a different time window. So, earlier you were looking between 0 to capital T; now you look at the time window T to 2T. Is that okay? And now one can see that a similar argument as before helps us conclude that this new IVP—this new initial value problem—also has a unique solution, right?





Now, we can stitch those two solutions. By 'stitch,' I mean that the initial solution, let us say, started at X0, okay, and at time t reached this Xt. Then, the next or subsequent solution started at Xt and, let us say, landed at X2t, right? You can stitch these two solutions together, and that would lead to a unique solution for this initial value problem. So, observe that in this case, the initial condition is x0, but now the time window is between 0 and 2t. So, the way we extended the solution trajectory from 0 to t to 0 to 2t was not special.



So, we can actually continue doing this from 0 to 2t, 0 to 3t, 0 to 4t, and in general, for 0 to infinity. And, you know, we can repeat this process also in the negative direction for negative values of t, and I encourage you to do that on your own. And by extending this idea both on the positive time axis and the negative time axis, one can conclude that there is a unique solution to this IVP for all t in R. And with that, we can conclude the proof of this theorem, which now guarantees that as long as your H function—which can be viewed as the driving function of your ODE—is Lipschitz continuous, your initial value problem will indeed have a unique solution. So, with this, let us end today's discussion.

Thank you.