STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Engineering

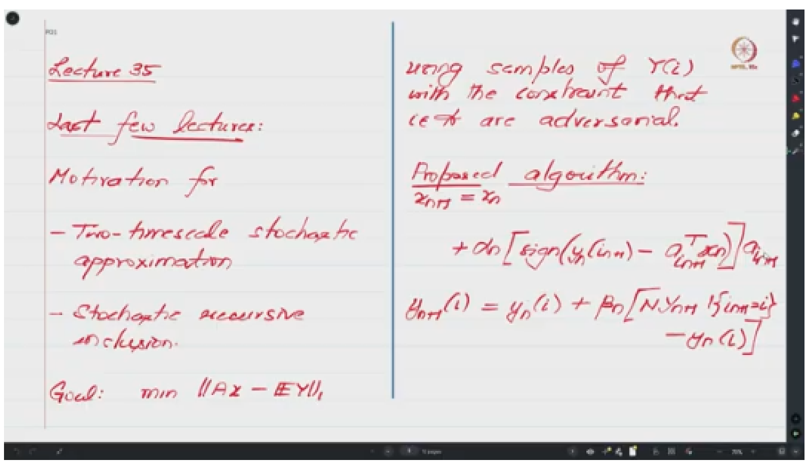
Indian Institute of Science, Bangalore

Week 9

Lecture 35

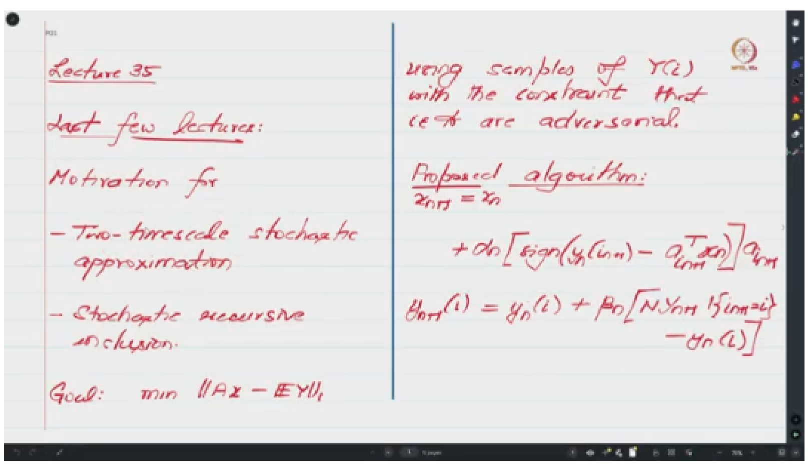
Well-Posedness of Differential Inclusions

Hello and Namaste everyone. Welcome to lecture 35 of this NPTEL course on Stochastic Approximation. If you recall, this week and the next, we are going to study some important variants of Stochastic Approximation algorithms. In particular, we are going to look at two-timescale Stochastic Approximation and Stochastic Recursive Inclusion. Over the past few lectures, I have used this example of a linear inversion problem.

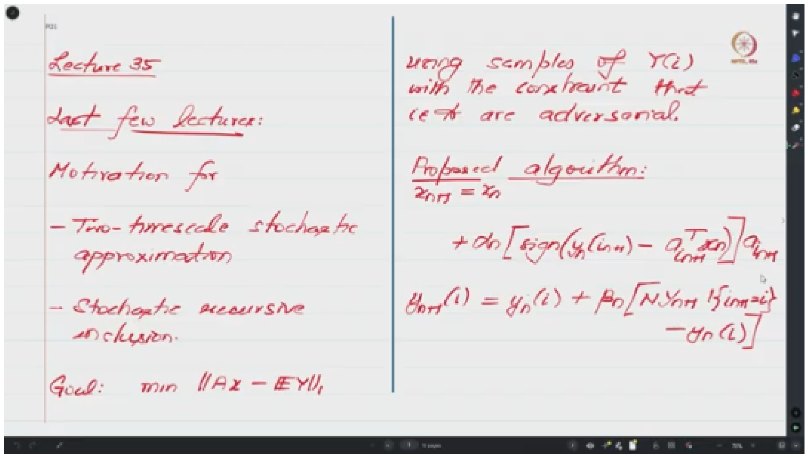


Right, and you know, looked at the case where some of the measurements come from adversaries and used that example, you know, to motivate solving this minimization of the L1 norm of AX minus the expected value of Y, right? And from there, coming up with the algorithm that necessitates the use of, on one hand, two-timescale stochastic approximation and, on the other hand, you know, taking a stochastic recursive inclusion view. Of the update rule, right? And that stochastic recursive inclusion perspective comes from the fact that, you know, each time you visit a particular point x, the adversaries could, you know, do whatever they like. So, in some sense, the adversaries, you know, play from a set, and that, in some sense, forces a set-valued nature to the dynamics. Right, and whenever you have a set-valued dynamics, the resulting algorithm is a stochastic recursive inclusion, right? So, you know, we then started asking questions about what the limiting dynamics of a stochastic recursive inclusion are. The limiting dynamics is,

what is known as a differential—the limiting dynamics is governed by what is known as a differential inclusion. The differential inclusion is a set-valued generalization of an ordinary differential equation, right? And in the last few minutes of the previous class, we sort of looked at conditions under which a differential inclusion guarantees the existence of a solution, and one of the sufficient conditions we stated was that of this set-valued map capital H being Marchaud. In today's example, we will check if, for this linear inversion algorithm that we have cooked up, the associated set-valued map H is indeed Marchaud or not. So, with that in place, let us begin our formal discussion. So, as I said, during this week, you know, we have been sort of looking at good motivation for studying two-timescale stochastic approximation and stochastic recursive inclusion, and you know, in some sense, the problem that we are interested in solving is motivated by, you know, the presence of adversaries.



We want to look at the minimization of Ax minus expected value of y right where you know instead of having access to expected value of y we have access to samples of y of i right and we have this constraint that an unknown subset of the coordinates are actually adversarial in nature which implies that you know they will provide us with values right and some of them could also be provided with the intention of derailing our algorithm right. So, towards that we proposed the following algorithm. So, Xn is updated in the following fashion and Yn is updated in this fashion right in particular you know even though we do not know which I's are adversarial and which are not.

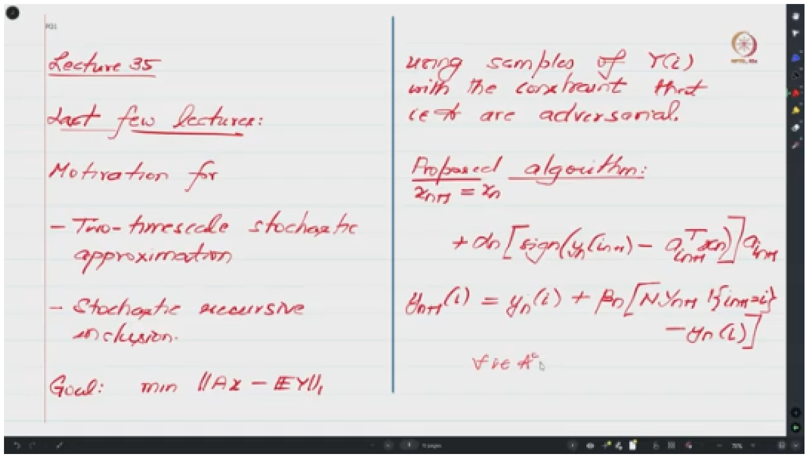


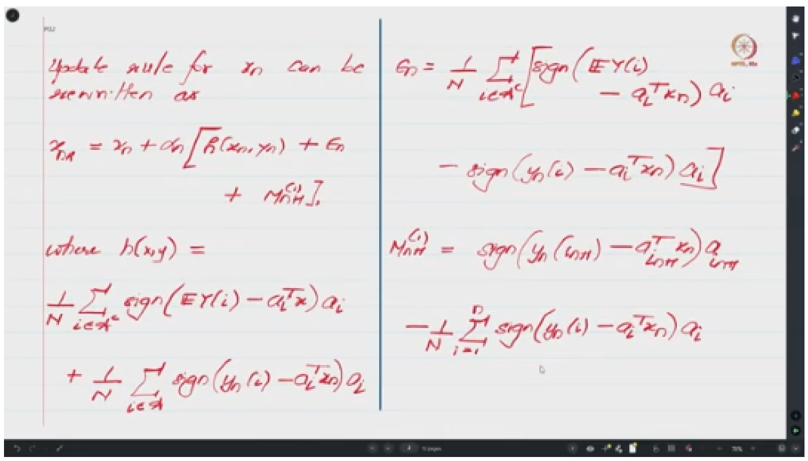
$$||Ax\ -\ Ey||_1$$

$$x_{n+1} = x_n + \alpha_n\left[sign\left(y_n\left(i_{n+1}\right) - a_{i_{n+1}}^T\ x_n\right)\right]a_{i_{n+1}}$$

$$y_{n+1}(i) = y_n(i) + \beta_n\left[Ny_{n+1}\ 1\left\{i_{n+1}\ =\ i\right\} - y_n(i)\right]$$

At least we know that for the unknown, you know, honest set of workers, this would be the update rule, right? And I should emphasize here that this is for all I in calligraphic A complement. And of course you know for I in A the update rule of Yn could have no meaning and Yn of I could be garbage values or I should not say garbage it could be adversarial values with you know designed in a way to derail our algorithm potentially. right and then we said that look the update rule for x in can be written in this form where h of x comma y can be broken down in this fashion right so notice that we break it down in a particular fashion to enable our analysis in particular this part of H of Xi is in some sense aligned with the ideal dynamics.

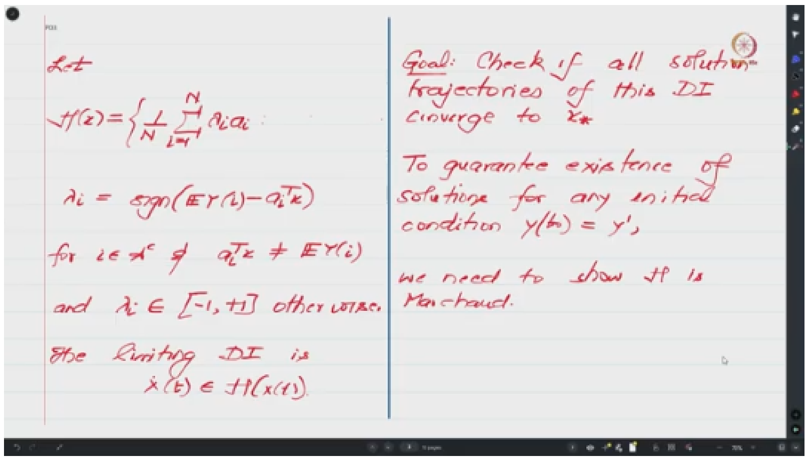




The ideal dynamics is that, you know, in some sense you should look at the gradient descent with respect to this objective function. So, this part is aligned with that. However, this part, you know, comes from adversaries. So, I think I need to be careful here. So, there is no N in here, right?

So, the inputs are X and Y. And whenever the inputs are X and Y, basically this will be there. And of course, when the input is Xn and Yn, wherever you see X and Y, you replace it with Xn and Yn, right? And this epsilon n over here is basically the difference between the, you know, idealized version. I think there is a mistake here.

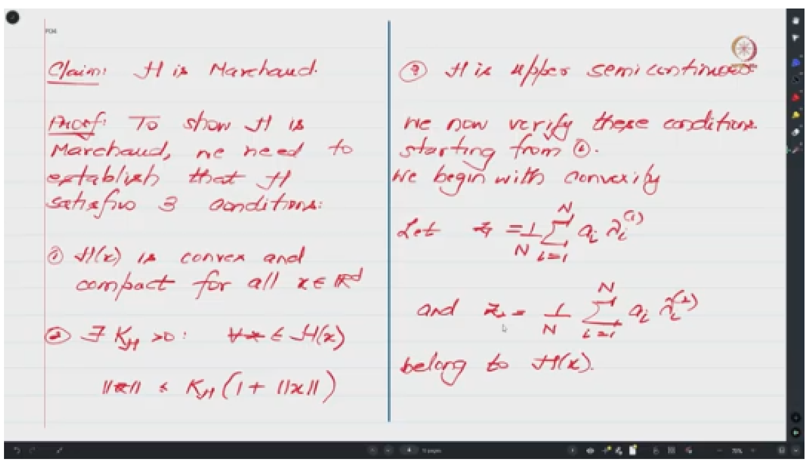
I think epsilon n should come with a negative sign so that this part has the positive sign and this part has the negative sign, so that this part cancels off with this. So, epsilon n, as I told you last time, is basically the perturbation bias. It basically compares what we would have ideally liked with what we would get using an estimate of the expected value of Y of I, and mn plus 1 is basically the martingale difference noise. Right, and in order to analyze this, you know, algorithm, we cooked up this set-valued function H and we defined it in the following way, right? So, this calligraphic



H of X, right, consists of all elements of this form where lambda I equals the sign of the expected value of Y of I minus A I transpose X, whenever I is non-adversarial and A I transpose X does not equal the expected value of Y. And for every other I, lambda I is an arbitrary number between minus 1 and plus 1. So, this is the way that we had, you know, defined this set H of X, right? And we said that the limiting differential inclusion that governs the stochastic recursive inclusion that we had over here, which is the one that is specified here, right? This is of this form, right? X dot of T belongs to capital H of X of T.

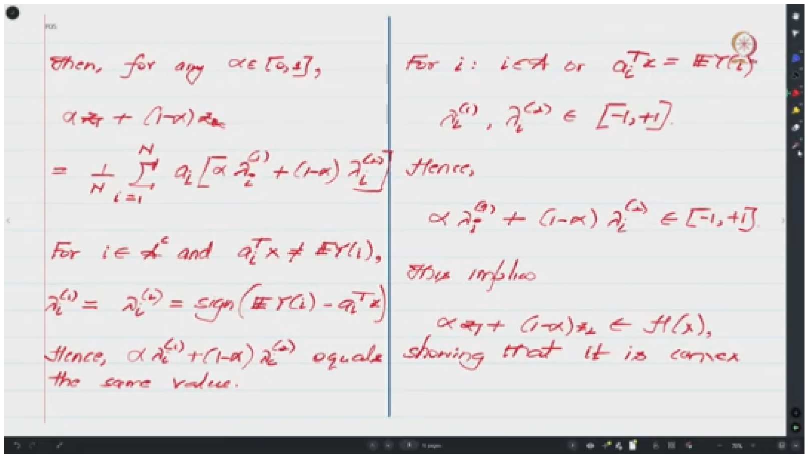
So, now before we start discussing the limiting behavior of the stochastic recursive inclusion from the previous slide, we would first like to check this idealized differential inclusion. So, notice that in this idealized thing. There is no notion of noise or there is no notion of adversaries, right? So, I mean, H of X is defined in this fashion, and we would like to ask whether the solution trajectories of this DI indeed converge to X star or not, right? Where recall that, you know, X—so I should be careful here—this should indeed converge to the expected value of X or not, which is the true expectation of the random variable X, right? Now, before we verify, you know, this fact, what we will do today is verify whether the limiting differential inclusion is well-posed or not, right? And in the next class, we will, you know, formally check whether the solution trajectories of the DI converge to the expected value of X or not, and subsequently, we will discuss the convergence of the stochastic recursive inclusion algorithm itself.

So, towards guaranteeing that this differential inclusion has solutions for any initial condition like this, we need to show that H is Marchaud. And if you recall, to show H is Marchaud, we need to verify three conditions. The first is that your H of X, for any X, is convex and compact. Furthermore, for any x—right, there, I mean, I should say not for any x—there exists some constant k h such that for any x and any z in H of x, right, the norm of z is upper-bounded by some constant times 1 plus the norm of x. So, this constant here should be independent of x. So, this, in some sense, ensures a linear growth condition on the size or norm of your vectors that are present in H of X. So, linear growth condition means something like this, and on top of that, we require that this function H be upper semi-continuous, right?



So, I will sort of elaborate what upper semi-continuous means when we have to verify this fact. So, now what we will do is we will verify each of these three conditions one by one, starting with this part 1, and, you know, the thing that we need to verify is that this H of X is convex and compact. So, let us begin by verifying that this H of X that we had defined over here is indeed convex. So, towards that, let us consider two arbitrary elements Z1 and Z2, which belong to capital H of X. And in order to show that this set is convex, we need to show that if I take any convex combination of Z1 and Z2, they also belong to H of X. So, in some sense, we want to show that if Z1 and Z2 are two arbitrary points in H of X, then the line segment connecting Z1 and Z2 also lies within H of X.

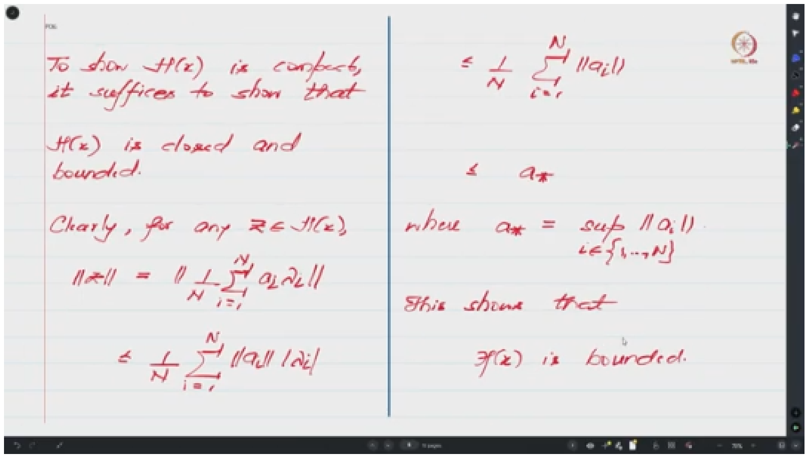
And in order to show that the line segment is contained within H of X, it suffices to pick an arbitrary value of alpha between 0 and 1 and show that this convex combination lies in H of X.



So, if you take this convex combination and since Z1 and Z2 have this form that is mentioned over here, it is easy to see that alpha Z1 plus 1 minus alpha Z2 would have this form over here, where the convex combination results into alpha times lambda I of 1 and 1 minus alpha times lambda I of 2. Now, since both Z1 and Z2 belong to H of X. It follows that for any I which is honest and AI transpose not equal to the expected value of Y of I, it must be the case that the corresponding coefficients equal the sign. So, lambda I1 and lambda I2 must equal the sign of this expression. And what that implies is that for such values of I, these two things will be the same.

So, alpha times this plus 1 minus alpha times this will precisely be this value over here. Now, for any I such that either I is adversarial or I is honest and AI transpose X equals the expected value of Y. Lambda I1 and lambda I2 could be arbitrary values between minus 1 and plus 1. However, this interval minus 1 to plus 1 is convex, which implies that if I take a convex combination of these coefficients, that convex combination will also lie in minus 1 to plus 1. In turn, that implies that for such an I, this belongs to minus 1 to plus 1, right?

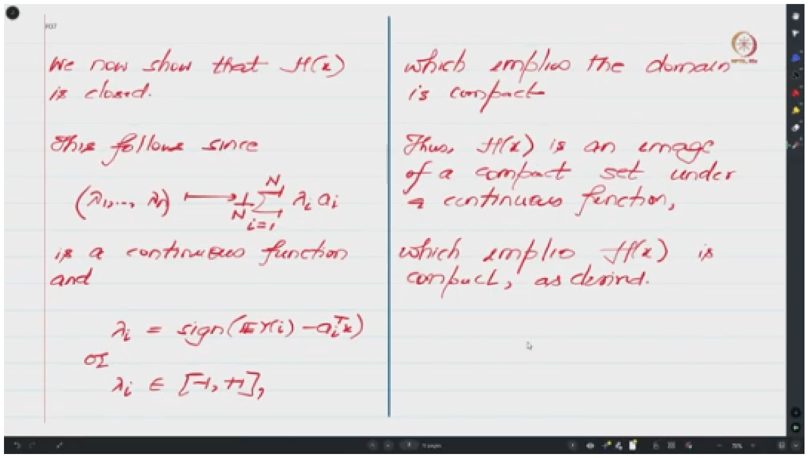
And hence, this convex combination indeed belongs to H of X, right? So, recall that for any vector to belong to H of X, it should have a form like this, right? And for every I for which both these conditions hold, lambda I should equal the sign, and for all other I, lambda I should belong to minus 1 to plus 1. So, which is what we have verified here, and hence this conclusion goes through, which finally shows that the set H of X is convex. So, now we need to show that this H of X set is compact.



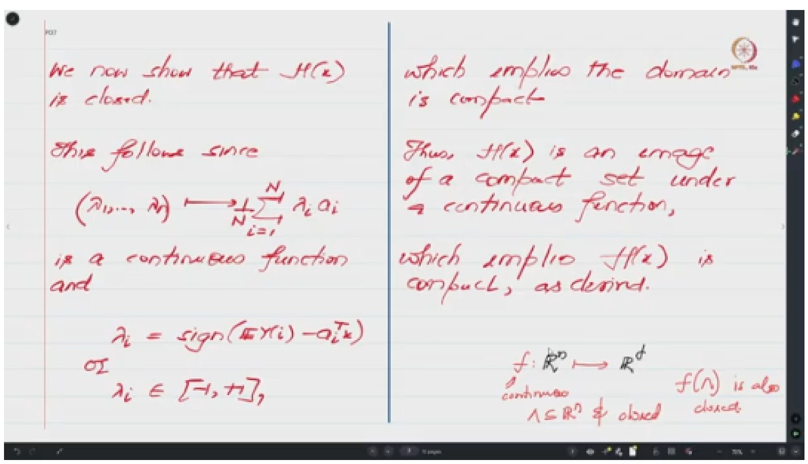
To show a subset of RD is compact, since RD is a Euclidean space, it suffices to show that H of X is closed and bounded. That is an equivalent characterization of a compact set. And in order to show that this H of X is closed and bounded, we will begin by showing, you know, for any X, H of X is bounded. So, towards that, observe that if we pick any Z in H of X, then Z must have a form like this. And hence the norm of this must equal the norm of this expression.

Now, from triangle inequality this norm can be upper bounded by 1 over n times the sum of these norms. Now, you know whatever be the value of lambda i since it lies between minus 1 to plus 1, this absolute value is upper bounded by 1. Hence, this whole expression is upper bounded by this. And one can now conclude that this sum is upper bounded by A star where A star is the maximum of these norms. So, if you replace every norm Ai by the maximum value amongst these values then you will have A star times n and this n and this n will cancel off resulting in this upper bound.

So, from this one can conclude that indeed this calligraphic H of X is bounded. And in order to show H of X is compact, it remains to show that H of X is closed. So, towards that we will make use of one property which says that let us say you have a continuous function between two spaces. Let us say calligraphic A and calligraphic B. And let us say this F is continuous. And let us say you know A in calligraphic A.



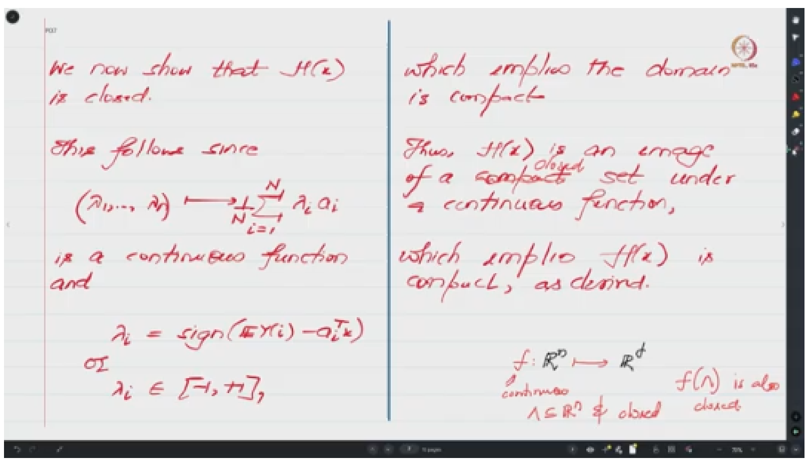
Okay, maybe what I will do is I will not use these notations because I have used calligraphic A to mean something. So, let us say I make use of, you know, let us say Rn to, let us say Rd. So, suppose you have a function f of this form, and let us say you have some capital lambda which is a subset of Rn, and this capital lambda is closed. Let us say f is continuous; then we know that f of lambda is also closed. So, this is a fact from functional analysis. We will be relying on this fact.



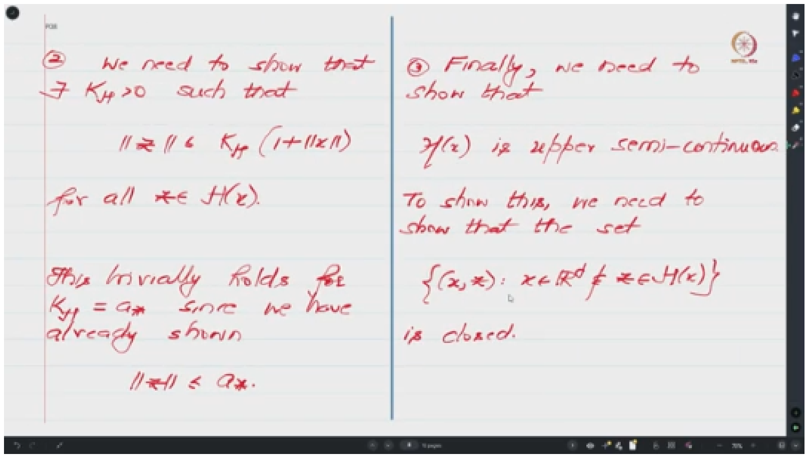
So, one can see that in order to invoke this, what we can see is that for any element h of x that can be constructed using a map of the following form. So, any element in h of x has this nature, right? And one can see that, you know, the characterization of this element is basically the values of these different lambda 1 to lambda n. Hence, one can, you know, construct a map which maps lambda 1 to lambda n to this element in h of x, right? So, this is your function f.

If you want to think in that way. And because lambda 1 to lambda n is mapped to this linear combination, what follows is that this function that we have over here is indeed continuous, right? And furthermore, you know, the input of this map to get h of x should have the property that whenever i satisfies you know, the fact that it is an honest coordinate and AI transpose X is not equal to the expected value of Y, in which case lambda I must equal this or lambda I should belong to minus 1 to plus 1. So, in this sense, one can see that, you know, the set H of X can be viewed as an image of a closed set. I should perhaps say closed set.

of a closed set under the continuous function. So, why do I say it is a closed set? Well, it is the Cartesian product of sets of this form. A singleton set is closed, and this set is also closed. So, if you take the Cartesian product of sets of this form along with



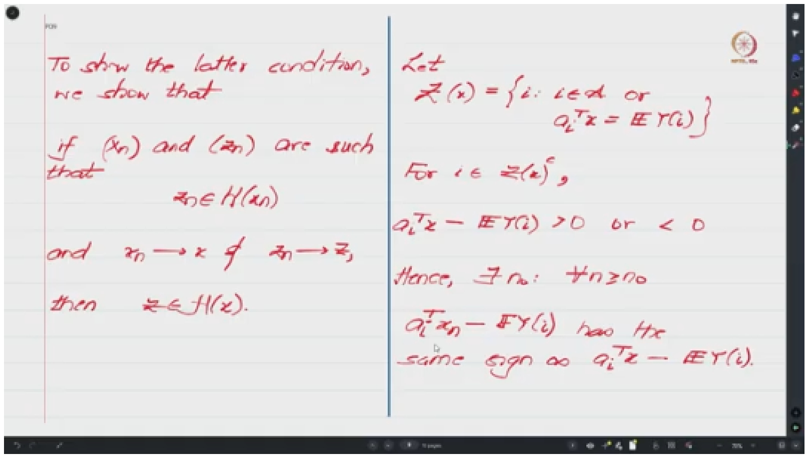
sets of this form, you know, one can see that H of X is indeed an image of a closed set under a continuous function, from which one can conclude that H of X is closed and hence H of X is compact, as desired, right. Now, this sort of verifies our first condition, you know, in the list of conditions needed to verify that this capital H function is merged out. Now, the next thing that we need to verify is this linear growth condition. But this trivially holds from what we have done so far. If you remember, for any element Z in H of X, we have already shown that the norm of Z is upper bounded by A star.



And because this thing is true, one can see that if I take kH to be A star, then this condition trivially holds. So, condition 2 is also verified. Now, the last thing that we need to verify is to show that this function H is semi-continuous. So, I think I need to say that the function H is semi-continuous. Now, let me first provide a definition of what a set-valued function being upper semi-continuous means, and then we will verify this condition.

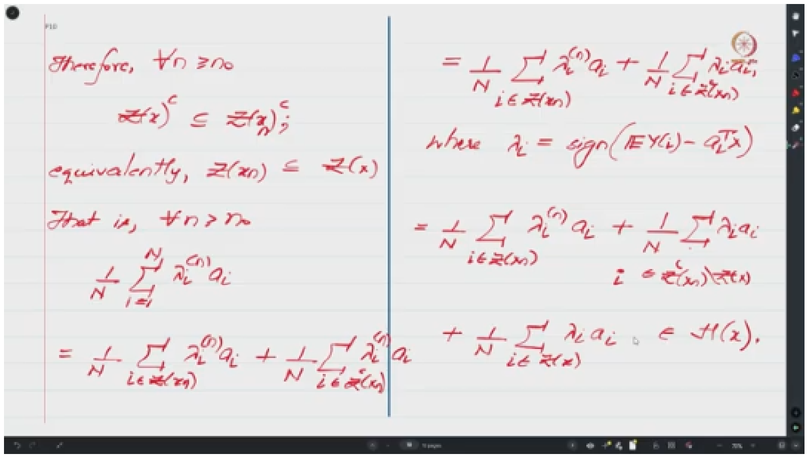
To show that your set-valued function H is upper semi-continuous, we need to show that the graph corresponds to your set-valued function H. You know, which is basically the collection of all x, z where x is d-dimensional, z is d-dimensional, and z belongs to h of x, right? So, you collect all such pairs of points. So, this is like x, z. So, it is a couple of two d-dimensional vectors, right? And they should have the property that x is in Rd and z is in h of x, right?

And this collection should be closed. So, in order to show that this collection is closed, it suffices to show the following: if Xn and Zn are two sequences such that Zn belongs to capital H of Xn, and Xn converges to X and Zn converges to Z, then this limit Z belongs to H of X. So, if you verify this condition, That is equivalent to showing that this set is indeed closed. So, we will now verify this thing. So, towards that, what we will do is we will verify, you know, some set of this form, right? For any input X, let Z of X be the collection of all I such that either I is A, I belongs to A—that is, either I is malicious or adversarial—or AI transpose X equals the expected value of Y. So, key—I mean, just to emphasize why we are also looking at this—this is because



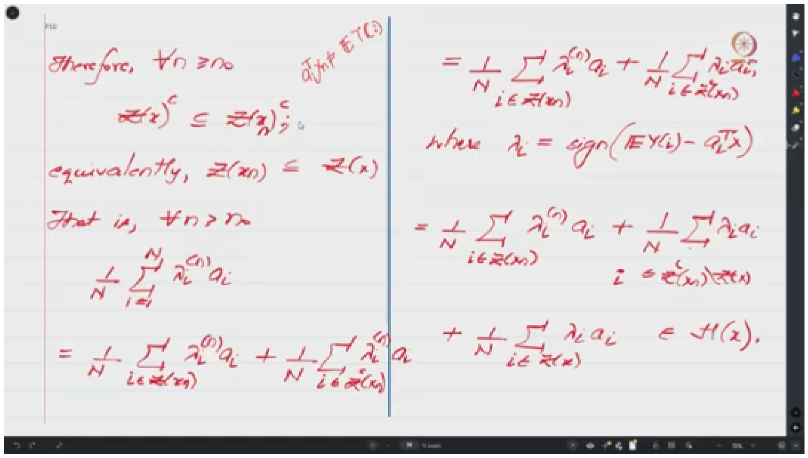
So, this condition would ensure that the input to the sine function is 0, right? In which case, you know, the sine function you can pick any value between minus 1 to plus 1, right. So, Z of X basically consists of all such I's, right. So, now let us see what happens if I belongs to ZX complement, which means that, you know, both it should be non-adversarial and AI transpose X should not be the expected value of Y. For such an I, it is trivial to conclude that AI transpose X minus the expected value of YI will either be greater than 0 or less than 0.

And since Xn converges to X, it follows that Ai transpose Xn minus expected value of Yi will also converge to this expression because this is a continuous function. So, if Xn converges to X, this expression should converge to this. And if this has a specific sign either greater than 0 or less than 0, then this expression eventually should also have the same sign. So that is what I have said that there exists some N0 such that for all N bigger than N0 this expression and this expression have the same sign. Consequently for all N bigger than N0 we can conclude that

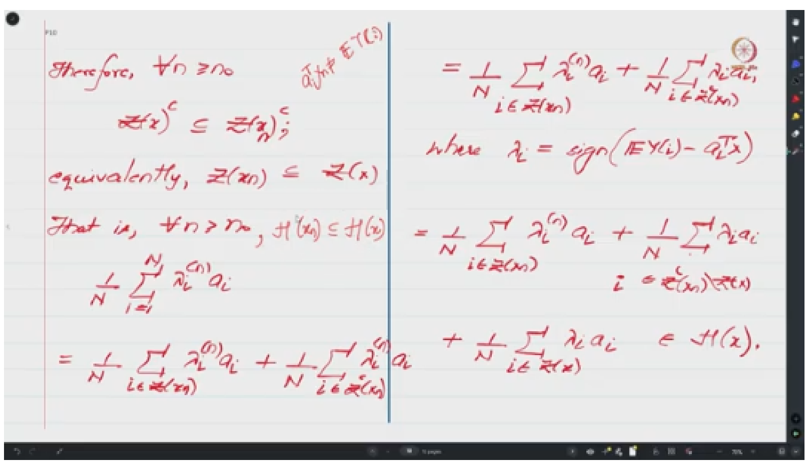


Z of x complement is a subset of Z of xn complement. Recall that Zx you know Zx complement is basically the collection of all those indices which are honest and Ai transpose you know x is not equal to expected value of y. Now what we have managed to show in the previous slide is that for all such i. You know the I will also be in ZXN complement and what is ZXN complement? It is basically the collection of all those I's which are honest and AI transpose XN minus expected value of I XN. is not equal to expected value of Y of i right.

So, for every i that belongs in Z x complement we have already shown that such a i also belongs to this and from this one can conclude that Z of x complement is a subset of Z xn complement. And you know from that one can conclude that ZXN must be a subset of ZX. So, I have removed the complement from here and you know flip the subset relation and hence Z of XN is a subset of Z of X. So, from this what I want to conclude is that For, you know, all n bigger than equal to n naught, your h of xn is a subset of h of x. So, this is something I want to conclude and you will soon see why this is important. So, you will soon see this.



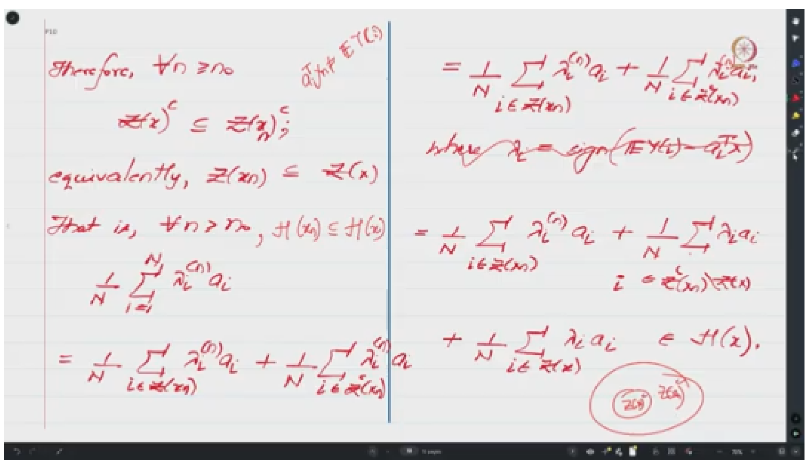
So, first let us verify that this H of Xn is a subset of H of X. So, towards that, what we will do is pick n bigger than n0 and let us consider an element in H of Xn. So, let us consider an element in H of Xn. So, any element in H of Xn would have the following form: that is, it will be 1 over n some i in 1 from n lambda i n a i. So, this is a generic representation of an element in H of Xn, and I want to conclude that any element that is in H of Xn for n bigger than n0 will also be in H of X. So, towards that, what we will do is we will break this sum into two parts.



One is where I belongs to Z of Xn and I belongs to ZXn complement or I does not belong to ZXn. So, you know, we can split this sum into these two parts, and what we have managed to show is that ZXn is a subset of ZX or equivalently ZX complement is a subset of ZXn complement. Is this okay? Now, let us try to understand these two terms here. In particular, let us try to understand what we can say about these lambda i n values, right?

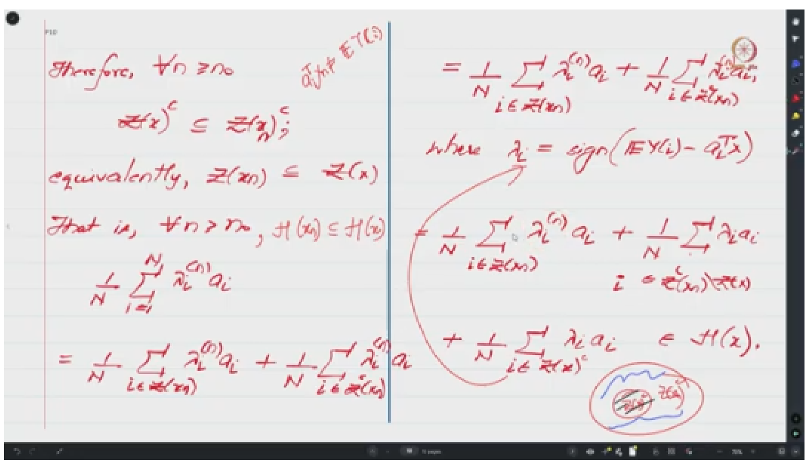
Okay, so for every lambda i n over here, right, for every lambda i n over here, right, if i belongs to ZXn, we leave it as it is, right, and for i belonging to ZXn complement, okay, so I think I have made a mistake in the argument here. What I should say is let this be this. Okay, and what I will do is I will further split this sum into this and this. Okay, so because your ZX complement is a subset of ZXn complement, right, we can split this ZXn complement into something that belongs to ZXn complement but does not belong to Z of X. And the last terms which belong to Z of X, right?

So, what I am trying to do is you have your ZX here, right? And you have your ZXN over here, right? So, ZX complement here, right? And ZXN complement here. So, we have managed to show from this fact that ZX complement is a subset of ZXN complement, right?

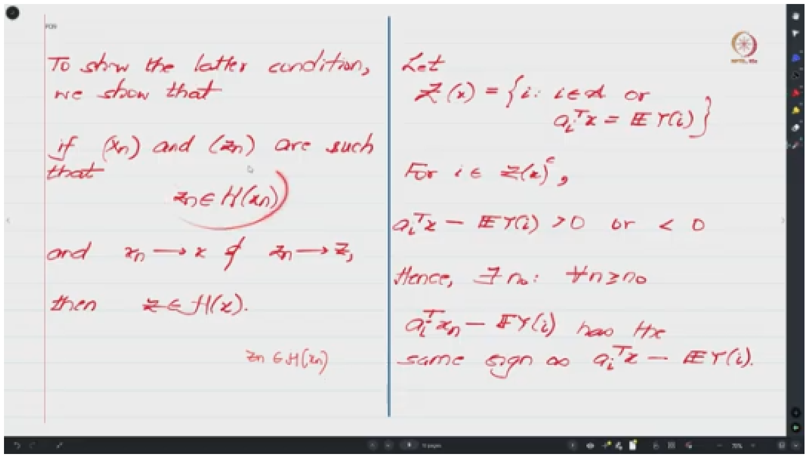


So, I should be able to split your i in ZXN complement in the following way: that is, either you belong to this blue region or you belong to this black region, right? And hence, if it belongs to this, then what we can do is we can split this up in the following way. Is this okay? So, we can split this up in the following way: so, i in ZXN complement can be split as i in ZXN complement minus Z of X. That is, it lies in the blue region or i belongs to ZX complement. Is this okay? So, in ZX complement, the value of lambda i should precisely be what is mentioned over here, right?

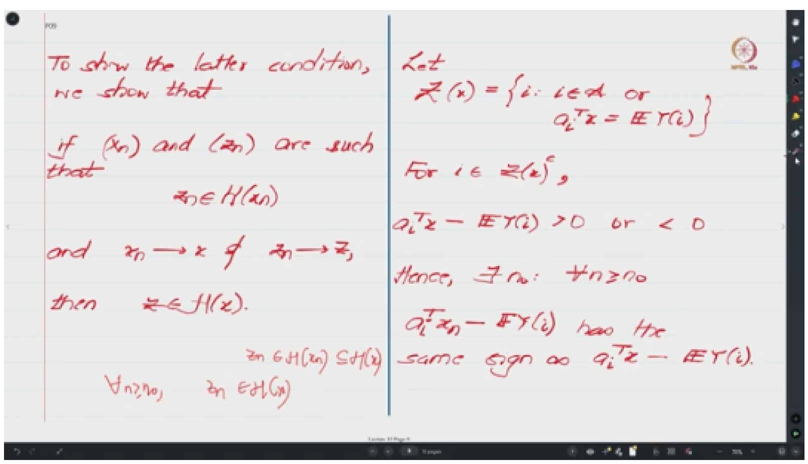
So, for every i over here, your lambda i should have this value, and here lambda i n can be some arbitrary value. And here, lambda i could either be, you know, plus 1 or minus 1, since it belongs to ZX complement; it belongs to plus 1 or minus 1. But you can see that, you know, for a sum of this form to belong to H of X, we require that for every i in ZX complement, this lambda i should match this sign, which is something we have shown already. I mean, that was the characterization of your n0. And since for all i in ZX complement, lambda i indeed equals this value, one can then conclude that because these lambda i n's and the lambda i over here I should again say this is lambda i n. So, these set of values can be anything between minus 1 and plus 1. One can use this fact to conclude that this expression over here indeed belongs to H of X. So, what we have managed to show is that for n bigger than or equal to n0, if I pick an arbitrary element in H of Xn, then that arbitrary element must indeed belong to H of X. And hence, one can show that H of Xn is a subset of X. So, what we have, you know, from this, what we can conclude is that eventually, so, you know, your Zn,



Belongs to H of Xn, right? You know, that is something that is required over here, right? And we have now managed to show that Xn is a subset of X, which implies that for all n, for all n bigger than n0, Zn belongs to H of X, right? And previously we had shown that this H of X set is actually closed, right?



And we now have that the tail of this Zn sequence belongs to H of X for all n bigger than n0, that is the tail, right? And we also know that Zn converges to Z. So, from the fact that H of X is closed, one can now conclude that Z belongs to H of X, as desired. So, this brings us to the end of today's class. So, let me quickly summarize what we have done in today's class. In today's class, we tried understanding whether the set-valued function H is Marchaud or not, and towards that, we verified whether H of X is compact and convex.



Then we verified whether, you know, H of X has this linear growth condition, and finally, we showed that the set-valued map H is upper semi-continuous. So, these are some technical conditions, and you know, I have verified such conditions for the example algorithm that we are considering. So, one can ask, you know, why do we need to show that this H is Marchaud? Well, you know, the purpose of H being Marchaud is to ensure that for any initial condition, the corresponding differential inclusion guarantees existence of solutions, right? So, this is the fact, and one can, you know, in some sense, imagine this Marchaud property being a set-valued generalization of the Lipschitz continuity that we had studied when we were discussing existence and uniqueness of solutions. So, there the Lipschitz continuity ensured uniqueness and existence of solutions, but in the context of differential inclusions, the Marchaud property only guarantees existence of solutions.

In the next class, what we will see is why any solution trajectory of this differential inclusion corresponding to this stochastic recursive inclusion that we are considering should behave. So, here, you know, the goal is to find the expected value of X. So, what we would like to ask is why the solution trajectories of this DI, you know, converge to the expected value of X. If that does not happen, then, you know, we cannot hope these corresponding stochastic recursive inclusions to do magic. However, if the solution trajectories of the differential inclusion are well-behaved, then, you know, we can identify conditions under which the limiting behavior of your stochastic recursive inclusion, okay, mirrors those of the solution trajectories of your differential inclusion. So, this is something we will verify in our next class. Until then, goodbye and namaste.

See you then. Bye.