NPTEL

NPTEL ONLINE COURSE

REINFORCEMENT LEARNING

Cauchy Sequence and Green's Equation

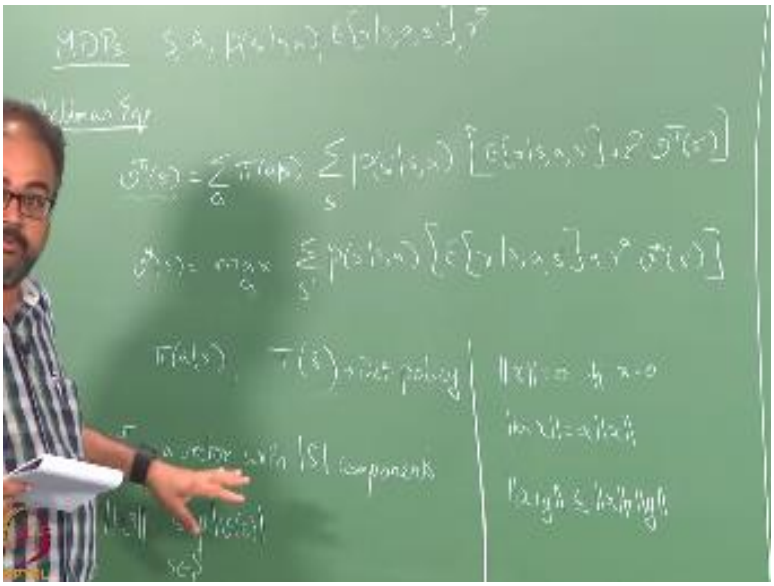
Prof. Balaraman Ravindran

Department of Computer Science and Engineering

Indian Institute of Technology Madras

So where were we in the last class we are looking at looking at.

(Refer Slide Time: 00:21)



Bellman I mean MDP s and you are taking it the bellman equation right so the MDP is are rectified by s a right so this is what we use for defining in MDP right and what is the bellman equation let us stick with the bellman equation for B for the simplicity sake tell me timber okay

likewise we had another optimality equation right so we have the optimality equations one thing which I want you to note about the optimality equation is that we are making an assumption that so the optimal policy would have the same max here right.

So that is fine correct another assumption that you will make as we go long trying to solve the MDP is that the optimal policy is deterministic right so and we will also use this following notation instead of saying π (a) given is I will use the notation π (s), so π (s) given is office is essentially a single action a that we will always do whenever you come to status right so by of a given s is a probability distribution right this is essentially saying that if my π is says that it is one for one action and 0for all the rest for every state right.

Then I can represent it as this is where I will return the action for which the probability is 1 okay so this is a deterministic policy so this is a notation that you use so this is the note sir right and later on well we will see that will make an assumption that the deterministic policy I mean the optimal policy is deterministic right, in fact more correctly you will assume that if the MDP has an optimal policy then there exists a deterministic optimal policy right there could be multiple stochastic optimal policies but that will exist at least one deterministic optimal policy.

So that is assumption that will make so later so in fact I will see if I can actually develop all the machinery required to prove that but we would not have to go do that right so we will see if I do not get to that then please take it on faith okay so a couple of things so now I am going to talk about so the existence and the uniqueness of solutions for these sets of equations right going to talk about the existence and uniqueness of solutions for this set of equations so for us to get into that we need a little bit of analysis right.

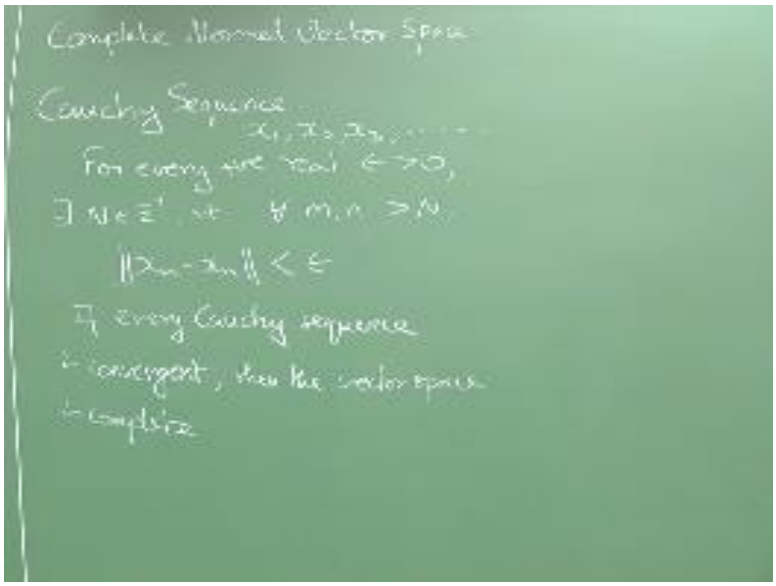
So we need to know a little bit of fun days from functional analysis and so I will I'm going to keep it very very light okay I am not worried I am not going to do anything very deep in this but for that people have to recall all their things about vector spaces and so on so forth remember vector spaces right so vector spaces you have a set of elements are closed under addition and scalar multiplication right, so we are going to consider at some point welcome to that so one way of thinking about the value functions right.

In a finite MDP you can think about it in a non-final a discrete state space also but for the simplicity sake let us assume that what I am going to talk about now is only about finite MDP so what are finite MDP these are MVP where s is finite right and is finite both s and a are finite and we already know that the expectation should be bounded right so these are the three conditions that we need is a finite MDP then s is finite a is finite and this expectation is bounded so if you think of a finite MDP I can think of my value function as a vector in a size of s dimensional vector space right.

So I can think of it as a vector which has s components in fact so it is it is a space on which I can define a notion of a norm that normally something like a size of the you can think of at roughly as equal to being size of the vector that I can define a norm on this vector space right but I am interested in using a specific kind of norm on this space which we will call the maximum right so people know norms so you can think of it as this kind of some kind of a measure of size right so let us put this down here okay, so what if and only if x is 0 norm x is 0 right and Alpha X is alpha where α is a scalar is alpha times norm of X and this is the triangle inequality okay great we have this now I am just seeing what notations I should introduce for you.

So that is fine so people know all about vector spaces okay so in vector space where you have a norm defined on the vectors is called that non vector space ok so all of you know that great.

(Refer Slide Time: 09:04)



So now I want to talk about something called a complete normed vector space we all know about water normed vector spaces so when do you say it is complete okay, so for that we need to understand something called was it not Cauchy-Schwarz inequality really need to understand what is called a Cauchy sequence okay, so a sequence of course is something like this okay right so what does Cauchy sequence it says that okay.

For any real epsilon that you give me I can find some point in the sequence ok beyond which if you pick any two elements they are within epsilon of each other selective says for any epsilon you give me right so what does this mean so this sequence should be such that the successive elements are coming closer and closer to each other right does it make sense is a sequence in which the successive elements are coming closer and closer to each other.

So if every Cauchy sequence in a normed vector space converges to a point in the vector space right then you call the vector space as a complete space X_1 X_2 the rescission x_1 and x_2 will be greater than X_2 and X_3 yeah so x_1 x_2 the distance will be greater than explain extreme possibly it should be can you further from the definition of a Cauchy sequence the Cauchy sequence just

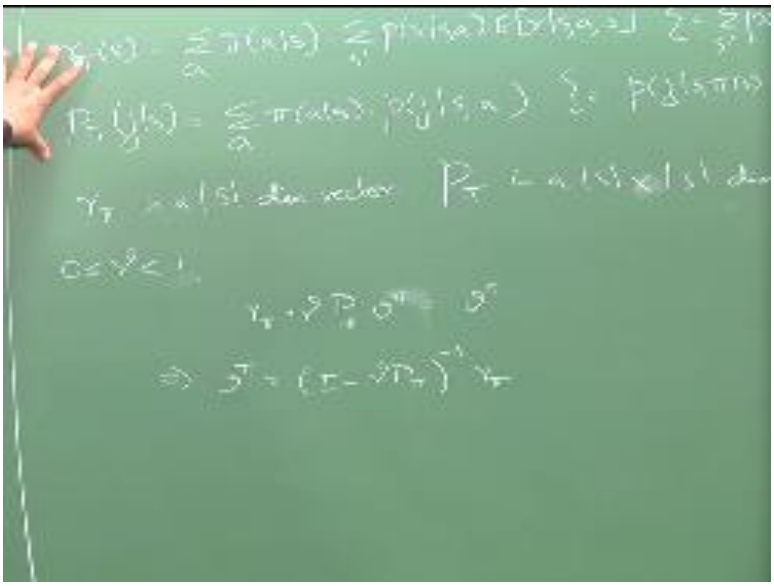
says the successive difference are getting smaller and smaller and smaller does not necessarily say there is a limit right.

Does not necessarily say there is a limit it just says it becomes smaller and smaller and smaller and smaller so you could have a limit right that is outside the space that you are considering it right so if you want to look at examples of Cauchy sequences that are not convergent go look at Wikipedia right so there are enough material out there that talk about all of these things right so you could have them becomes smaller and smaller and smaller right but the point to which they converge might be a point which does not belong to the space that you originally are defining right.

So these kinds of things would happen because when we are dealing with infinitely long sequence in all kinds of crazy things happen okay I should have noted down one such example anyone can recall an example of the top of their head non convergent Cauchy sequence something is like 1 by n and it converges to 0 but then we can define of space which excludes 0 right can you can you define a vector space at exclude 0no then they have to come up with some other example say something that converges to some other value 1 -1/x.

We converse to and now define a space that excludes Monday so but that people get the idea right so Cauchy sequences might not be convergent in the in the domain that you are operating in right so the if every Cauchy sequence okay good so I think that is all we need from that so when introduced a few more vector notation.

(Refer Slide Time: 14:45)



So that we can translate all these bellman equations into in to some nice underlying vector space and then show some results there okay so we already have VS a vector. So I am going to depend r π of s essentially I am saying that what is the reward I will get in-state case for following policy pipe and what is expected reward in state tests for following policy pie it makes life a little simple as you can see some of this thing here I am kind of compressing and writing it as a single vector essentially the first term here right.

I have compressed it and I am writing it as ARP is okay which is the reward I expect to get starting in state s and following policy π just like the immediate one-step probe all right it is not it is not expected reward for return we need the v π γ v π just the one-step reward that I will get starting from is following policy π, so likewise I am going to write so this gives me the one step transition probability under policy π right I fixed policy pie right so this will give me the one step transition probability under policy file so what is the probability.

I will end up in state Jif I start from state tests and take one action according to policy π and there could be many ways in which I can get to state J from status S right these are the two things that we needed and so just as an aside if I am using a deterministic thing this becomes a little easier

so this will essentially be the deterministic policy so at some points it makes it easier for us to explicitly consider deterministic policies so I am just writing out the notation here so whenever I want to used deterministic policies I might actually switch to this notation the summation over a disappears right.

So likewise here this will be just I because only one action I will take under policy pie right therefore the probability of going to jay from yes is a probability of going to gay from is under that action π of s just a notational convenience okay so do not do not work cookers of too much for it so one thing to notice this Rπ is ayes dimensional vector again so our π sorry Rπ is dimensional vector and pπ is a s-cross s dimensional matrix it is also a special kind of metric is a stochastic matrix.

So what is the stochastic matrix those are up to one and anything else all the negativism on-oh no hit an equal to zero okay so all the all the rows have greater than equal to 0 essentially each row you can think of it as a discrete the probability distribution right they are the entries in a row sum up to 1 and they all have to be non-negative so basically they lie between 0 and 1 then so the normal essential we are going to make is that γ lies between 0 included and one okay.

This assumption we will make and if you think about it that one step returned it I can write it as tourism I will come to that in a minute right so what do you think this is so if you think of an MDP right where I am making one decision right so the decision is going to cost me to make a transition to some state right I will starting from some sadly some yes okay making a decision it is going to cause a transition to a state J right there will be a reward along the way and I also have a terminal cost or a terminal reward for landing in j.

That terminal reward is given by the function V so if you think of it this way so my entire problem consists of taking only one step right I pick a step I take an action I perform it I get my choose an action we perform the action I am going to make one transition and that is it Island right in what will happen is for that one transition I will get some reward right and for ending in that particular state J I am going to get some payoff for ending in j right and that pair of literacy is given by V then this is the total reward I am going to get right.

So r is the vote for the one transition and v is the final cost right and where I am going to land up is determined by pπ, now you can think of it as a one-step transition function where the terminal costs is given by V, so what we really want to do is instead of assuming that there is an arbitrary function V that is giving us this terminal cost we want to find a function v π right which in some sense substitutes for the terminal cost of following policy π.

Right so we want a v π that satisfies this equation so you can see why v π is a terminal cost side so it is like what is v π so value I am going to get expected return and going to get starting from the next state and following I mean starting from the state and following all this impact right, so one way of thinking about it is my decision making problem ended with making one decision then I made the choice according to π I stopped there and whatever in the future regardless I am going to consider that as one value and then add it to main function so that is be fine so I am thinking take in treating v π as a terminal cost for a one-step decision problem and then what I really want Leslie prior to satisfy it should be the total cost right.

Now starting from here so this is essentially the equation that I want to solve for v π so we'll come to we solve later this is the equation I want to solve for B π X so people are all onboard with this right and so you can substitute that for R π and this for p π and you write it out you see that it is essentially this expression like I started from this expression and simplified it there but I wanted you to look at the other way of interpreting this expression that is why I started with this terminal cost interpretation okay.

So if the terminal cost interpretation does not appeal to you can do algebra start from this expression which way wish I hopefully convinced you in the last class is correct right simplified you will get that expression okay great so this implies that okay, as simple as if I just took this to other side and took multiply it by the inverse on both sides I get this so what can you say about this matrix is it invertible will remember determinants anyway fine.

So well you know what the Eigen values are IR all ones right and what about p π it is so caste x so the largest Eigen value for a stochastic matrix is 1 okay, so the largest Eigen value for a

stochastic matrix is 1 and I multiplied gamma so the largest Eigen value is going to be less than 1 so the whole matrix that cannot have a zero Eigen value that for this thing will be invertible it in there for there exists a unique solution for V π was it well.

So for stochastic matrices right the Eigen values the largest Eigen value for a stochastic matrix is 1 this is why I said this is a particularly mentioned that this thing is a stochastic matrix right because the rows all sum to 1 you can show with little bit of work that the largest Eigen value for stochastic matrix is 1 okay so I am multiplying it by γ which is less than one right so this whole thing is going to be going to have an Eigen value largest Eigen value which is less than 1 right and the Eigen values for I are all one right so if you take the difference matrix and try to find the Eigen values for those right it cannot have an Eigen value of zero right as long as the matrix does not have an Eigen value of zero it is invertible.

This is basic linear algebra properties I right so σ table therefore this is unique there is a unique solution to this and we are all set correct so this is sometimes known as the green equation but not always because there is something else also called the greens equations I do not want to write it down that so this is one way of showing that p π is unique right sorry r is equal to dimension for a unique solution yeah, that must be equal to dimension for it to be invertible okay right this is one way of convincing people that this works right so another way of convincing people is to look at the whole dynamics of what happens right in that equation right.

So think about something like this right I start with some arbitrary function for v π I start with an arbitrary guess for v π okay, so I plug this in here right and I am going to solve this equation not solve there is going to run this computation so I am going to get another expression for I mean other value for v π let us start with some arbitrary gates let us call it be not okay I will compute this I will get say v_1 right and then I will go back and plug v_1 back into this expression okay I will compute this I will get V_2 and I keep doing this iteratively right I am going to claim that this will converge to V π.

Right we are going to make several claims I am going to say I made several claims in that statement first claim was is going to converge so you can claim was third claim is it is unique

right so we already know it is unique from this expression this thing this thing we know it is unique but I can also try to show it in a different way right so I am going to show that it is going to converge right and it will converge to be pi and that v π is unique right so then I am going to do it this way is it makes it easier for us to handle the optimality equation as well later right.

So if I show you how to do this in the v pi case it makes it easier for the stand in a optimality equation later great so people have heard me before know that I am going dead slow today I would have normally finished everything I wanted to do by now but I will probably go to a second class because I wanted to make sure that people get what is what I am saying right so let us see so if you have any questions or anything are you are completely lost how many of you are completely lost that us basic basically energy world basic linear so go do me a favor go Google up relationship of Eigen values to invert ability of matrices they probably get a stack over flocked okay.

I do no stack state x change it stack exchange is here no that is a stat exchange right yeah you would probably get a stat exchange hit on this yeah anyway so that is very basics term right so you know about determinants and products of Eigen values being equal to determinant right so a matrix is not invertible only if the determinant is 0 right and if know the Eigen values 0 then the determinant cannot be 0 therefore the matrix will be invertible that convince you maybe a bit more okay.

Anyway right let us go look up so I am not going to do basically in algebra now so let us now take you kicking and screaming into something more I need to erase this part which part can a race okay so I probably end up erasing this part so everyone has copied down what this our pie and R π and P π rare right, so everyone has it in the notebook so I can erase this because I am going to use it but I want that expression also to stay I do not want a race that so everyone has this great so I am going to raise it.

IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development

Government of India

www.nptel.ac.in