

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**SHAN-YUAN HO:** OK. So today's lecture is going to be on Finite-state Markov Chains. And we're going to use the matrix approach. So in last lecture, we saw that the Markov chain, we could represent it as a directed graph or as a matrix.

So the outline is we will look at this transition matrix and its powers. And then we'll want to know whether this  $p$  of  $n$  is going to converge for very, very large  $n$ . Then we will extend this to Ergodic Markov chains, Ergodic unichains, and other finite-state Markov chains.

So remember in the Markovity, these Markov chains, the effect of the past on the future is totally summarized by its state. So we want to analyze the probabilities of properties of the sequence of these states. So whatever the state you are in, all the past is totally summarized in that state. And that's the only thing that affects the future.

So an ergodic Markov chain is a Markov chain that has a single recurrent class and is aperiodic. So this chain doesn't contain any transient states. And it doesn't contain any periodicity. So an ergodic unichain is just ergodic Markov chain, but it has some transient states in it.

So the state  $x_{sub n}$  of this Markov chain at step  $n$  depends only on the past through the previous step. So for  $n$  steps, we want to be at state  $j$ . And then we have this path.  $x_{sub n}$  minus 1 is  $i$ , and so forth, up to  $x_0$ . It's just the probability from  $i$  to  $j$ , from state  $i$  to state  $j$ . So this means that we can write the joint probability of all these states that we're in, so  $x_0, x_1$ , all the way up to  $x_n$ , as a function of these transition probabilities.

So in this transition probability matrix, we can represent these transition probabilities. We see that here, in this example, this is a 6-state Markov chain. So if I want to go from, say, state 2 to state 1 in one step, it would just be  $p_{2,1}$ . If I want to go from state 6 to itself-- this is last one, which is  $p_{6,6}$ .

So this is a probably transition matrix. So if we condition on the state at time 0 and then we define this  $p_{ij}$  is equal to the probability that we're in state  $j$  at the  $n$ -th step, given that we start  $x_0$  is equal to  $i$ , let's look at what happens when  $n$  is equal to 2. So in a 2-step transition, we go from  $i$  to  $j$ . It's just the probability that at step 2,  $x_2$  is equal to  $j$ ,  $x_1$  is equal to some  $k$ , and  $x_0$  is equal to  $i$ .

So remember, we started in state  $i$ . But this has to be multiplied by probability that  $x_1$  is equal to  $k$ , given that  $x_0$  is equal to  $i$ . And we have to sum this over all the states  $k$ , in order to get the total probability from--

Oh, stand back? OK. There. OK.

So this is just probability of  $ij$  in two steps. It's just the probability of  $i$  going to  $k$  times the probability of  $k$  going to  $j$ , summed over all  $k$  states. So we notice that this term right here, the sum over  $k$  or  $\sum_k p_{ik} p_{kj}$  is just the  $ij$  term of the product of the transition matrix  $P$  with itself. So we represent this as  $P^2$ .

So we multiply the transition matrix by itself. This gives us the 2-step transition matrix of this Markov chain. So if you want to go  $i$  to  $j$ , you just look at  $ij$  element in this matrix. And that gives you the probability in two steps, going from state  $i$  to state  $j$ .

So for  $n$ , we just iterate on this for successively larger  $n$ . So for  $n$  state to get from state  $i$  to state  $j$ , we just have this probability  $x_n(j)$ , given  $x_{n-1}$  of  $k$ , given  $x_{n-2}$  of  $k$ , given  $x_{n-3}$  of  $k$ , given  $x_{n-4}$  of  $k$ , given  $x_{n-5}$  of  $k$ , given  $x_0$  is equal to  $i$ , summing over all  $k$ . So this means that we broke this up for  $n$ -th step.

In the  $n-1$  step, we visited state  $k$ . And then we multiplied that one-step transition from  $k$  to  $j$  because we want to arrive at  $j$  starting at  $i$ . But again, we have to sum over all the  $k$ 's in order to get the probability from  $i$  to  $j$  in  $n$  steps.

So  $p$  of  $n$  right here, this representation is just the transition matrix multiplied by itself  $n$  times. And this gives you the  $n$ -th step transition probabilities of this Markov chain. So computationally, what you do is you take  $p$ ,  $p$  squared,  $p$  to the fourth. If you wanted  $p$  to the 9th, you'd just take  $p$  eighth multiplied by  $p$ , to to multiply by this.

So this gives us this thing called the Chapman-Kolmogorov equations, which means that when we want to go from step  $i$  to step  $j$ , we can go to an intermediate state and then sum up all the states that would go into the intermediate state. So in this case, if the step is  $m$  plus  $n$  transition, we can break it up into  $m$  and  $n$ . So it's the probability that it goes from  $i$  to  $k$  in exactly  $m$  steps and  $k$  to  $j$  in  $n$  steps, summing over all the  $k$ 's that it visits on its way from  $i$  to  $j$ . So this is very useful a quantity that we can manipulate our transition probabilities when we get higher orders of  $n$ .

So the convergence of  $p$  to the  $n$ . So a very important question we like to ask is as  $n$  goes to infinity whether this goes to a limit or not. In other words, does the initial state matter, all initial states matter in this Markov chain? So the Markov chain is going to go on for a long, long, long, long, long time. And at the  $n$ -th state where  $n$  is very large, is it going to depend on  $i$ ? Or is it going to depend on  $n$ , which is the number of steps?

If it goes to this quantity, some limit, then it won't depend on this. So let's assume that this limit exists. If this limit does exist, we can take the sum of this limit and then multiply it by  $p$  of  $jk$ , summing over all  $j$ . So we do a sum of over  $j$ . So we're going from  $j$  to  $k$  on both sides, and we sum over all  $j$ .

So we take this limit right here. We notice that this left side going from  $i$  to  $k$  to  $n$  plus 1 is just that this limit at state  $k$  exists. Because we saw assumed up here that this exists for all  $i$  and all  $j$ . So therefore, if we take the  $n$  plus 1 step, we take this  $n$  going to infinity of  $i$  to  $k$ , it has to go to  $p_i$  of  $k$ .

So when we do this, we could simplify this equation up here. And if it doesn't exist, we have this  $p_i$  sub  $k$  for all the states in the Markov chain. So this is just a vector.

So  $\pi_k$  is equal to  $\sum_j \pi_j P_{jk}$ , summed over all  $j$ . So if you have an  $m$  state Markov chain, you have exactly  $m$  of these equations. And this one, we'll call it the vector  $\pi$ , which consists of each element of this equation, if the limit is going to exist. But we don't know whether it does or not, at this point in time.

So if it does exist, what's going to happen? So that means I'm going to multiply this probability matrix,  $P$  times  $P, P, P, P, P, P, P$  all the way. And if the limit exists, then that means for each row, they must be all identical. Because we said the limit exists, then going from 1 to  $j$ , 2 to  $j$ , 3 to  $j$ , 4 to  $j$ , they should be all exactly the same.

This is also the equivalent of saying, when I look at this large limit, as  $n$  is very, very large if the limit exists, that all the elements in the column should be exactly the same as well, or all the rows. So the elements are equal to each other, or all the rows, if I look at the row, which is going to be this  $\pi$  vector. They should be the same.

So we define this vector. If this limit exists, the probability vector is this vector  $\pi$ . Because we said it was an  $m$  state Markov chain. Each  $\pi_i$  is non-negative, and they obviously have to sum up to 1. So this is what we call a probability vector, called the steady-state vector, for this transition matrix  $P$ , if it exists.

So what happens is this limit is easy to study. In the future in the course, we will study these  $\pi P$ , this steady-state vector for various Markov chains. And so you see, it is quite interesting, many things that could come about it. So we notice that this solution can contain more than one. It may not be unique.

So if it contains more than one, it's very possible that it has more than one solution, more than one probability vector solution. But just because a solution exists to that, it doesn't mean that this limit exists. So we have prove that limit exists, first.

So for ergodic Markov chain, here we have another way to express this that this matrix converges is that the matrix of the rows-- the elements in the column are all

the same for each  $i$ . So we have this theorem. And today's lecture is going to be completely this theorem.

This theorem says that if you have an ergodic finite-state Markov chain-- so when we say "ergodic," remember it means that there's only one class, every single state in this is recurrent, you have no transient states, and you have no periodicity. So it's an aperiodic chain. And then for each  $j$ , if you take the maximum path from  $i$  to  $j$  in  $n$  steps, this is non-increasing in  $n$ .

So in other words, this right here, this is non-increasing. So if I take the maximum path from state  $i$  to  $j$ , it gives you exactly  $n$  steps. So that means this is maximized over all initial states  $i$ . So it doesn't matter what state you start, and I take the maximum path.

And if I increase  $n$ , and I take maximum of that again, the maximum path, this is not increasing. And the minimum is non-decreasing in  $n$ . So as we take  $n$ , the path from  $i$  to  $j$ , this  $n$  getting larger and larger, we have that the maximum of this path, which is the most probable path, is non-increasing. And then the minimum of this path, the least likely path, is going to be non-decreasing.

So we're wondering whether this limit is going to converge or not. In this theorem it said that for an ergodic finite-state Markov chain, this limit actually does converge. So in other words, the  $\limsup$  is equal to  $\liminf$  of this and will equal  $\pi_j$ , which is the steady-state distribution. And not only that, this convergence is going to be exponential in  $n$ . So this is the theorem that we will prove today.

So the key to this theorem is this pair statements, that the most probable path from  $i$  to  $j$ , given  $n$  steps-- so this is the most probable path-- is non-increasing at  $n$ , and the minimum is non-decreasing in  $n$ . So the proof is almost trivial, but let's see what happens in this.

So we have a probably transition matrix. So this is the statement right here. And the transition is just one here and one here, with probability 1, 1. In this case, we want to say, what is the maximum path that we're in state 2, given  $n$  steps?

So we know that this probability alternates between 1 and 2, it's non-increasing, it's not decreasing, it's always the same. So those two bounds are met with equality. So in this here.

So the second example is this. We have a two-state chain again. But this time, from 1 to 2, we have the transition of  $3/4$ . So that means that we have a chain here of  $1/4$ . See, the minute we put a self-loop in here, it completely destroys the periodicity.

Any Markov chain, you put a self-loop in it, and the periodicity is destroyed. So here we have  $3/4$ . So this has to come back with  $1/4$ .

All right. So in this one, let's look at the  $n$  step going from 1 to 2. So basically, we want to end up in state 2 in exactly  $n$  steps. So when  $n$  is equal to 1, what is the maximum?

The maximum is if you start it in this state and then you went to state 2. The other alternative is you start at state 2, and you stay in state 2. Because we want to end at state 2 in exactly one step. So the maximum is going to be  $3/4$ , and the minimum is going to be  $1/4$ . You get  $n$  is equal to 2.

Now we want to end up in state 2 in two steps. So what is going to be the maximum? The maximum is going to be if you visit state 1 and then back.

So  $n$  is equal to 1. Then  $P_{1,2}$  from 1 to 2 is equal to  $3/4$ . So the probability from 1 to 2 in two steps is equal to  $3/8$ .

So it goes  $1/4$  plus  $3/4$ ,  $3/4$  plus  $1/4$ . It should be equal to  $3/8$ , right? Is that right?  
OK.

And then it when  $P_{1,2}$  to 3, if there are three transitions from 1 to 2, then it's equal to  $9/16$ . So if for 2, if I want to transition from 2 to 2  $n$  steps-- so  $P_{2,2}$  is equal to  $1/4$ . So it just stayed by itself.

So  $P_{2,2}$  in two steps, you don't have a choice. You have to go from  $3/4$  to  $3/4$ . So

that's  $9/16$ . But For thing is I can also stay here by  $1/4$  times  $1/4$ . So that gives me  $5/8$  and so forth.

So basically, the sequence going from 1 to 2 is going to be oscillating between  $3/4$ ,  $3/8$ ,  $9/16$ , and so forth. And then going from 2,2, it's going to be oscillating too. We can see that's  $1/4$ ,  $5/8$ ,  $7/16$ .

So what happens is this oscillation is going to converge-- it's going to approach, actually,  $1/2$ . So if we take the maximum of these two, so  $P_{1,2}$  and  $P_{2,2}$ -- because that means that we're going to end at state 2. And maximum over  $n$  steps, then we just look at these two numbers, the  $3/4$  and  $1/4$ , if we want the maximum, then it's going to be  $3/4$ . For the  $3/8$  and  $5/8$ , the maximum is going to be  $5/8$ , the  $9/16$  and  $7/16$ , the  $9/16$  will be the maximum.

And similarly, we compare it, and we take the minimum. And the minimum is  $1/4$ ,  $3/8$ , and  $7/16$ . So we see that the maximum is going to be-- it starts high. And then it's going to decrease toward  $1/2$ . And the minimum, what happens is it's going to start low, and then it's going to increase to  $1/2$ . So this is exactly this one here. So  $P$ 's transition makes this an arbitrary finite-state Markov chain.

Therefore, each  $j$ , this maximum path, the most problem path from  $i$  to  $j$  in  $n$  steps is non-increasing  $n$ . And the minimum is non-decreasing  $n$ . So you take  $n$  plus 1 steps from  $i$  to  $j$ . So we're going to use that Chapman-Kolmogorov equation. So we take the first step to some state  $k$ . And then we go from  $k$  to  $j$  in  $n$  steps. But then we sum this over all  $k$ .

But this  $P^n$  for state to  $j$  to  $k$  in  $n$  steps, I can just take the maximum path. So I take the most probable path, the state that gives me the most probable path, and I substitute this in. When I substitute this in, obviously every one of these guys is going to be less than or equal to this. Therefore, this outside term is going to be less than or equal.

So now this is just going to be a constant. So I sum over all  $k$ , and then this term

remains. So therefore, what we know is if I want to end up in state  $j$ , and for  $n$  steps, if I increase the step more, to  $n + 1$ , we know that this probability is going to stay the same or decrease. It's not going to increase. So you could do exactly the same thing for the minimum.

So if this is going to be true, then of course, if I think the maximum of this, it's also going to be less than that. Because this limit's true for Markov chain. It doesn't matter. It just has to be a finite-state Markov chain. So this is true for any finite-state Markov chain.

So if I take the maximum of this, it's less than or equal to the maximum of the  $n$ -th step. So  $n + 1$  steps, the path is going to be less probable when I take the maximum path, the fact that I end up at state  $j$  than  $n$ .

So before we complete the proof of this theorem, let's look at this case where  $P$  is greater than zero. So if we say  $P$  is greater than zero, this means that every entry in this matrix is greater than 0 for all  $i, j$ , which means that this graph is fully connected. So that means you could get from  $i$  to  $j$  in one step with nonzero probability.

So if  $P$  is greater than 0-- and let this be the transition matrix. So we'll prove this first, and then we'll extend it to the arbitrary finite Markov chain. So let  $\alpha$  here is equal to the minimum. So it's going to be the minimum element in this transition matrix. That means it's going to be the state that contains the minimum transition. So let's call  $\alpha$ -- it's the minimum probability. Excuse me.

So let all these states  $i$  and  $j$ . And for  $n$  greater than or equal to 1, we have these three expressions. So this first expression says this, that if I have an  $n + 1$  walk from  $i$  to  $j$ , I take the most probable of this walk over  $i$ . So my choices, I can choose my initial starting state. In  $n + 1$  steps, I want to end in state  $j$ . So I pick the most probable path.

If I minus this, which is the least probable path-- but you get to minimize this over  $i$ ,

over the initial starting a state. So this is less than or equal to the  $n$  step. It's exactly this term here, the  $n$  step times  $1 - 2\alpha$ . So  $\alpha$  is the minimum transition probability in this probability transition matrix.

So this one, it's not so obvious right now. But we are going to prove that in the next slide. So once we have this, we can iterate on  $n$  to get the second term.

So for this term inside here, the most probable path to state  $j$  in  $n$  steps, minus the least probable path to state  $j$  in  $n$  steps, is equal to exactly the same thing in  $n - 1$  steps times  $1 - 2\alpha$ . So we just keep on iterating this over  $n$ , and then we should get this.

So to prove this to this, we prove it by induction. We just have to prove the initial step, that the maximum single transition from  $i$  to  $j$ , minus the minimum single transition from  $i$  to  $j$ , is less than or equal to  $1 - 2\alpha$ . So this one is proved by induction.

So as  $n$  goes to infinity, notice that this term is going to go to 0. because  $\alpha$  is going to be less than a  $1/2$ . Because if it's not, then we can choose  $1 - \alpha$  to be this minimum. So if this is going to 0, this tells us the difference between the most probable path minus the least probable path, the fact that we end up in state  $j$ .

So if we take the limit as  $n$  goes to infinity of both of these, they should equal. Because the difference of this, we notice that it's going down exponentially in  $n$ . So this shows us that this limit indeed does exist and is equal.

We want to prove this first statement over here. So in order to prove this first statement, what we're going to do is we're going to take this  $i, j$  transition in  $n + 1$  transitions. And then we're going to express it as a function of  $n$  transitions.

So the idea is this. We're going to use the Chapman-Kolmogorov equations to have an intermediary step. So in order to do this  $i$  to  $j$  in  $n + 1$  steps, the most probable path, we're going to go to this intermediate step and then on to the final step. In this intermediate step, it's going to be a function of  $n$ . So we're going to take one step and then  $n$  more steps.

So what we're going to do is, the intuition is, we're going to remove the least probable path. So we remove that from the sum in this Chapman-Kolmogorov equation. And then we have the sum of everything else except for that path. And then the sum of everything else, we're going to bound it.

Once we bound it, then we have this expression. The probability of  $i$  to  $j$  in  $n + 1$  steps is going to be a function of a max and a min over  $n$  steps with a bunch of terms. So that's the intuition of how we're going to do it.

So the probability of  $ij$  going from state  $i$  to state  $j$  in exactly  $n + 1$  steps is equal to this. So it's the probability of going from  $i$  to  $k$ , this intermediate step. We're going to take one step to a state  $k$ . And then we're going from  $k$  to  $j$  in  $n$  steps, summing over all  $k$ . So this is exactly equal to this with Chapman-Kolmogorov.

So now what happens is we're going to take-- Before we get to this next step, let's define this  $l_{\min}$  to be the state that minimizes  $p$  of  $ij, n$  over  $i$ . So  $l_{\min}$  is going to be the state that's going to be such that the choices  $l$  pick over  $i$  that in  $n$  steps  $l$  arrive at  $j$  that's going to be the least probable. So this is  $l_{\min}$  over here. It's the  $l_{\min}$  that satisfies this.

Then I'm going to remove this. So this is one state.  $l_{\min}$  is just one state that  $i$  is going to go to in this first step. So we're going to remove it from the sum. So then, this is just here.

So that path goes from  $i$  to  $l_{\min}$  times  $l$  to  $j$  in  $n$  steps. So remove that one path from here. Now we have the sum over the rest of the cases because we just removed that.

So we have  $ik, kj$  to  $n$ , where  $k$  is not equal to that element. So we removed that path, the one that goes to that state. But  $p$  of  $kj, n$ , the path that goes from  $k$  to  $j$  in  $n$  steps, we can just bound this term by the maximum over  $l$  from  $l$  to  $j$  of  $n$ .

So then we're going to take the most probable path in  $n$  steps such that we end up

in state  $j$  in  $n$ . So this term right here is bounded by this term. Becomes is bounded by this, that's why we have this less than or equal sign. So we just do two things from this step, the first step, to the second step. So we took out the path that's going to minimize that right at the  $j$ -th node in  $n$  steps. And then we bounded the rest of this sum by this.

So when we sum this all up, this is just a constant here. And  $i_k$  here is just all the states that  $i$  is going to visit except for this one state,  $l$  min. Since it's just all of them except for that, it's just  $1$  minus the probability that it goes from state  $i$  to  $l$  min. So this sum here is just equal to this sum here. So this arrives here. And this term is still here.

So going from here, what happens is we just to rearrange the terms. So nothing happens right here. It's just rearranging.

Now we have this term here. So we look at this term,  $P$  from  $i$  going to  $l$  min-- Remember, we chose  $\alpha$  to be the minimum single transition probability, single transition in that probability transition matrix. So  $i$  to  $l$  has to be greater than that. But the minus of this has to be less than, the negative has to be less than. So this we can substitute here. So now we have this.

So the maximum over  $i$  of this  $n$  plus  $1$  step actually shows you the probability. Because this  $l$  can write as an  $n$  plus  $1$  step path from  $i$  to  $j$ . So if this is less than this entire term, of course  $l$  can write the maximum path from  $i$  to  $j$ . It also has to be less of this because this is satisfied for all  $i, j$ . So therefore, we arrive at this expression here.

So now we're kind of in good business because we have the  $n$  plus one step at transition, the maximum path from  $i$  to  $j$  in  $n$  plus  $1$  steps as a function of  $n$ , which is what we wanted, and a function of this  $\alpha$ .

So we repeat that last statement. And the last one is here, the last line. So now we have the maximum. So now we want to do is we want to get the minimum. So we do

exactly the same thing, with the same proof.

And with the minimum, what we're going to do is we're going to look at the  $ij$  transition in  $n + 1$  steps. And then what we're going to do is we're going to pull out the maximum this time. So we pull out the most probable path in  $n$  steps such that it arrives in state  $j$ . Then we play the same game. Would bound everything-- above, this time-- by the minimum of the  $n$  step transition probabilities to get to  $j$ .

So once we do that, we get this expression, very similar to this one up here. So now we have the maximum path, which is  $n + 1$  steps to  $j$ , and the minimum of  $n + 1$  steps to  $j$ . So we could take the difference between these two.

So if you subtract these equations here, so this first equation minus the second equation, we have this on the right-hand side here and then these terms over here on the left-hand side. So these terms over here on the left-hand exactly proves the first line of the lemma. So the first line of the lemma was here.

So now, to prove the second of the lemma, remember, we're going to prove this by induction. In order to prove this by induction, we need to be first initial step. So the initial step is this. So if I take the minimum transition probability from  $i$  to  $j$ , it has to be greater than here with the  $\alpha$ . Because we said that  $\alpha$  was the absolute minimum of all the single-step transition probabilities. Then the maximum transition probability has to be greater than or equal to  $1 - \alpha$ . It's just by definition of what we choose.

So therefore, if I take this term, the maximize minus the minimum is just  $1 - 2\alpha$ . So that's your first step in the induction process. So we iterate on  $n$ . When we iterate on  $n$ , one arrives at this equation down here.

So this shows us from here that if we take the limit as  $n$  goes to infinity of this term, this goes down exponentially in  $n$ . And both of these limits are going to converge, and they exist, and they're going to be greater than 0. So they'll be greater than 0 because of our initial state that we chose this path with a positive probability.

Yeah, go ahead.

**AUDIENCE:** It seems to me that alpha is the minimum, the smallest number in the transition matrix, right?

**SHAN-YUAN HO:** Alpha is the smallest number, correct.

**AUDIENCE:** Yeah. How does it fall from that, like that? So my is, the convergence rate is related to  $f$ ?

**SHAN-YUAN HO:** Yes, it is, yeah. In general, it doesn't really matter because it's still going to go down exponentially in  $n$ . But it does depend on that alpha, yes.

Any other questions? Yes.

**AUDIENCE:** Is the strength that bound it proportional to the size of that matrix, right?

**SHAN-YUAN HO:** Excuse me?

**AUDIENCE:** The strength of that bound is proportional to the size? I mean, for a very large finite-state Markov chain, the strength of the bound is going to be somewhat weak because alpha is going to be--

**SHAN-YUAN HO:** Alpha has to be less than  $1/2$ .

**AUDIENCE:** OK, yes. But the strength of the bound, though, it's not a very tight bound on  $\max$  minus  $\min$ . Because in a large--

**SHAN-YUAN HO:** Yes. This is just a bound. And the bound is what when we took it that minimum-probability path, the  $\min$ , remember? The bound was actually in here. So we took the minimum-probability path in  $n$  steps, this  $\min$  that minimizes this over  $i$ . And then this is where this less than or equal to here is just a substitution.

Any other questions?

So what we know is that what happens is that these limited-state probabilities exist. So we have a finite ergodic chain. So if the probability of the elements in this transition matrix are all greater than 0, we know that this limit exists. But we know

that in general, that may not be the case. We're going to have some 0's in our transition matrix.

So let's go back to the arbitrary finite-state ergodic chain with probability transition matrix  $P$ . So in the last slide, we showed that this transition matrix  $P$  of  $h$  is positive for  $h$  is equal to  $M$  minus 1 squared plus 1. So what we do is, we can apply lemma 2 to  $P$  of  $h$  with this  $\alpha$  equals to minimum going from  $i$  to  $j$  in exactly  $h$  steps.

So why is this  $M$  minus 1 squared plus 1? So in the last lecture-- so what it means is this. So what it says is here. This was an example given in the last lecture. It was a 6-state Markov chain.

So what it says is that if  $n$  is greater than or equal to  $M$  minus 1 squared plus 1-- in this case, it's going to be 6. So if  $n$  is greater than or equal to 26, then I take  $P$  to the 26th power, it means it's greater than zero. That means if I take  $P$  to the 26th power, every single element in this transition matrix is going to be non-zero, which means that you can go from any state to any state with nonzero probability, as long as  $n$  is bigger than that. So basically, in this Markov chain, if you go long enough, long enough. Then I say, OK, I want to go from state  $i$  to state  $j$  in exactly how many steps, there is a positive probability that this is going to happen.

So how did this bound come across? Well, for instance, in this chain, if we look at  $P_{1,1}$  so we have here? So I'm going to look at the transition starting at state 1. And I want to come back to 1. So you definitely could come back at 6, because these are all positive probability 1. So 6 is possible. So  $n$  is equal to 6 is possible.

So what's the next one that's possible?  $n$  is equal to 11, right? Then the next one is what? 16 is possible, right? So 0 to 5 is impossible, is 0. So if I pick  $n$  between 0 and 5, and 7 and 10, you're toast. You can't get back to 1.

And so forth. So 18 is possible. 21-- let's see, is 17 possible? Yeah, 17 is also possible.

**AUDIENCE:** Why is 16 possible?

**SHAN-YUAN HO:** So I go around here twice, and then the last one. Is that right? So if I go from here to here to here to here, if I go twice, and then one more in the final loop.

**AUDIENCE:** That's 12.

**SHAN-YUAN HO:** Oh, it's 12? No. I'm going to go this inner loop right here. So if I go from 1 to 2 to 3 to 4 to 5 to 6, down to 2. Then I go 3, 4, 5, 6, 1. That's 11, isn't it? So 16 is I'm going to go around the inner loop twice.

OK. Go ahead. Question?

**AUDIENCE:** So everything 20 and under is possible, right?

**SHAN-YUAN HO:** No. Is 25 possible? Tell me how you're going to go 25 on this. You just do the 5 loop 5 times.

**SHAN-YUAN HO:** Yeah, but I want to go from 1 to 1. You're starting in state 1.

**AUDIENCE:** Oh, oh, sorry. OK.

**SHAN-YUAN HO:** 1 to 1, right?

**AUDIENCE:** OK, cool. OK, I see.

**SHAN-YUAN HO:** So you know for this one that this bound is actually tight. So 25 is impossible. So  $P_{1,1}$  of 25 is equal to 0. There's no way you can do that. But for 26 on, then you can. So what you're noticing is that you need this loop of 6 here and that any combination of 5 or 6 is possible.

So basically, in this particular example, if  $n$  is equal to  $6k$  plus  $5j$ , where  $k$  is greater than or equal to 1-- because I need that final loop to get back-- or  $j$  is greater than or equal to 0-- So any combination of this one, then I can express  $n$ . I can go around it to give me a positive probability of going from state 1 to state 1.

So I'm going to prove this using extremal property. So we're going to take the absolute worst case. So the absolute worst case is that for  $M$  state finite Markov

chain is if you have a loop of  $m$  and you have a loop of  $m - 1$ . You can't just have a loop of  $m$ . The problem is now this becomes periodic.

So we have to get rid of the periodicity. If you add a single group here, that doesn't help you. Then after 6, then it I get 7, 8, 9, 10, 11, 12. That didn't have this.

So the absolute worst case for an  $M$  state chain is going to be something that looks like this. 1 that goes to 2-- you're forced to go to 2-- so forth, until state  $M$ . And then this  $M$  is going to go back to 2 or is going to go back to 1. So in other words, the worst case is if you have--  $n$  has to be some combination of  $Mk + M - 1 - j$ . So this will be the worst possible case for  $M$  state Markov chain.

So it'll be  $Mk + M - 1 - j$ . So  $k$  has to be greater than or equal to 1. And then  $j$  has to be greater than or equal to 0, because you need to come back. So I'm just looking at the case probability that I start in state 1 and I come back in state 1. So all right.

So how do we get this bound? Well, there is an identity that says this. If  $a$  and  $b$  are relatively prime, then the largest  $n$  such that it cannot be written-- so we want to find the largest  $n$  such that  $ak + bj$ -- but this is  $k$  and  $j$  greater than or equal to 0-- that it cannot be written in this form. The largest integer that it cannot be written is  $ab - a - b$ . This takes a little bit to prove, but it's not too hard. If you want to know this proof, come see me offline after class.

This is the largest integer. If  $n$  is equal to this, it cannot be written in this form. But if  $n$  is greater than this, then it can. So all we do is substitute  $M$  for  $a$  and  $M - 1$  for  $b$  because  $M$  and  $M - 1$  are relatively prime. But remember, we have a  $k$  here that has to be greater than or equal to 1. We need at least one  $k$ . But this so identity is for  $k$  and  $j$  greater than 0. So therefore, we have to subtract out that  $k$ .

So therefore, we have  $M$  times  $M - 1$ , minus  $M - 1 - M - 1$ . But the thing is we have to add the extra  $M$ , because this  $k$  is greater than or equal to 1. So we have to add up one of the  $M$ 's because of this. So this is just equal to  $M - 1$ ,

squared.

So this number, if  $n$  is equal to this, it's the largest number that it cannot be written like that. So therefore, we have to add 1. So that's why the bound is equal to 1.

So the upper bound that  $n$  can be written is going to be  $M$  minus 1, squared plus 1.

**AUDIENCE:** Why did you add the 1 at the end?

**SHAN-YUAN HO:** This one?

**AUDIENCE:** No, we've got to do the 1 at the end.

**AUDIENCE:** We already have that in there.

**SHAN-YUAN HO:** Oh, where is it? No, it's in here, right?

**AUDIENCE:** No, it's not here.

**SHAN-YUAN HO:** Did I-- What are you talking about? Where's the 1?

**AUDIENCE:** At the end, the last equation.

**SHAN-YUAN HO:** This one?

**AUDIENCE:** Yes.

**SHAN-YUAN HO:** OK. This is the "cannot," largest  $n$  which you cannot write. You cannot write this. So this bound is tight. It means that this is the one that you can.

So if  $n$  is greater than or equal to this, then it's possible. This is the largest one it cannot. Based on this, it cannot. So we have to add the 1. So therefore, in here, you could do 26. So starting from 26, 27, 28, you can do that. Any questions?

**AUDIENCE:** Relatively prime, what do you mean by "relatively"?

**SHAN-YUAN HO:** There is a greatest common divisor of 1.

So if we take  $h$  here,  $h$  is going to be positive. So if  $h$  is equal to  $M$  minus 1, squared

plus 1, then now all the elements are positive. Because we just proved that we can write this-- every state can be visited by any other state, with positive probability. So we say, looking at  $P$ , we know that  $P$  of  $h$  is positive for  $h$  greater than or equal to this bound.

So what we do is we applied this lemma 2 probability to this transition matrix  $P$  of  $h$ , where we have picked  $\alpha$ -- remember,  $\alpha$  is the single-step transition probability. So instead of the single transition, we have lumped this  $P$  into  $P$  to the  $h$  power. So it's  $h$  steps. Because we proved the result before for positive  $P$ . So this  $P$  to the  $h$  is positive, so we take  $\alpha$  as the minimum from  $i$  to  $j$  of  $P$  to the  $h$  in this matrix.

So it doesn't really matter what the value of  $\alpha$  is, only that it's going to be positive. And it has to be positive because it's a probability. So what happens is, if we follow the proof of what we just showed in the lemma, then we show that the maximum path from  $i$  to  $j$ --  $h$  times  $M$ . So  $M$  is going to be an integer, so in multiples of  $h$ -- this upper limit is going to be equal to the lower limit. So the most probable path is equal to the least probable path.

So this is multiple of  $h$ 's. So if we take this as  $M$  goes to infinity, this has got to equal to-- Oops, this should be going to  $\pi$  sub  $j$ , excuse me. This little temple here. And this is going to be greater than 0.

So the problem is now we've shown it for multiples of  $h$ 's, what about the  $h$ 's in between? But the fact is that lemma 1, we showed that this maximum path from  $i$  to  $j$  in  $n$  is not increasing in  $n$ . So all those states, all those paths, the transition probability for the paths in between these multiples of  $h$ 's, in between them it's going to be not increasing in  $n$ . So even if we're taking these multiples of each of  $h$  and  $n$  here, here, here, and we know that this limit is increasing, we know that all the ones in between them are also going to be increasing to the same limit because of lemma 1.

To remember, the maximum is going to be not increasing, and the minimum is going to be non-decreasing in any one path. So this must have the same limit as

this multiple of this. So the same limit applies. So any questions on this? So this is how we prove it for the arbitrary finite-state ergodic chain when we have some 0 probability transition elements in the matrix  $P$ . So the proof is the same.

So now for ergodic unichain. So we see that this limit as  $n$  approaches infinity from  $i$  to  $j$  of  $n$  is going to just end up in the steady-state transition  $\pi_j$  for all  $i$ . So it doesn't matter what your initial state is. As  $n$  goes to infinity of this path, as this Markov chain goes on and on, you will end up in state  $j$  with probability  $\pi_j$ , where  $\pi$  is this probability vector.

So now we have this steady-state vector, and then we can solve for the steady-state vector solution. So this  $\pi P$  is equal to  $\pi$ .

Yeah? Go ahead.

**AUDIENCE:** Where did you prove that the sum of all the  $\pi_j$ 's equal to one? Because you say that we proved that this is the probability vector. But did prove only that it is non-negative?

**SHAN-YUAN HO:** It's non-negative. But the thing is because as  $n$  goes to infinity, you have to land up someone, right? This is a finite-state Markov chain. You have to be somewhere.

And the fact that you have to be somewhere, your whole state space has to add up to 1. Because it's a constant, remember? For every  $j$ , as  $n$  goes to infinity, it goes to  $\pi_j$ . So you have that for every single state. And then you have to end up somewhere.

So if you have to end up somewhere, the space has to add up to one. Yeah, good question.

So why are we interested in this  $\pi_j$ ? The question is that because in this recurrent class, it tells us that as this goes to infinity, we see this sequence of states going back and forth, back and forth. And we know that as  $n$  goes to infinity, we have some probability,  $\pi_j$ , of landing in state  $j$ ,  $\pi_i$  of landing in state  $i$ , and so forth. So it says that in the  $n$  step, as  $n$  goes to infinity, that this is the fraction of

time that, actually, that state is going to be visited. Because at each step, you have to make a transition.

So it's kind of the expected number of times per unit time. So it's divide by  $n$ . It's going to be that fraction of time that you're going to visit that state. It's the fraction of time that you're going to be in that state. It's this limiting state as  $n$  gets very, very large. So we will see that in the next few chapters when we do renewal theory that this will come into useful play. And we give a slightly different viewpoint of it.

So it's very easy to extend this result to a more general class of ergodic unichains. So remember the ergodic unichains, now we have increased these transient states. So before, we proved this. We just proved it for it contains exactly one class. It's aperiodic, so we have no cycles, no periodicity in this Markov chain. And so we know that the steady-state transition probabilities have a limit. And the upper limit and the lower limit of these paths as they go to infinity-- in fact, they end up in a particular state-- has a limit. And we have this steady-state probability vector that describes this.

So now we have these transient states. So these transient states of this Markov chain, what happens is there exists a path that this transient state is going to go to a recurrent state. So once it leaves this transient state, it goes to recurrent state. It's never going to come back.

So there is some probability,  $\alpha$ , of leaving the class at each step. So there's some transition probability in this transient state that's going to be  $\alpha$ . And the probability of remaining in this transient state is just  $1 - \alpha$  to the  $n$ . And this goes down exponentially.

So what this says is that eventually, as  $n$  gets very large, it's very, very hard to stay in that transient state. So it's going to go out of the transient state. And then it will go into the recurrent class.

So when one does the analysis for this, what happens in the probability in this steady-state vector is those transient states, this  $\pi_i$ , will be equal to 0. So this

distribution is only going to be non-zero for recurrent states in this Markov chains. And the transient states will have probability equal to 0. In the notes, they just extend the argument. But you need a little bit more care to show this. And it divides the transient states into a block and then the recurrent classes into another block and then shows that these transient states' limiting probability is going to go to 0.

So let's see. So this says just what I said, that these transient states decay exponentially, and one of the paths will be taken, eventually, out of it. So for ergodic unichains, the ergodic class is eventually entered, and then steady state in that class is reached. So every state  $j$ , we have exactly this. The maximum path from  $i$  to  $j$  in  $n$  steps-- and the minimum path. We look at the minimum path in  $n$  steps and the maximum path in  $n$  steps. And for each  $n$ , we take the limit as  $n$  goes to infinity. These guys, these limits are exactly equal, and it equals to this  $\pi_j$ , which is equal to the  $j$  state.

So your initial states, how you went the paths that you have gone is completely wiped out. And all that matters is this final state, as  $n$  gets very large. So the difference here is that  $\pi_j$  equals 0 for each transient state, and it's greater than 0 for the recurrent state.

So other finite Markov chains. So we can consider a Markov chain with several ergodic classes. Because we just considered it with one ergodic class. So if the classes don't communicate, then you just consider it separately. So you figure out the steady-state transition probabilities for each of the classes separately.

But if you have to insist on analyzing the entire chain  $P$ , then this  $P$  will have  $m$  independent steady-state vectors and one non-zero in each class. So this  $P^n$  is still going to converge, but the rows are not going to be the same. So basically, you're going to have blocks. So if you have one class, say 1 through  $k$  is going to be in one class, and then  $k$  through  $l$  is going to be another class, and then  $l$  through  $z$  is going to another class, you have a block. So this steady-state vector is going to be in blocks.

So you can see the recurring classes only communicate within themselves. Because

these don't communicate, so they're separate. So you could have a lot of 0's in limiting state, if you look at this,  $P^n$  goes to infinity. So there is a set of rows, one for each class. And a row for each class  $k$  will be non-zero for the elements of that class.

So then finally, if we have periodicity. So now if we have a periodic recurrent chain with period  $d$ . We had the two where it's just a period of 2. So with periodicity, what you do is you're going to divide these classes into  $d$  different states. So you have to go to one state-- So if there's  $d$  states, this is a period of  $d$ , you separate or you partition the states into  $d$  of them,  $d$  subclasses, with a cycle rotation between them. So basically, each time unit, you have to go from one class to the next class.

And then we do that, then for each class, you could have the limiting-state probability. So in other words, you are looking at this transition matrix,  $P^d$ . Because when it cycles, it totally depends on which one you start out at. But if you look at the  $d$  intervals, then that becomes the ergodic class by itself. And there are exactly  $d$  of them. So the limit as  $n$  approaches infinity of  $P^{nd}$ , this thing also exists, but exists in the subclass sense of there is  $d$  subclasses if it has a period of  $d$ .

So that means a steady state is reached within each subclass, but the chain rotates from one subclass to another. Yeah, go ahead.

**AUDIENCE:** In this case, if we do a simple check with 1 and 2, with 1 and 1, it doesn't converge.

**SHAN-YUAN HO:** No, it does. It is 1, converges to 1. So it's 1, and then it's going to be 1.

**AUDIENCE:** It's 1, 1, 1, 1, 1, 1, 1, 1. So you go here? Like, it's reached--?

**SHAN-YUAN HO:** No, no. It converges for here. But this  $d$  is equal to 2, in that case. So you have to do  $nd$ , so you've got to look at  $P^2$ . So if I look at  $P^2$ , I'm always a 1-- 1, 1, 1, 1, 1, 1, 1, 1. That's converging. The other one is 2, 2, 2, 2, 2, 2. That's also converging.

OK. So is there any other questions about this? OK, that's it. Thank you.