INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

REINFORCEMENT LEARNING

Bellman Equation

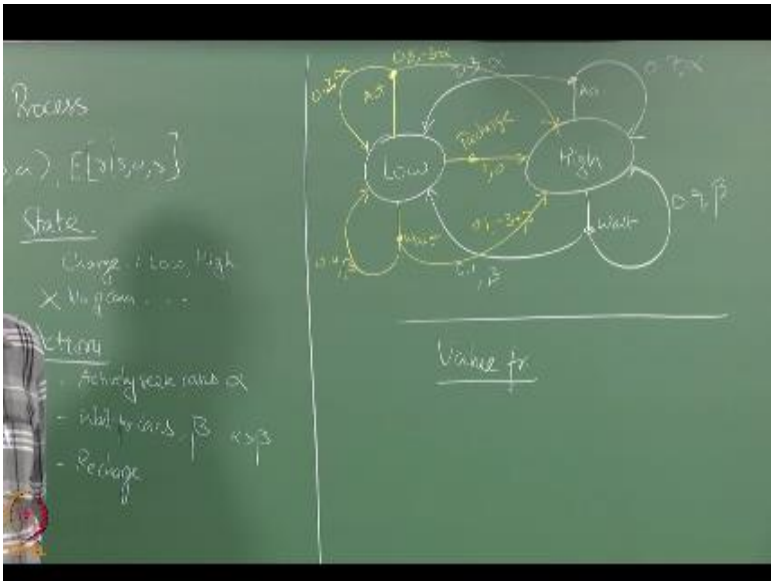
with

Prof.Balaraman Ravindran

Department of Computer Science and Engineering

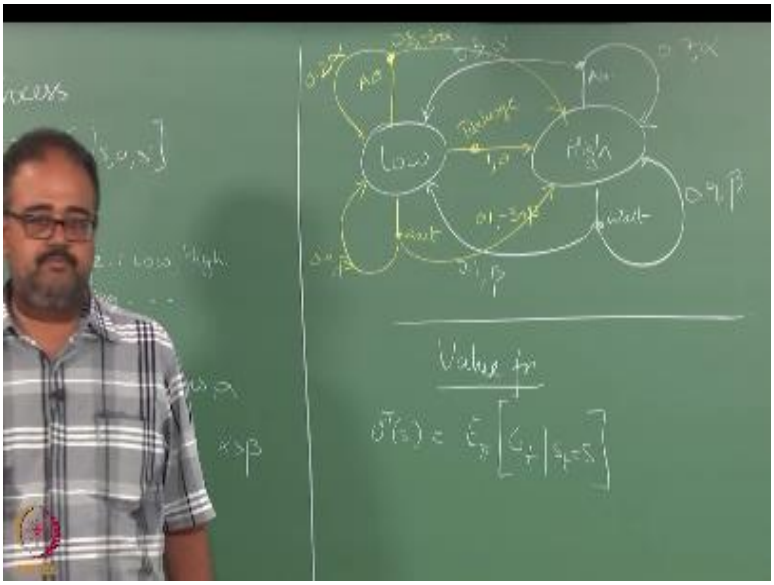
Indian Initiation of technology, Madras

(Refer Slide Time: 00.15)



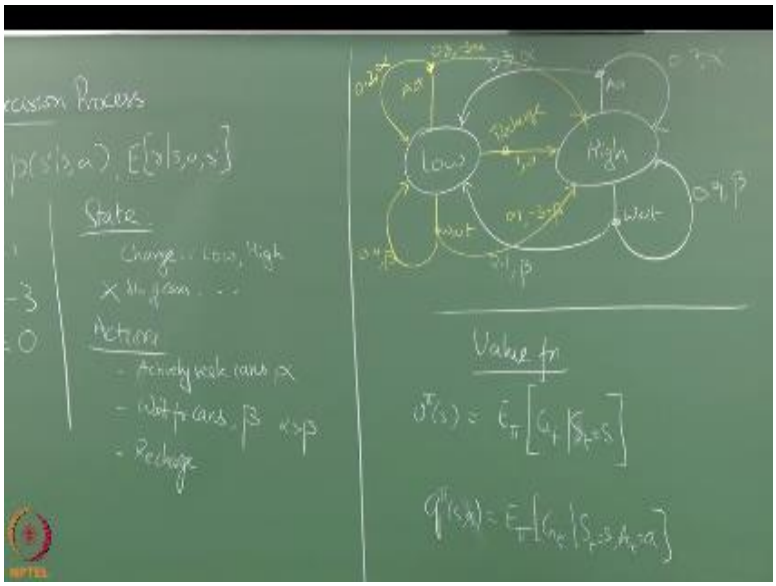
So we are talking about value functions okay people remember value functions is we define value functions we said

(Refer Slide Time:00.39)



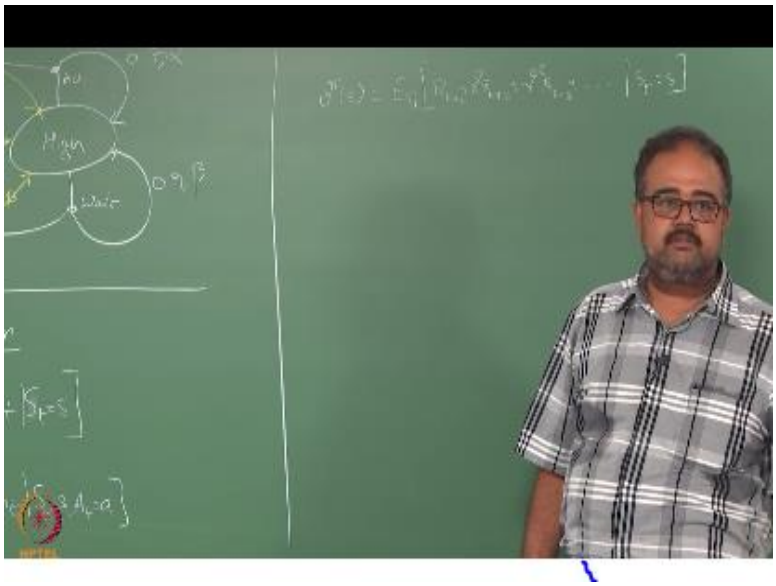
V^{Π} (S)= to expected value under Π of right.

(Refer Slide Time:00.46)



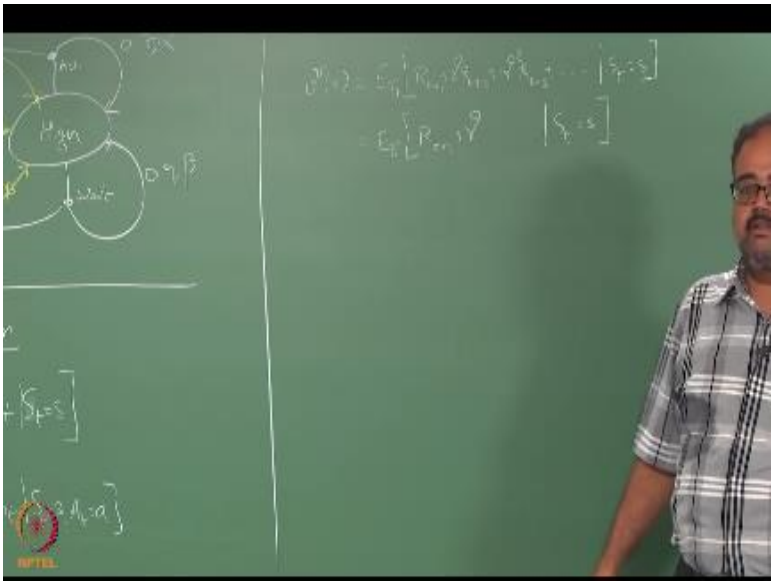
And then we defined Q^{JI} .Right good, so now we are going to start unrolling their expectation and so people remember this expression right so all of your player but I don't have to explain what is happening there again right so V^{JI} (s) is means starting in status following policy Л . Thereafter and taking the expectations of the, the returns that I am getting I am going to start unrolling this right so what does it what does it look like so

(Refer Slide Time:1.48)



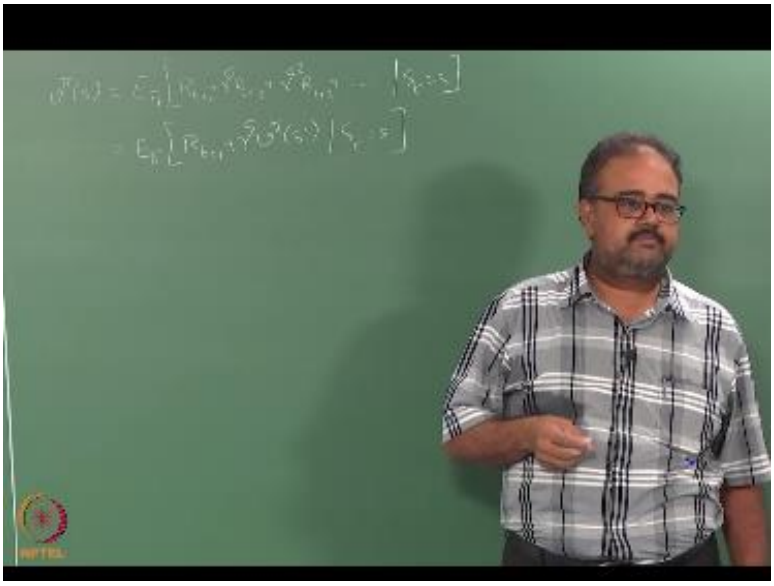
i am going to say V^{JI} (s) equal to, yes

(Refer Slide Time:2.31)



What can right here, I GT+1 right but what is what would be the value of GT+ for the expected value of GT+1.V^{JI} off whatever state you land up in right so let us go I am going to say that this is equal to

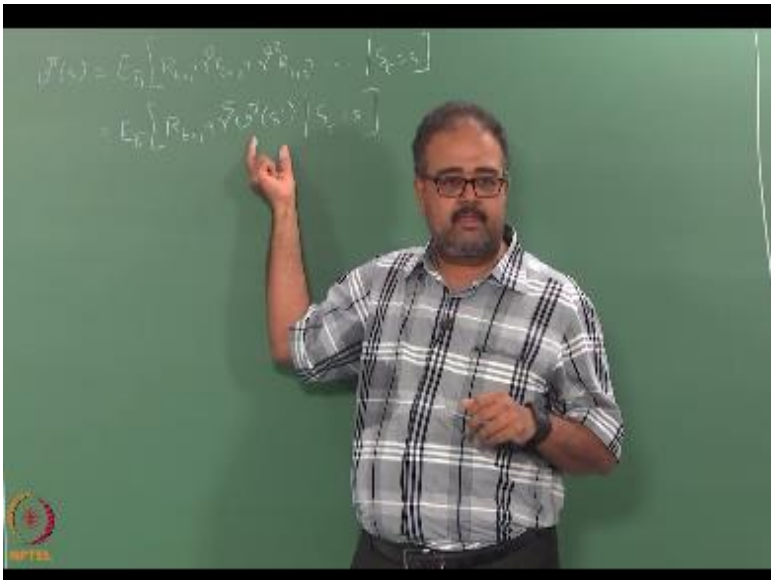
(Refer Slide Time:2.59)



VJI (s) prime were s prime is the state I am going to add up and we will quantify as Prime in a minute right so, so people agree with me on this right is s prime is the state I go to from s right so I can take this summation and write it like this. Everybody on board okay great so now I need a slight slight of hand here right so I took GT and then I wrote it as one term plus the expected value of GT but that is fine because I am anyway going to take a expectation eventually right so this is slight reordering of expectations right instead of taking an expectation of the full expression.

I took the expectation of the inner one alone but one thing you have to remember is it is only a partial expectation because s prime can change right so for me to get the full expectation i need to condition on s prime also

(Refer Slide Time:04.04)

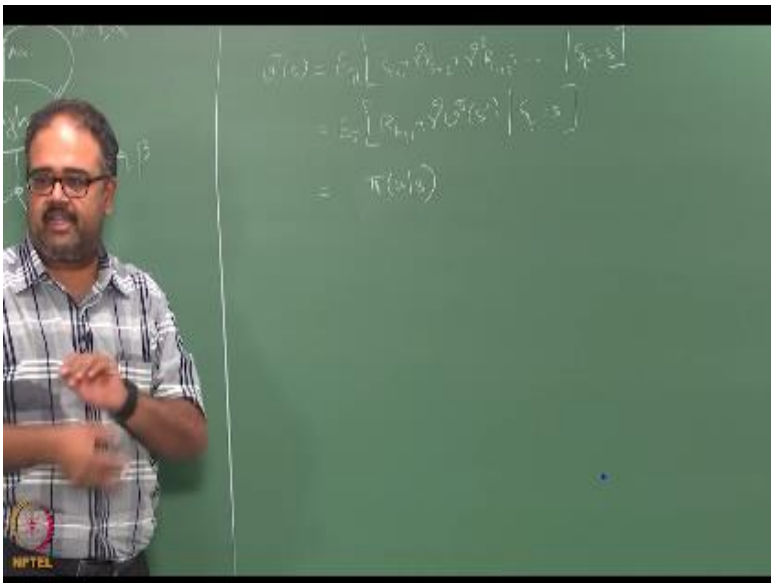


That is why this expression is still inside the expectation so if this has been the full expectation we are taking it out, out of the expectation since it is still only partially conditioned so I need to

conditioned on the full thing okay so now I have to write it out so that will get rid of the expectations okay so one way of writing out this expectation is to think of it like a generative process okay.

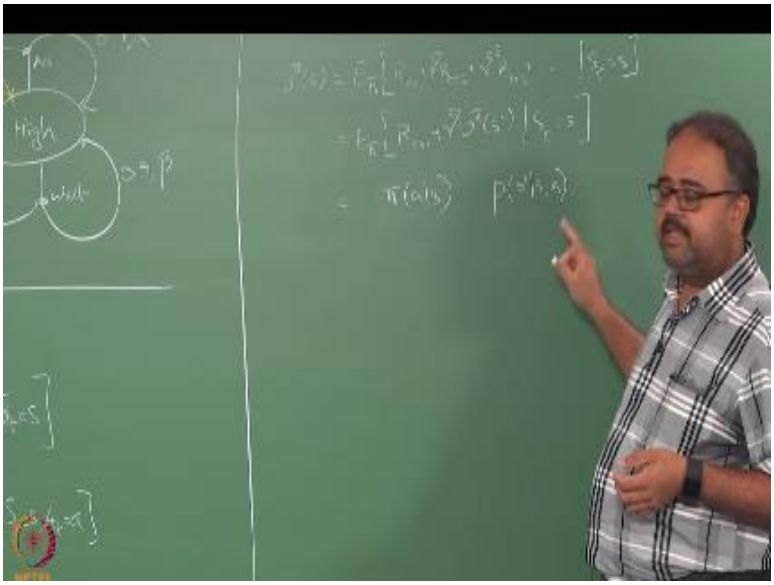
So remember what did I say we are going to do we'll start in state s right then select actions according to policy Л right and measure the, expected return right so the starting in state is and the selecting actions according to policy Л.

(Refer Slide Time:4.58)



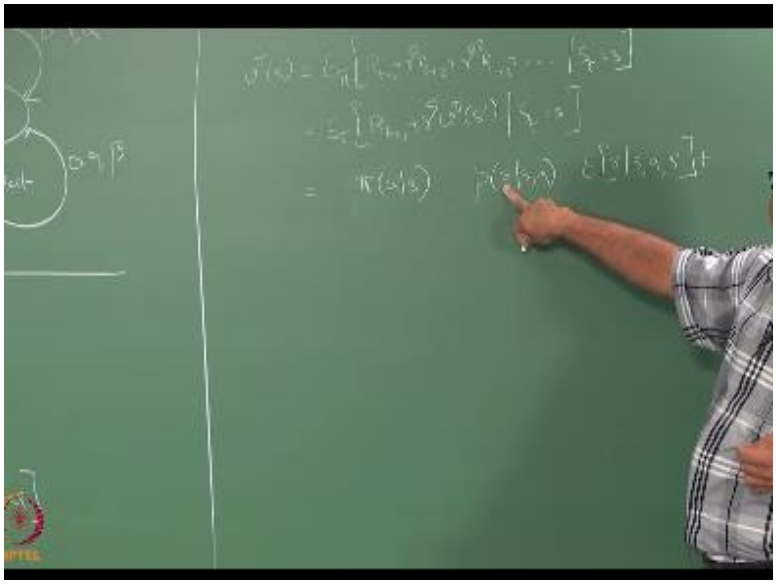
So what does that mean using Л right and I am selecting actions according to policy Л so let us say I pick some action a right so now what I am going to do I am going to get some reward corresponding to that RT+ yeah the RT+1 will be something that corresponds to that reward so but then.Before I can find out what RT +1 this I need to specify what the next state is going to be so I need this.

(Refer Slide Time :5.37)



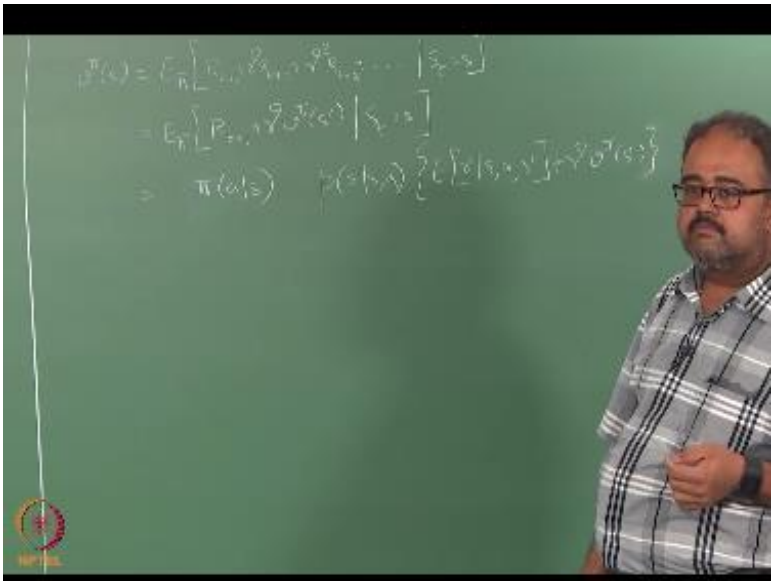
So taken action a then I determine what s prime is and then I determine

(Refer Slide Time :5.45)



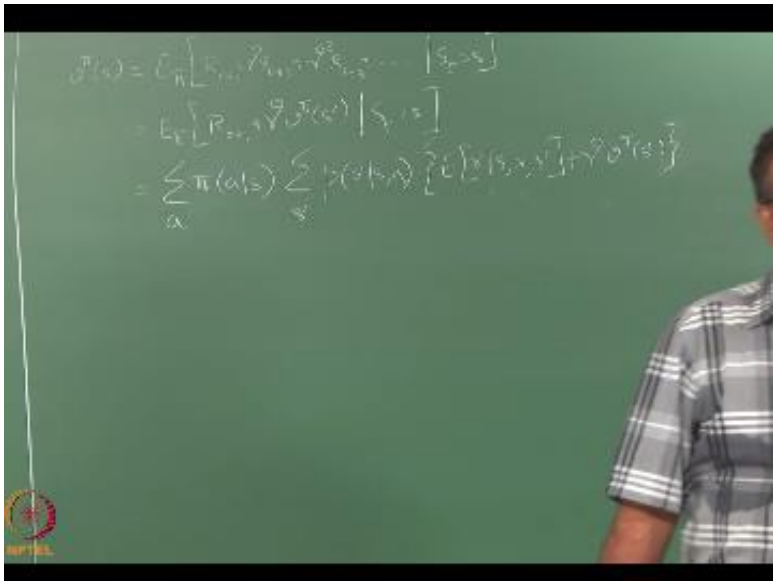
What but I already determined what s prime is it so

(Refer Slide Time:5.55)



I can do this right so this is one step I am taking a pick an action a according to Л right now that is going to cause a transition to some s Prime according to this P that we have that right and once I know what the S prime is I can figure out what the expected reward will be as well as $V^{Л}$ (s) prime because I know what is frame is so what is the probability that this will be my return is this times the probability i will pick action so this whole thing gives me the probability that this will be my expected return okay so now i have to do this overall all possible outcomes.

(Refer Slide Time :6.54)

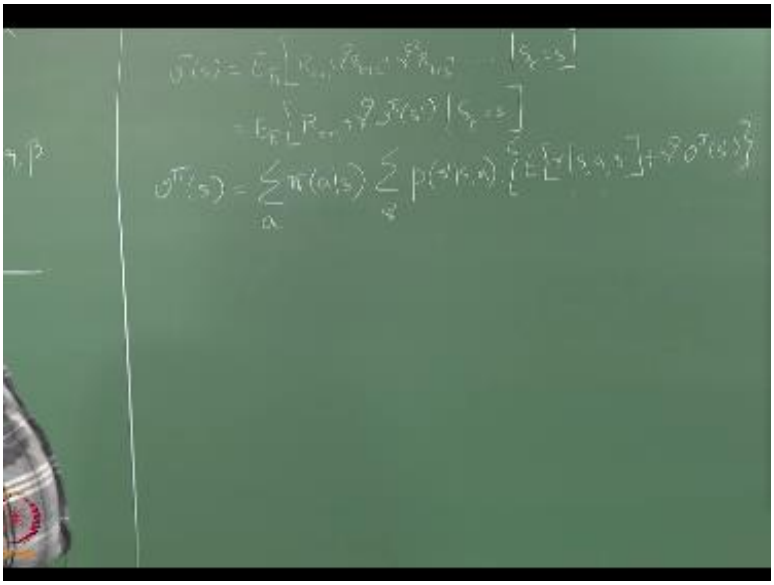


So basically a sum over spray okay is it clear how we got this right this one right so i have a s a right that i picked a a first then I pick an s Prime according to p now I have a SA S prime so what I need to do is figure out what this our expectation of RT+1 would be right so that is given in my MDP right the expectation of RT+1 would be given SA S prime.

What is the reward so that is this expectation I've written it here expectation of are given SAS prime right and then the future is γ VJI$^{(s)}$ prime so that S Prime has already been fixed so I can just add the γ beep is prime here. So the probability that this will be the outcome right is the probability of s prime given S and A time's the probability of picking A given is S .

So all these things put together gives me the probability that this will be the outcome right so that is the expected outcome for one choice of A and one choice of S prime so I'll have to sum this overall choices of S prime and all choices of A for me to get the overall expectation so that is essentially what I've done here $\sum$xI $\sum$a .

(Refer Slide Time:8.16)



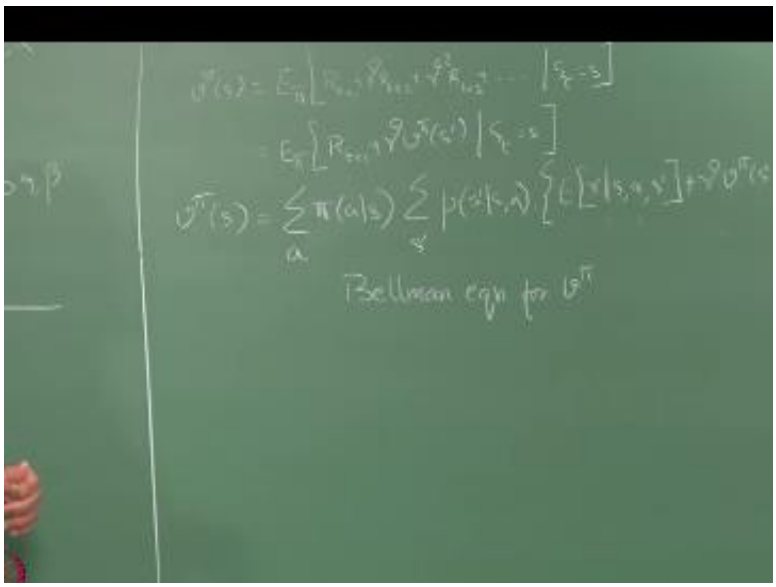
This is nice right so we written VJI in terms of VJI. So this, this system of it is a system of equations right you can see there is a system of equations that you have one equation for every, every state how many variables are there number of as many variables as there are number of states the number of states is not the variable that is also fixed as many variables as number of states and what is a very we here VJI(s) is the variable.

So essentially so in some thinking of it as a function now I can think of it as a collection of variables right one for each state right and I have N such equations and N such variables and can I solve it. Depends on what, its linear N variables in equations so when can you not solve it yeah then ,then then yeah certainly get dependent columns and so it becomes redundant so you have multiple solutions and sorts of three so it turns out as I'll show you in the next class right the system of equations have a unique solution.

Always right by virtue of the fact that the P that you are using is stochastic so the rows of that P matrix will come to one right so by virtue of that fact we can actually go back and show that there will be a unique solution for this need to write, write is slightly different in a different

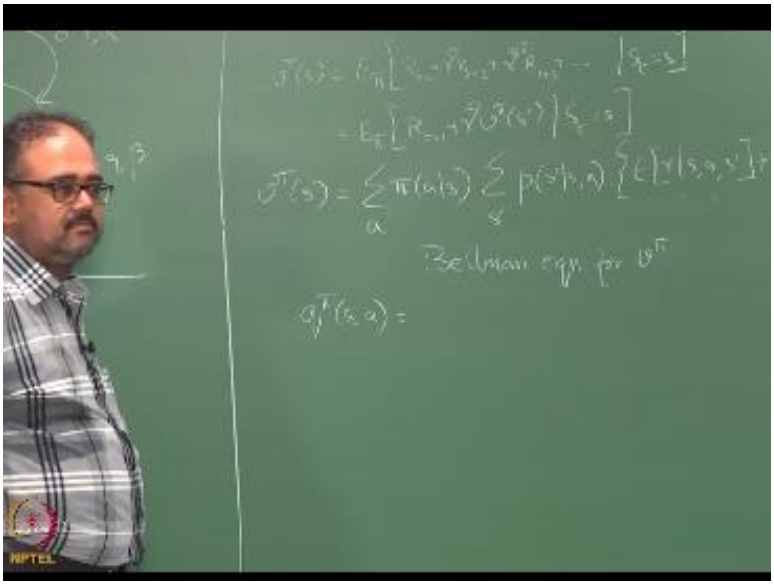
notation for us to see that so I will do that in the next class right so this is free which has a unique solution and this is sometimes known as

(Refer Slide Time:10.40)



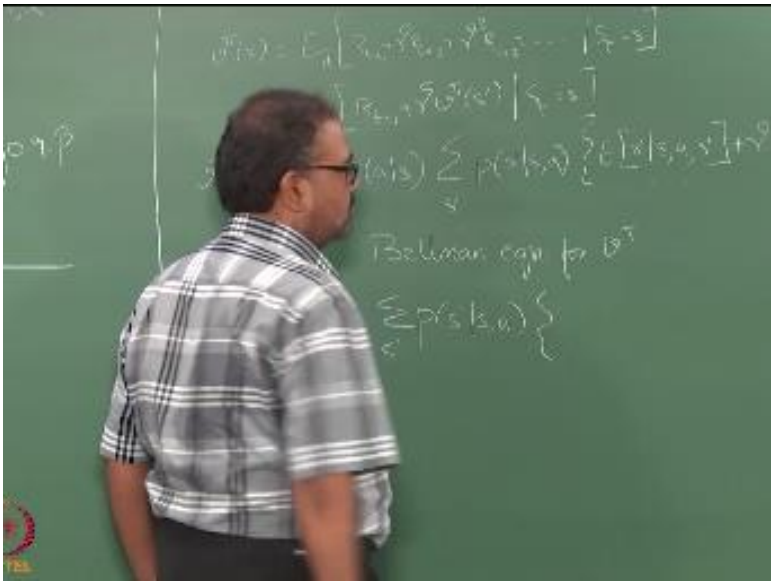
The Bellman Equation for VJI after the Richard Bellman who's one of the pioneers in the field of stochastic dynamic programming okay. Like ways you can write a bellman equation for QJI okay. Satan wants to tell me about the bellman equation for QJI would be

(Refer Slide Time:11.07)



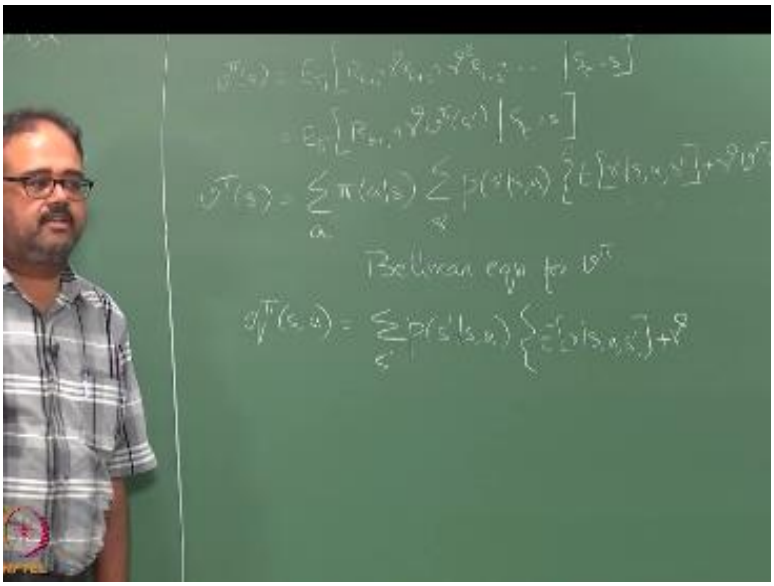
So think of it in the generative process right so what is the first thing you do when you're thinking about QJISA .They start with yes I pick a right I mean it is fixed so there is no summing over there right it's S and A is picked right so what do I have to sum over next only the Express because this Л(a/s) is out of the picture here right because I have already picked A . So I do not need to I do not need to sample from Л they already pick case so I just start my summing over s prime.

(Refer Slide Time :11.51)



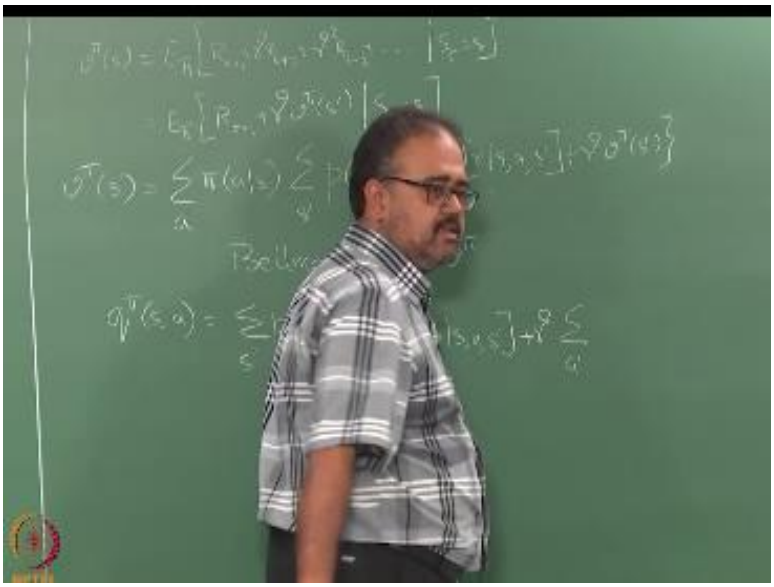
But now, now SA is prime are all fixed right so I can write expected value of R given SA s prime plus γ times.

(Refer Slide Time:12.08)



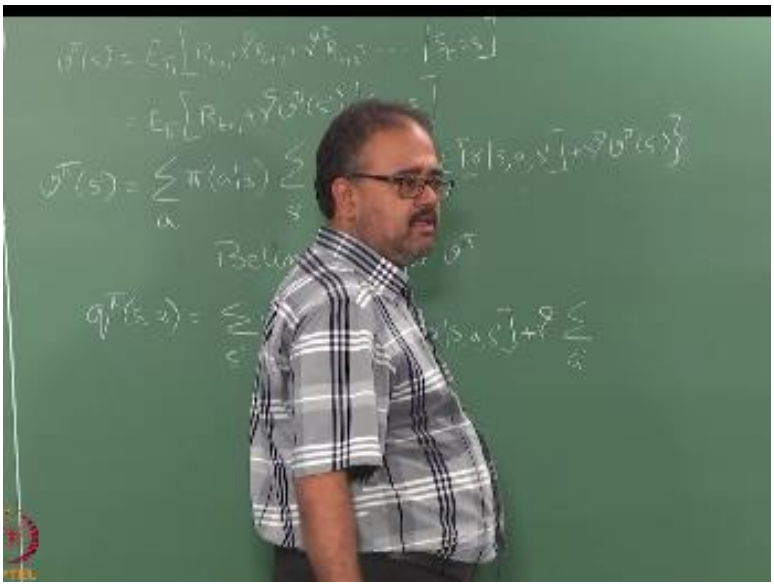
Now the interesting question is what goes there I could write VJI and it will be perfectly correct but I will not be writing QJI in terms of QJI.

(Refer Slide Time:12.25)

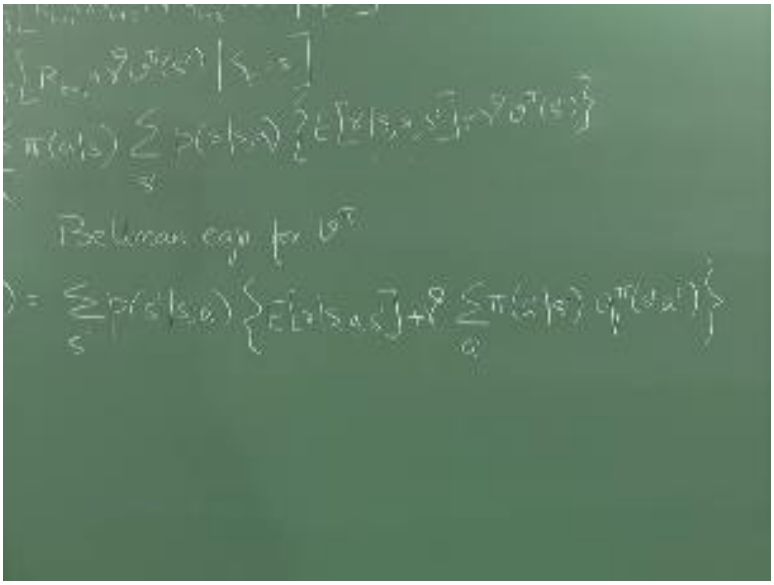


Sum over all actions

(Refer Slide Time:12.26)

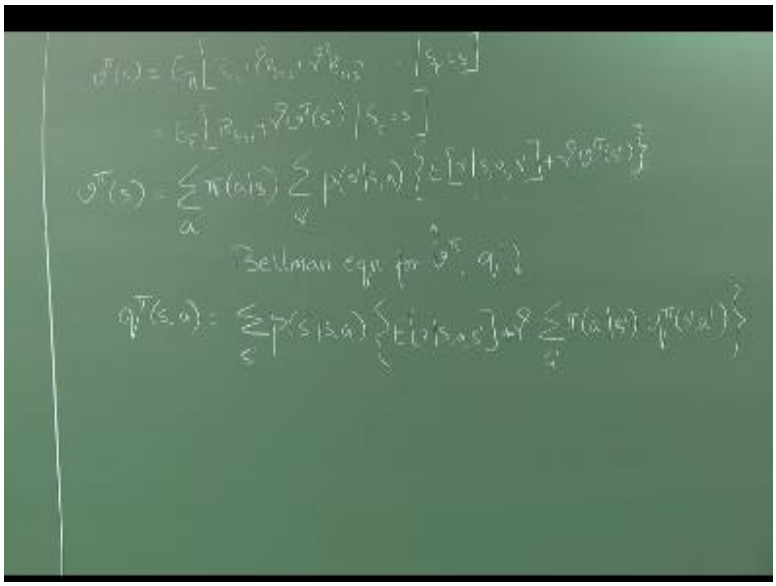


(Refer Slide Time:12.42)



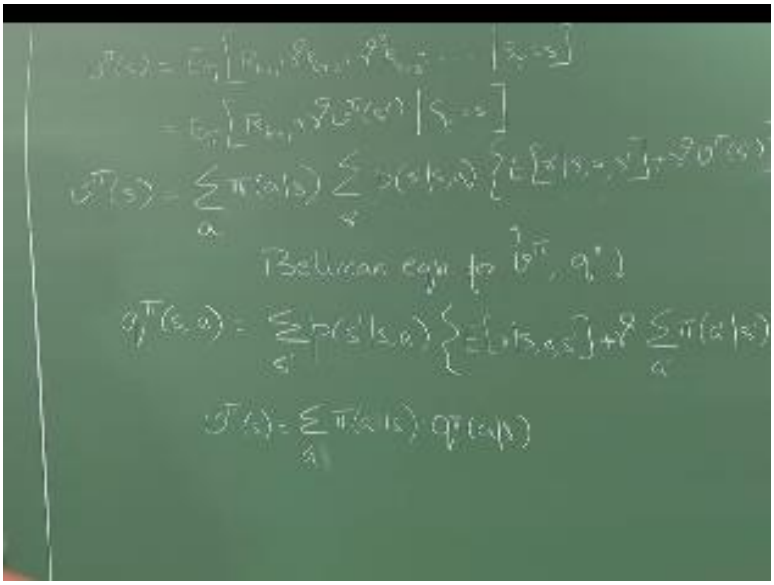
Use a Bellman Equation for QJI okay.

(Refer Slide Time:12.59)



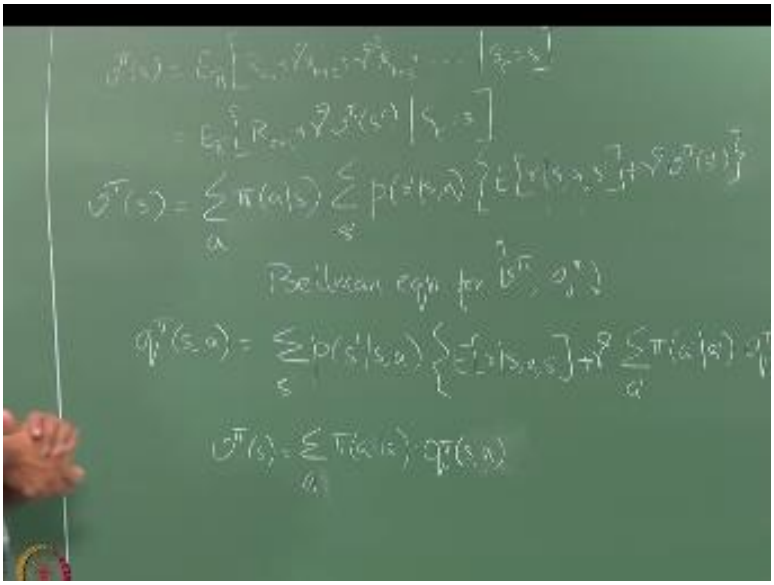
So this should also give you one, one more ancillary equation there which is.

(Refer Slide Time:13.13)



And that is a way Q and V are related and what about QJI can you write QJI in terms of VJI .But exactly what the earlier equation was the Bellman Equation substitute the last summation with VJI and that is essentially QJI in terms of VJI. Right so I cannot write QJI SE in terms of VJIs it has to be in terms of VJI of the subsequent state right but VJI is I can write in terms of QJI oops sorry.

(Refer Slide Time: 14. 17)



IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development

Government of India

www.nptel.ac.in