

ENPTEL

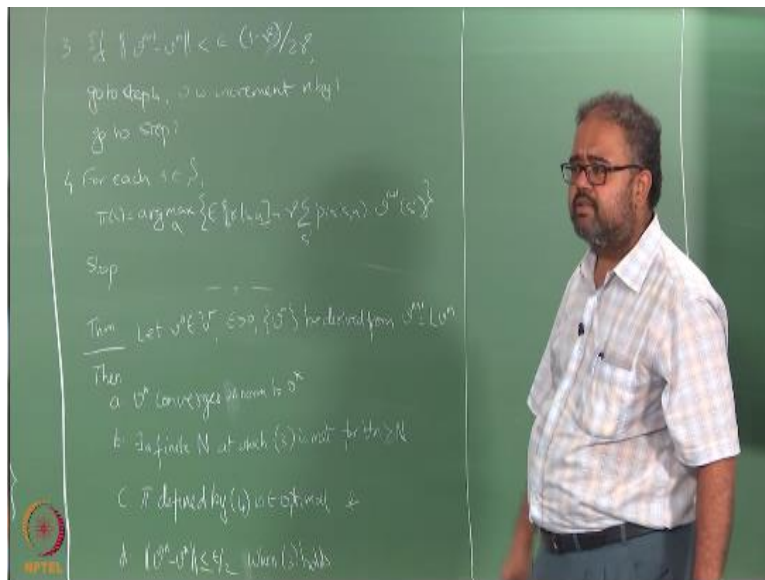
NPTEL ONLINE COURSE

REINFORCEMENT LEARNING

Value Iteration Proof

Prof. Balaraman Ravindran
Department of Computer Science and Engineering
Indian Institute of Technology Madras

(Refer Slide Time: 00:13)



So here is a value iteration theorem can read all of it I apologize I go τ down the burden handwriting series dignity rates at which this condition in three is met for all N greater than capital in okay so Π defined by step four is ϵ optimal and $V^{N+1} - V^*$ is less than or equal to $\epsilon/2$ when 3 holds so when this holds V^{N+1} is ϵ close to be V^* right so what is it A and B of kind of easy we already seen that A and B as essentially Brouwer's fixed point theorem right.

I am sorry yeah the Brouwer's fixed point theorem gives us A and B correct what is the difference between C&D why do I have two of these things so they look contradict they look contradictory not order it Reaper safe but we did not anyone gives me the right answer I will give

you a bonus mark that essentially which have been following whatever I have been saying so far so further we will have to take down the roll number and note down that now I got their attention two people who do not know the class averages for my wishes tend to be in like 2 or 3 out of 20 or 50 it does not matter what the denominator is does not seemed to change but so every mark matters so tell me what is the difference why do I need C and D sorry this ϵ optimal means it is so maybe she phase say it in a different way then you will get it so what I am saying is that $\|V\pi - V^*\|$ is less than ϵ here I am saying $\|V_{N+1} - V^*\|$ is less than $\epsilon / 2$.

Remember my space V does not necessarily contain only value functions it is just a space of functions right V_{N+1} need not be the value function corresponding to the policy π even though I recover π by being greedy with respect to V_{N+1} does not mean if I actually solve for value of π if I find $V\pi$ it need not be V_{N+1} right so what three is telling me is that $\|V\pi - V^*\|$ is ϵ not what C is telling me is $\|V\pi - V^*\|$ epsilon optimal what D is telling me is that $\|V_{N+1} - V^*\|$ is $\epsilon / 2$ optimal okay so fun let me give you the work how about you enough I never offered them before right.

This is this is on offer nobody took it that is easy enough right all you need to do is just mean following the multiple times I kept saying that V is not the space of value function is the very beginning when somebody said V is a space of value function I corrected you of this class very beginning of this class anyway so let us look at the proof well A and B are immediate the follower Brouwer's fixed point theorem so we will start off with C and D right so just assume that 3 is made for some n okay so n it is already met that condition has been met right and that we π the π has been derived from doing this the π has been derived from doing this.

And now I can use my triangle inequality right it does not matter that then this actually does not matter right whether these conditions are met or not I can take an arbitrary $V\pi$ and a V^* and I can use a triangle inequality and write this because it is just about three points in space right we talked about how I can keep using the triangle inequality in an arbitrary fashion because all I am talking about are just some 3 points in space it so happens that one of them is $V\pi$ other one is V_{N+1} the third one is V^* but this will hold regardless of which three points I pick okay.

So here is a tricky thing I want you to look at so let us look at L_π say I want L_π to act on V_{n+1} right I want L_π to act on B_{n+1} right so what will be L_π acting on V_{n+1} is it space of what I π is an operator corresponding to the bellman equation for Π right sort of way I said last class we left off by saying L_π is a contraction so that L_π so L_π is just a bellman value equation operating on V_{n+1} so what would that be will be this right this but instead of $R \max$ here I will have a summation over π Sa say allow summation over π Sa but what I have done here by π itself is an $R \max$ a right just one action.

So if you remember I told you about the deterministic policy is right so this whole thing will go I will just replace this with π S is this will be expected value of R given S, π S probability of s' given s come up is V_{n+1} so that will essentially be the L_π acting on V_{n+1} right is it clear this is L_π on V_{n+1} I just read it out I just write in unit known as mnemonic expected value of R given s, $\pi s + \gamma$ times summation over s' probability of s' given $\pi s \times V_{n+1}$ okay so now thing that you have to notice how would I get to this πs is by using max I use the max operator over this right.

So if I had done L on V_{n+1} what would I have obtained L on V_{n+1} is essentially this right next right this is this is what this one L or V_{n+1} is essentially this right so what I have done here because I chose my π to be the max action here so whether I apply L_π on V_{n+1} or whether I apply L on V_{n+1} it is the same because the π was chosen to be the max action right so I can say that this is equal to sorry not equivalent so the two are equal that makes it easier.

(Refer Slide Time: 12:24)

(c) Suppose (3) is met for some n , & π satisfies R_n . Then:

$$\|v^\pi - v^*\| \leq \|v^\pi - v^{n+1}\| + \|v^{n+1} - v^*\|$$
$$L_{v^{n+1}} = E[v^{n+1} | s, \pi(s)] \leq \sum_{s'} p(s' | s, \pi(s)) v^{n+1}(s')$$
$$L_{v^{n+1}} = L_{\pi} v^{n+1}$$

Now we are going to start simplifying stuff.

(Refer Slide Time: 12:27)

$$L_{\pi} v^{n+1} = E[v | \mathcal{F}_{\pi}(s)] \cdot v \leq \rho(s | \mathcal{F}_{\pi}(s)) v^{n+1}(s)$$

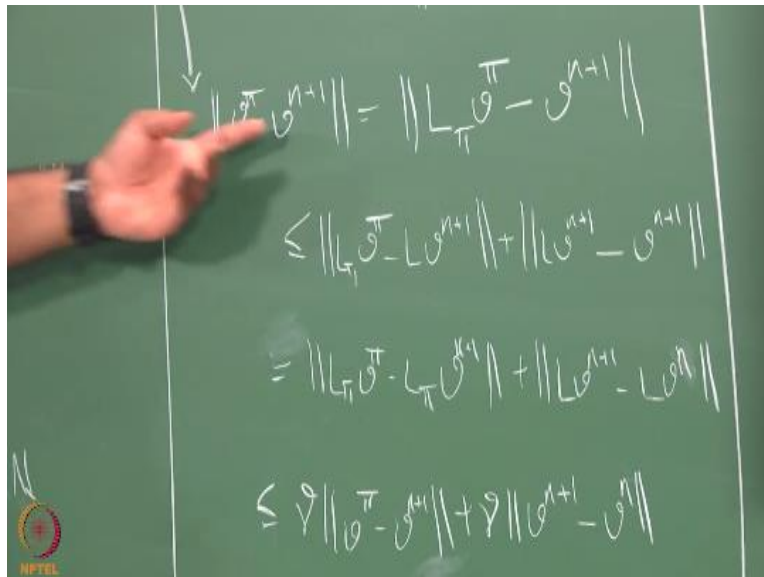
$$L v^{n+1} = L_{\pi} v^{n+1}$$

$$\|v^{n+1} - v^{n+1, \pi}\| = \|L_{\pi} v - v^{n+1}\|$$

$v^{n+1} = L v^n$

So I am going to take this expression I am going to try and simplify that okay we can I can replace $v \pi$ with $L \pi v \pi$ dot $y \pi$ is a fixed point nothing will change I can just replace it to the LP $v \pi$ now so can I do this same application of the triangle inequality as we had before so three points are $L \pi v / v^{n+1}$ and $L v^{n+1}$ so $L v^{n+1}$ is just another point in space between three points I can write this triangle inequality so this is less than or equal to H this is equal to right so this is the reduction we talked about here so from $L v$ I went to $L \pi v^{n+1}$ I cent $L \pi v^{n+1}$ and from v^{n+1} I went to $L v^{n+1}$ because that is how v^{n+1} was generated by applying L on v^n okay so I have written that.

(Refer Slide Time: 15:06)



A hand is pointing to the first equation on a chalkboard. The equations are:

$$\begin{aligned} \|\mathcal{L}_T^{\pi^{n+1}}\| &= \|\mathcal{L}_T^{\pi} - \mathcal{L}_T^{\pi^{n+1}}\| \\ &\leq \|\mathcal{L}_T^{\pi} - \mathcal{L}_T^{\pi^{n+1}}\| + \|\mathcal{L}_T^{\pi^{n+1}} - \mathcal{L}_T^{\pi}\| \\ &= \|\mathcal{L}_T^{\pi} - \mathcal{L}_T^{\pi^{n+1}}\| + \|\mathcal{L}_T^{\pi^{n+1}} - \mathcal{L}_T^{\pi}\| \\ &\leq \gamma \|\mathcal{L}_T^{\pi} - \mathcal{L}_T^{\pi^{n+1}}\| + \gamma \|\mathcal{L}_T^{\pi^{n+1}} - \mathcal{L}_T^{\pi}\| \end{aligned}$$

Now I can simplify this so I have $\epsilon \pi^{n+1}$ here $v \pi - v \pi^{n+1}$ here sorry and $v \pi - v \pi^{n+1}$ here so I can take it to that side simplify things so I will just get okay that is just the first term now we have to simplify the second term towards the second term.

(Refer Slide Time: 16:07)



I just like we did earlier I have done a repeated application of the triangle inequality is only tricky thing here is a summation goes to infinity because only when I repeat this till infinity will go to B^* right but that is fine because this is anyway it is a vanishing quantity as we go along right so this of this sum will actually okay exist limit of a summation yeah well I am you can think of this as $v^* - v_{N+1}$ right because I am just talking about the norm right so as $v^* - v_{N+1}$ and then I have written this up okay.

So this is v_{N+1} that is v_{N+2} then v_{N+2} v_{N+3} blah this one so forth so what will this be you only said it to be a limit of a summation right guys I need to take $1 - \gamma$ out okay so what do we have so we have the first component is less than equal to $\gamma \times 1 - \gamma \times v_{N+1} - v_N$ right but when three holes what is $v_{N+1} - v_N \leq \epsilon \times (1 - \gamma/2)$ right so if I plug that then what will I get $\epsilon/2$ to plug it in there what will I get $\epsilon/2$ so I have few components here so $v_N - v^*$ is less than or equal to $\epsilon \times 2 + \epsilon \times 2$ which is ϵ right $v_N - v^*$ is less than or equal to okay when we have actually to complete the proof we have to do that but I am asking you to do that what was the other thing that we wanted to show D or later right.

So D essentially was to show that V_{N+1} will be at least ϵ to close so we already done so we substitute that here so we get $\epsilon/2$ so is that so $V_N - p.m.$ sorry $V_{N+1} - v^*$ is less than or equal to $\epsilon/2$ and what we are showing there is $v_{\pi} - v^*$ is less than or equal to ϵ like so that is basically done so we have shown this theorem so essentially what we have here is that value iteration converges with that stopping criterion it converges to a ϵ optimal policy so you have to pick an ϵ first remember γ comes as part of the problem definition.

So once you pick an ϵ you can find a stopping criterion using that expression there and you are all set okay so there is one other very interesting thing that people talk about from a theoretical analysis of algorithms for solving these kinds of MDPs right so it is called the rate of convergence so I have shown you that it converges so there is a whole concept called rate of convergence right how quickly does it approached right so I am not going to get into a huge discussions on rate of convergence and like I mentioned with the deterministic policy thing I actually put up a small right upon rate of convergence as well you can read it.

One thing which I want to tell you is that so you can with a little bit of not much right so L is a contraction mapping L is a contraction so at every iteration right so my so what is the contraction factor for L we did this γ right so for every iteration I can show that my iterates that is basically I start off with v_0 then I become take me one then V_2 V_3 and so on so forth every iteration I can so that the successive iterates will become γ closer to v^* right.

So I have some so v_1 was some x close to v^* then v_2 will be γX closer to so γ is small it will converge faster if γ is very large it will converse lower right so this is called linear convergence with the rate of γ right so if it falls as γ^2 then it will be quadratic convergence right so but linear convergence with the rate of γ so this is essentially the thing so the interesting aspects I mean if you want to get into the more into the theory of all of these things is to start studying the rates of convergence of different algorithms and so on so forth which I am not going to get into but there is something for as an aside so that you can look at it.

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in

Copyrights Reserved