STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Engineering

Indian Institute of Science, Bangalore

Week 11

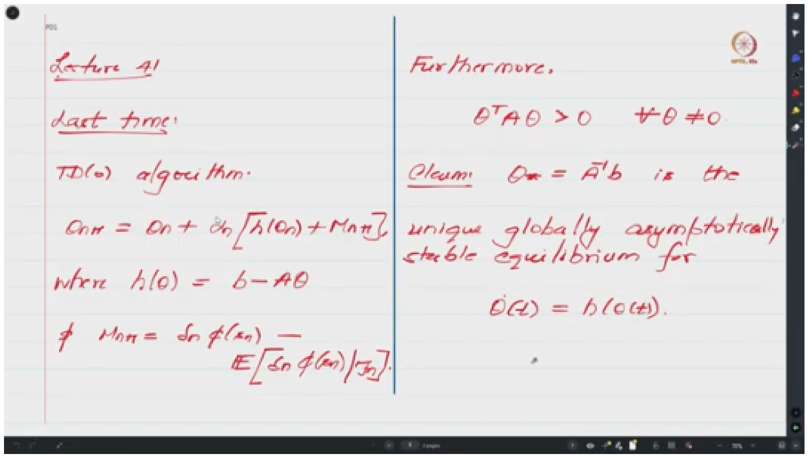
Lecture 41

How Good Is the TD Solution? Fixed Point Analysis in Linear Approximation

Hello and Namaste everyone. Welcome to lecture 41 of this NPTEL course on Stochastic Approximation. So, as you recall in this week's lectures and the next couple of lectures, we are looking at applications of Stochastic Approximation, in particular in the context of reinforcement learning. And over the past two lectures, we have focused on this problem called policy valuation, where given a policy, we would like to quantitatively say how good that policy is. And one way to do that is via what is known as the value function.

And in the last two lectures, we have tried you know, identifying the value function or estimating the value function of a policy pi when the state space is very large, and in that case we have been forced to work with this concept of function approximation, right? And in particular, we looked at the linear function approximation case where we are given a matrix V and the goal is to find within the column space of V the best approximation to your V pi vector. That was the goal, and towards that we came up with the TD0 algorithm. and we looked at the limiting ODE associated with this TD0 algorithm and then we asked what can we say about the, you know, asymptotic behavior of the solution trajectories of this limiting ODE. So, we will continue the discussion from here, and you know, over this lecture and the next, not only will we try to understand the limiting behavior of the asymptotic limiting ODE, but I mean limiting ODE, but also about the asymptotic behavior of the TD0 algorithm itself.

With this background, let us begin our formal discussion. So, here is the update rule for the TD0 algorithm. So, the TD0 algorithm has the update rule theta n plus 1 equals theta n plus alpha n times h of theta n plus m n plus 1, where h of theta has this linear description, that is b minus A theta, and m n plus 1 is your martingale difference noise which has this expression that is given over here. And in the previous class, we showed that this matrix A is positive definite, and positive definite means that theta transpose A theta is greater than 0 for all theta which is not equal to 0. And I would again like to highlight that.



This matrix A is not symmetric, but nevertheless, we call it positive definite because it satisfies this condition over here, right?

$$\theta_{n+1} = \theta_n + \alpha_n\left[h\left(O_n\right) + M_{n+1}\right]$$

$$h(\theta) = b - A\theta$$

$$M_{n+1} = \delta_n \phi\left(s_n\right) - E\left[F_n\right]$$

$$\theta^T A\theta > 0, \quad \forall\ \theta \neq 0$$

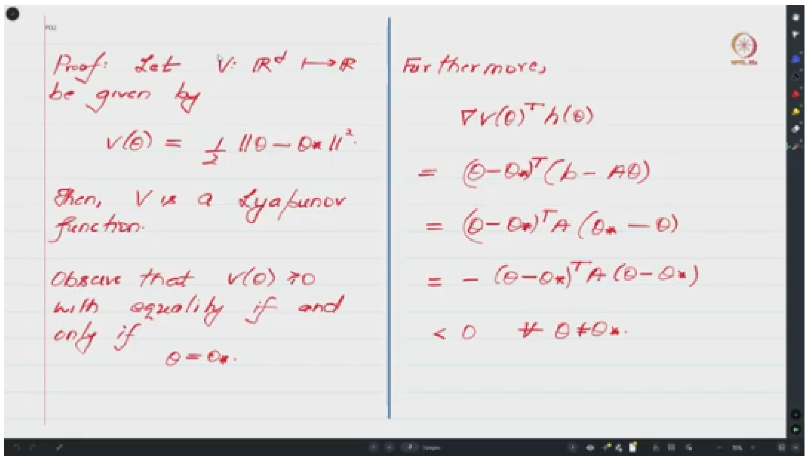
And the claim that we would like to formally show today is that theta star, which is defined to be A inverse B, right? This is the unique globally asymptotically stable equilibrium for this ODE. Theta dot of T equals H theta of T, right?

$$\theta_* = A^{-1} b$$

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

So, there are multiple ways of showing this, and one of the ways is via Lyapunov, I mean, is by using Lyapunov functions.

So, towards that, we will try to cook up the following Lyapunov function. So, let V be a function which goes from R^D to R. Recall that your theta star lies in the R^D space. So, V is this function from R^D to R which is given by half of the Euclidean distance between theta and theta star. So, our goal now is to show that this function is a Lyapunov function.



In particular, we need to, you know, satisfy different properties of a Lyapunov function, right? And that is what we are going to do. So, towards that, this is the claim that V is a Lyapunov function. So, towards that, the first thing that we will establish is that V of theta is non-negative and it equals 0 if and only if theta is theta star. But this is obvious from the definition.

Next, we will look at the inner product between grad V of theta and H of theta, and one can see from this description of V of theta that grad V of theta is simply theta minus theta star, and since there is a transpose, I have a transpose here, and H of theta is B minus A theta. Now, if you recall, b, I mean theta star is A inverse b, hence b equals A theta star. So, if you take A out as common, we would end up with theta star minus theta. You can see that here you have theta minus theta star, and here you have theta star minus theta. So, I will take the negative sign out, and that will lead me to the expression minus of theta minus theta star transpose times A theta minus theta star, right? And we have previously shown that.

Any expression of this form is positive, and because of the negative sign here, one can conclude that this expression is strictly less than 0 for all theta not equal to theta star because when theta equals theta star, this will be 0. So, except for the case where this quantity is 0, we know that this quantity is positive, and hence this overall quantity will be negative. So, we have managed to show that the inner product of the Lyapunov function and the driving function is negative for all theta not equal to theta star.

$$V(\theta) = \frac{1}{2}\left|\left|\theta - \theta_*\right|\right|^2$$

$$\theta = 0_*$$

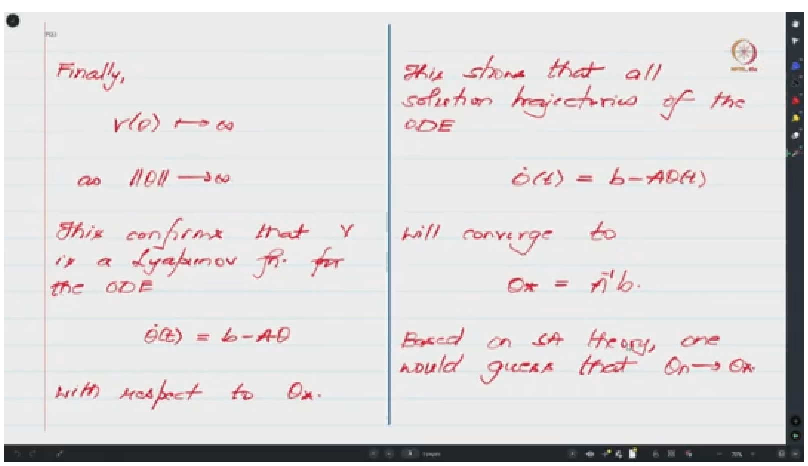
$$\nabla V(\theta)^T h(\theta) = \left(\theta - \theta_*\right)^T (b - A\theta)$$

$$= \left(\theta - \theta_*\right)^T A\left(\theta_* - \theta\right)$$

$$= - \left(\theta - \theta_*\right)^T A\left(\theta - \theta_*\right)$$

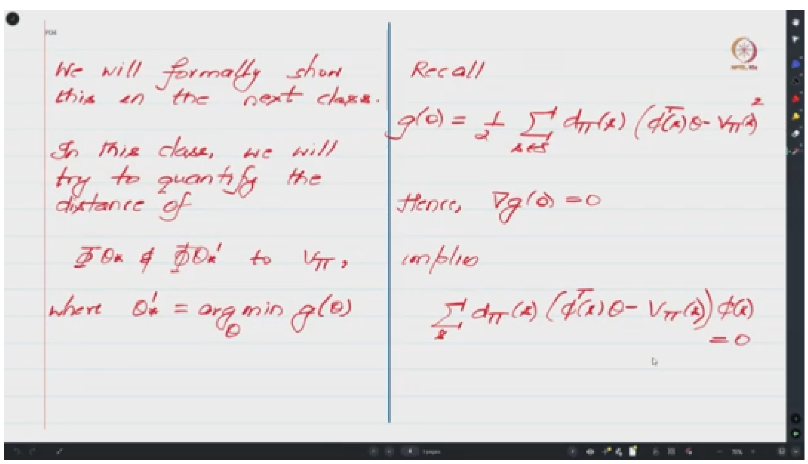
$$< 0, \quad \forall \ \theta \neq \theta_*$$

Finally, it is also easy to see that the value of V of theta goes to infinity as the norm of theta grows to infinity. Because all these conditions hold true, one can now conclude that this V is actually a Lyapunov function for the ODE theta dot of t equals b minus A theta with respect to your theta star. Now, again, I would like to recall for you what is the significance of a Lyapunov function.

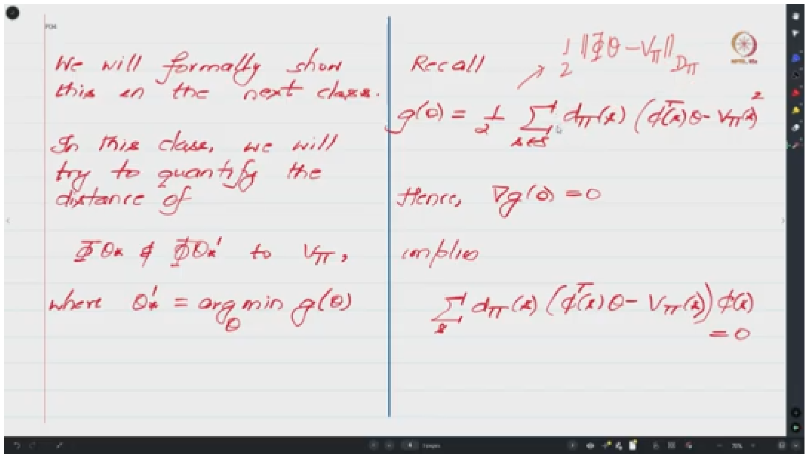


It means that if you take any solution trajectory of this limiting ODE and study the behavior of the value of V along the solution trajectory, then the value of V will keep decreasing. And since it is going to strictly decrease at every point in time, one can show that the value will continue decreasing until it reaches its minimum value, which is 0 in this case. And recall that V of theta can be 0 if and only if theta equals theta star. So, from that perspective, one can conclude that all solution trajectories of your limiting ODE, which is given over here, that is theta dot of t equals b minus A theta of t, will indeed converge to theta star, which is A inverse b. So, in the next class, what we will do is we will use stochastic approximation theory to show that your theta n converges almost surely to theta star.

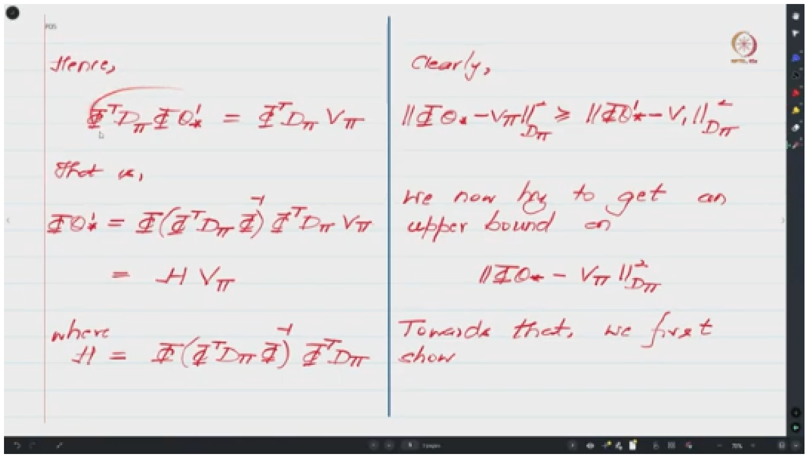
But in this class and perhaps also a part of the next class, what we will ask is or what we will focus on is Okay, if your iterates converge to theta star, right, is this good or not? So, we will try to answer that question first and then, you know, try to establish that theta and we will go to theta star. Okay, so what we will now try to do is. You know, theta star, if let us say this is the quantity that we manage to estimate, then phi theta star would be our estimate of V pi, right?



And now what we want to do is we want to ask ourselves how good an approximation is phi theta star of V pi. In particular, we will try to compare the distance of phi theta star to V pi, right? As against the distance of phi theta star prime to V pi where theta star prime we will define it to be the quantity that minimizes g of theta right and g theta recall is this function over here right. So, g theta if you recall can also be viewed in the following sense it is half times the norm of phi theta minus V pi right. So, d pi is the diagonal matrix whose diagonal entries are made up of these little d pi's and recall that little d pi is the stationary distribution associated with the Markov chain induced by your policy pi.



Right, so g theta is this quantity, and if we somehow manage to minimize this quantity, right, then we have the best possible approximation to V pi in the column space of phi under this metric d pi, right? And you know, in the previous class, I had done this calculation, but let me recall it for you. Let us quickly find what is theta star prime, right? So, if I want to minimize this quantity— Then, I will take its gradient, and because this is a quadratic function, if we set the gradient equal to 0, that would be the value—I mean, solving this expression, whatever theta we get, that will be the quantity that will minimize this objective function. Hence, let us set this quantity to 0. Right, and one can see that this expression compactly leads us to the following result: that is, your theta star prime, which minimizes your g theta, satisfies the equation phi transpose d pi phi theta star prime equals phi transpose d pi V pi, right? And if I take this matrix over here— That will give me the inverse and now if I multiply by phi on both sides one can then see that phi theta star is equal to this expression over here.



And for convenience, what we will do is, this quantity that is the quantity just before V pi, we will denote it as H. So, H is the matrix that is given by this expression over here. So, let us pause here and try to understand. The meaning of theta star prime, right. So, in some sense, phi theta star prime is the best proxy for V pi in the column space of your matrix phi, right. So, it is best with respect to this metric that we define, that is the metric with this, you know, induced matrix D pi, right.

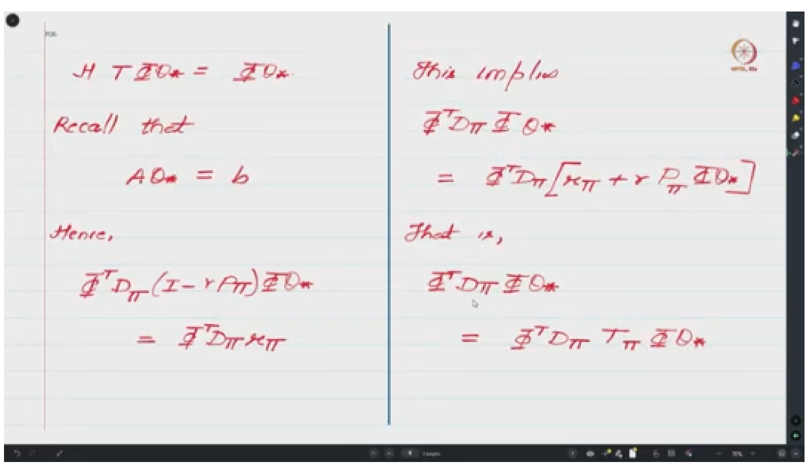
Or the metric that is induced by this matrix d pi. So, under this norm, your phi theta star prime is the best proxy for v pi, right. And if you know the concept of projections, one can see that your phi theta star prime would be the projection of v pi onto the column space of And hence, you can view your H as a projection matrix. So, whatever vector you place over here, H times that vector will be the projection of that original vector onto the column space of phi. And this d pi and all those things come into the picture here because instead of looking at the Euclidean projection, we are looking at the projection with respect to the norm that is induced by this matrix d pi, right?

So, you know, the best proxy for v pi within the column space of phi would basically be phi theta star prime, which is, you know, denoted as the projection of v pi. And now, the question that we would like to ask is, how good is phi theta star, right, compared to phi theta star prime? And we are going to derive some intermediate bounds now, and in the next class, we will formally obtain the relationship of the distance between phi theta star V pi and phi theta star prime and V pi. So, I should have V pi over here. So, let us first look at this inequality.

One can see that this inequality is straightforward. This is straightforward because of whatever I told you that theta star prime is the vector that minimizes this distance. And theta star is some vector that is not theta star prime and hence from the definition of theta star prime one can immediately conclude that the distance of phi theta star to V pi will be greater than or equal to the distance of phi theta star prime to V pi. Now we would like to understand How worse can this distance be, right?

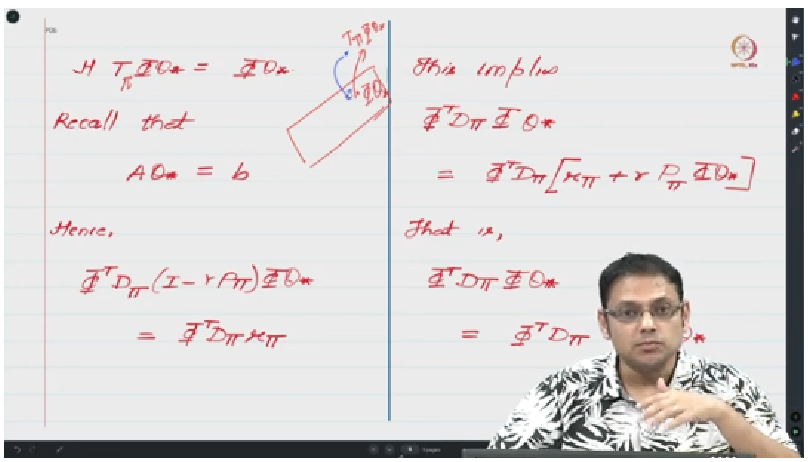
So, recall what is the TD0 algorithm and how did we end up with this theta star? Ideally, we wanted to minimize this expression, right? But we saw that the SGD algorithm with respect to this objective has V pi, which is something that we do not know, right? And then we sort of replaced V pi by phi theta n and that led to you know some variant of the SGD algorithm defined with respect to this objective.

And this modified algorithm we are now guessing that it converges to theta star and the question that we would like to ask is the limit of this TD0 algorithm how does it compare to the best approximation to V pi that we can get. So, towards answering this question, this lower bound is obvious but now we would like to get an upper bound on this distance that is phi theta star minus V pi d pi square. So, towards that we need some intermediate results which is what we are going to do now. So, the first intermediate result that we would need is that this phi theta star that we have defined satisfies this equation over here. So, there is a typo here.



This should be T subscript pi. So, the first equation says that H times T pi of phi theta star equals phi theta star. So, in other words, if you think of H T pi as an operator, that is you take a vector V. Apply first T pi on it. So, recall that T pi is the Bellman operator associated with your policy pi.

So, you apply T pi onto it and then whatever vector you get, you project that vector onto the column space of phi. So, what this relationship tells us is that phi theta star is the fixed point of the operator H T pi that is you start at phi theta star apply T pi that is the Bellman operator and then you apply the projection of phi theta star you will get back to phi theta star. So pictorially if you sort of imagine this to be your column space of phi right so phi theta star let us say is over here right you apply T pi it will take you outside you know the column space of phi and now what you do is you try projecting whatever is this vector back onto the column space of phi that is find the expression in the column space of phi that is the best

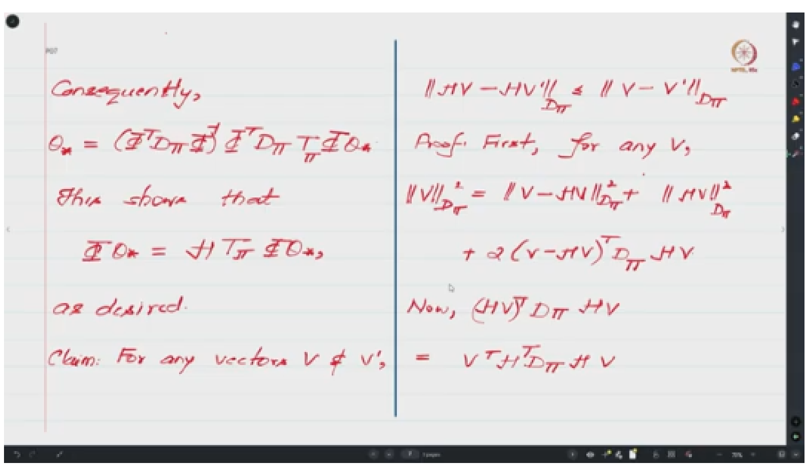


representative of this expression that is t pi phi theta star and again best over here means that you know you define this metric that is induced by your diagonal matrix d pi and try to find the vector within this column space of phi that minimizes the distance to t pi phi theta star. And that would indeed be the projection that is the one that is obtained by multiplying by H. And what this result says is that your phi theta star is the fixed point of this H T pi operator. So, now we are going to verify this and we will verify this in an algebraic fashion. So, recall that A theta star equals B. This is how we had defined theta star. So, theta star was A inverse B which implies that A theta star must be B. So, now if you substitute the definition of your A matrix.

So, a matrix recall is phi transpose d pi i minus gamma p pi phi. So, if I multiply this or put this in place of A, I will get this expression on the left-hand side and the B vector is phi transpose d pi r pi. So, by you know expanding this multiplication observe that I will end up with phi transpose d pi times the identity matrix times phi theta star which is what I have written here. Then I have phi transpose d pi r pi which is what I have here and I will also have phi transpose d pi gamma times p pi phi theta star. Right?

So, if you take this phi transpose d pi common from this expression and this expression, you will get it out common here. This r pi fits in here and this gamma p pi phi theta star is what you would have over here. Right? And this expression is basically your t pi phi theta star. You can go back and check the definition of t pi and one can see that this is exactly your t pi phi theta star and hence whatever you have over here, I can write it here which will give me

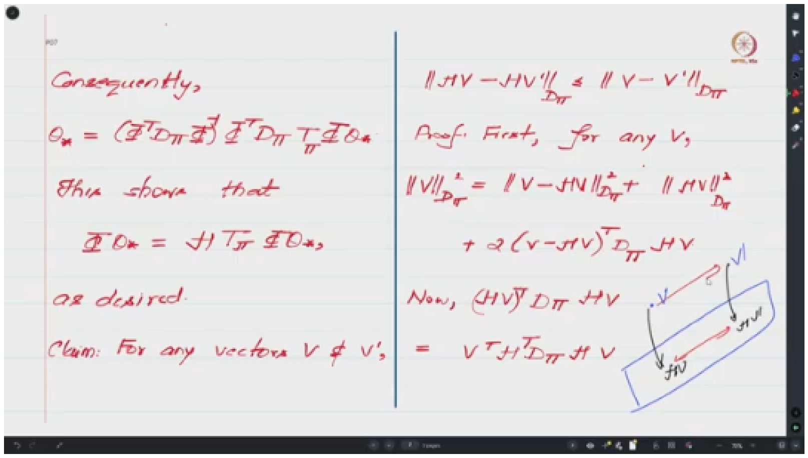
phi transpose d pi phi theta star equals phi transpose d pi d pi phi theta star. So, this is the expression that we end up with. Now, what we will do is this phi transpose d pi phi as I told you if you assume your phi matrix is having full column rank and your Markov chain is ergodic which ensures that the d pi matrix is having strictly positive diagonal entries and hence it is invertible. You know this phi transpose d pi phi then will be invertible and hence I can take it to the right-hand side in the form of an inverse and then I can multiply both sides by phi which will show that phi theta star is indeed equal to phi times this expression times T pi phi theta star and phi times this expression is precisely H and hence we end up with the relation phi theta star equals H T pi phi theta star, right?



And this was the relation that we wanted to prove, and we are over there. So, let us just summarize what we have managed to show, right? So, right now we are guessing that based on the limiting ODE associated with your TD0 algorithm, we are guessing that the iterates of the TD0 algorithm will go to theta star, and what we have now shown is that this theta star is special in that phi theta star is the fixed point of the projected Bellman operator T pi, right? So, this is what we have managed to show, right, and recall what we are trying to do now.

We are trying to, in some sense, see what is the distance of phi theta star to V pi and how does it compare to the smallest possible distance between V pi and the column space of phi, which is basically the distance between phi theta star prime and V pi, right, and we are proving some intermediate results and we will prove the final bound in the next class. So, the next intermediate result that we need to show is the following. If you have any two vectors v and v prime and if you look at the projection of v, that is H v, and the projection of v prime, that is H v prime. So here, projection means you take a vector and, you know, try to find the best proxy for this vector in the column space of your phi matrix, and of course here best proxy means with respect to this norm that is induced by this matrix d pi. So, if you look at H v and H v prime, that is the distance between the projections of v and v prime under this norm, then the claim is that the distance between the projected quantities is upper bounded by the distance between the unprojected quantities.

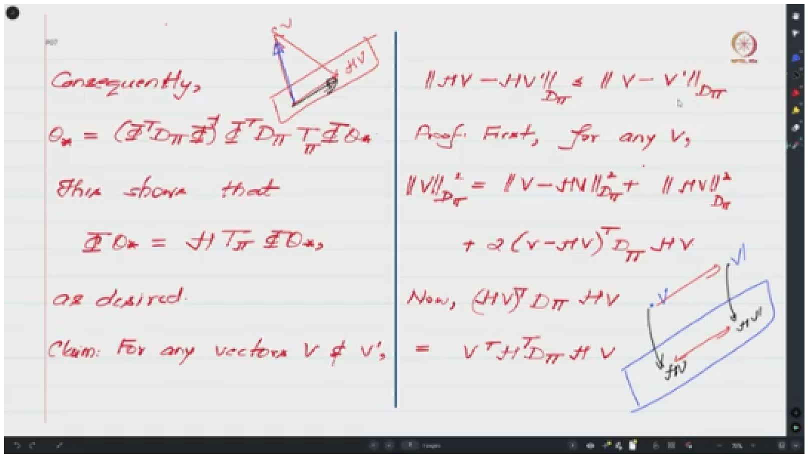
So, pictorially what we are saying is that if this is your column space of V and let us say you had v here and v prime here, and if you project these quantities onto the column space of V so that you end up with H v here and H v prime what we are claiming is that if you look at the distance between these two points and compare that distance to the distance between v and v prime, then this relationship that is stated over here holds. So, let us try to prove this thing; I mean, this result is common to any of your Pythagoras theorem, so not very different. So, for any vector v, I hope you agree that if you take the norm of v induced by this matrix d pi.



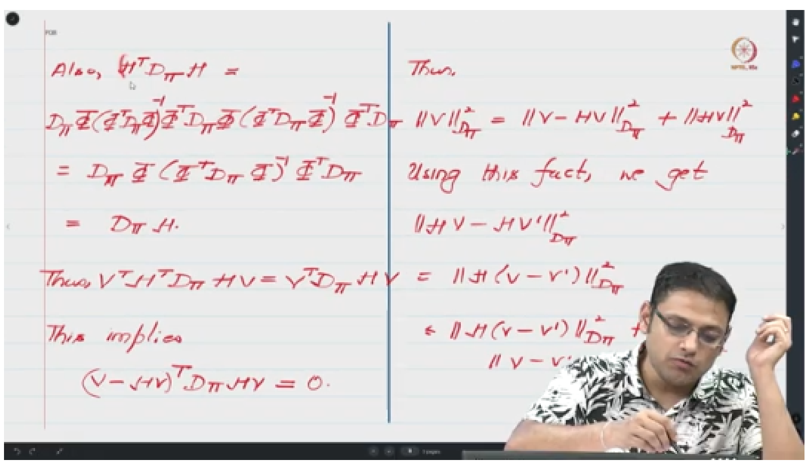
So, if you take the square of it. So, if I add and subtract hv, one can see that this norm squared equals the square of this distance plus the square of this distance plus the cross term. That is, you take v minus hv. And you take HV and then you have to multiply d pi. The reason you have to multiply d pi over here is that you are working with this metric that is induced by this matrix d pi.

So, hence you will end up with this cross term. Now what we will do is we will try to show that this cross term is actually 0. So, once we show this cross term is 0, one will get the usual Pythagoras theorem that is the sum of the hypotenuse squared equals the sum of squares of the different sides of a triangle. So, if you think of this, what this result is trying to tell is basically that you have some column space of V here, you have V, you project it, you will end up with HV. Okay, and v minus hv is this vector over here and here is your origin, right? So hv is basically this vector and then you have v, okay? So this is these three vectors, okay? So maybe I'll put this as one color and maybe this in another color, right? So if you think of this as a right angle triangle, so notice that this

Angle over here, I have not drawn it as a right angle. This is because this is not your usual Euclidean projection. It is projection with respect to this norm that is induced by this matrix over here. So, here one can see that this is your hypotenuse. So, V squared is what you have over here.



This equals this square plus this distance square. This is the result that we are aiming to show, and we will show that result by showing that this inner product is 0. So, towards that, let us consider this expression: that is, H V transpose d pi H V. So, we will look at this expression over here, and simple algebra shows that this expression will equal V transpose H transpose d pi H V. And now, if you recall the expression for H, we have H equals—sorry, H equals this expression over here, right?



And hence, H transpose will be this expression over here, right? And you have d pi in between. So, you have d pi over here. So, one can see that this expression will cancel off because you have phi transpose d pi phi and phi transpose d pi phi inverse over here.

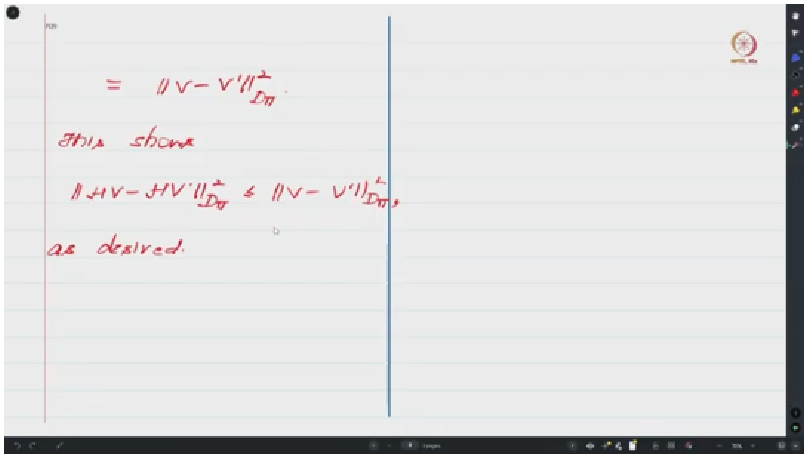
So, this and this will cancel off, and this will lead us to the expression d pi phi phi transpose d phi phi inverse phi transpose d pi. So, this is what we end up with, and one can immediately see that this expression is precisely H, and hence one can conclude that

H transpose d pi H is precisely d pi H, and from this one can see that V transpose H transpose H V equals V transpose d pi H V, and from this fact one can see that this inner product is 0 because the first term will be V transpose d pi H V, which is this quantity, and the second term which is what is written as the left-hand side here—this also equals this quantity. Hence, this inner product will be 0. So, because the inner product is 0, one can now immediately conclude that the norm of V, or the square of the norm of V, equals the square of these two norms, right? And from this fact we can easily see that if you look at the square of the distance between H V and H V prime, which was the quantity, you know, which we want—I mean, of interest with regards to the statement. So, if you look at the distance between H V and H V prime, right, and look at the square of this distance with respect to this norm, right, and because H is your matrix, I can pull H in common and say that this distance is basically the length

of this vector. So what is this vector? You take v minus v prime and take its projection onto the column space of phi. So this is this distance. Now what I will do is I will take this quantity and I will add this quantity to the right-hand side.

And since this quantity is non-negative, this quantity will be upper bounded by this thing. But whatever you have over here, if you compare it with what you have over here, you can see that this expression equals this expression if instead of V, we had V minus V prime. So you can see that this quantity is precisely this quantity with V replaced by V minus V prime and here again this quantity mirrors this quantity if you replace V with V minus V prime at both places. So, from the Euclidean, sorry, from the Pythagorean theorem that we have shown above, one can see that this right-hand side actually will equal, you know, the distance between V minus V prime squared.

So, what we have managed to show now is that the square of the distance between HV and HV prime is actually upper bounded by the distance between V and V prime. So, now we have shown it with respect to squares. However, since these quantities will be non-negative, one can show that the conclusion also holds true without the squares and this is what we wanted to show. So, now let me quickly summarize what we have done in today's class. So, in today's class, we first of all showed that your theta star indeed is the globally asymptotically stable equilibrium for the limiting ODE associated with your TD0 algorithm.



That is the first thing that we showed, and we showed this by cooking up a Lyapunov function. Thereafter, we wanted to compare the distance between phi theta star and the projection of V pi onto the column space of phi, which we denoted by phi theta star prime. So, we wanted to see how good phi theta star is compared to phi theta star prime, and towards that, we derived some intermediate results and we will make use of these intermediate results in the next class. To, you know, obtain a relation or an upper bound between the distance of phi theta star to V pi in terms of the distance of phi theta star prime to V pi. Recall that the distance of phi theta star prime to V pi will be the smallest possible because that is how we have defined theta star prime.

So, in the next class, we will discuss these things. Until then, goodbye and namaste.