

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.

PROFESSOR: OK, let's get started again on finite-state Markov chains. Sorry I was away last week. It was a long-term commitment that I had to honor. But I think I will be around for all the rest of the lectures. So I want to start out by reviewing just a little bit. I'm spending a lot more time on finite-state Markov chains than we usually do in this course, partly because I've rewritten this section, partly because I think the material is very important. It's sort of bread-and-butter stuff, of discrete stochastic processes. You use it all the time. It's a foundation for almost everything else.

And after thinking about it for a long time, it really isn't all that complicated. I used to think that all these details of finding eigenvalues and eigenvectors and so on was extremely tedious. And it turns out that there's a very nice pleasant theory there. You can find all of these things after you know what you're doing by very simple computer packages. But they don't help if you don't know what's going on. So here, we're trying to figure out what's going on.

So let's start out by reviewing what we know about ergodic unit chains and proceed from there. An ergodic finite-state Markov chain has transition probabilities which, if you look at the transition matrix raised to the n th power, what that gives you is the transition probabilities of an n -step Markov chain. In other words, you start at time 0, and at time n , you look at what state you're in. P_{ij} to the n th power is then the probability that you're in state j at time n , given that you're in state i at time 0. So this has all the information that you want about what happens to Markov chain as time gets large.

One of the things we're most concerned with is, do you go to steady state? And if you do go to steady state, how fast do you go to steady state? And of course, this

matrix tells you the whole story there, because if you go to steady state, and the Markov chain forgets where it started, then P_{ij}^n goes to some constant, π_j , which is independent of the starting state, i , and independent of n , asymptotically, as n gets big. So this π is a strictly positive probability vector. I shouldn't say so it is. That's something that was shown last time.

If you multiply both sides of this equation by $\sum_k P_{jk}$, then what do you get? You get $\sum_k P_{ik}^{n+1}$. That goes to a limit also. If the limit in n goes to infinity-- then the limit as $n+1$ goes to infinity is clearly the same thing. So this quantity here is the sum over j , of $\pi_j P_{jk}$. And this quantity is equal to π_k , just by definition of this quantity. So $\sum_k P_{jk}$ is equal to $\sum_j \pi_j$.

$\sum_k P_{jk}$, what does that say? That's the definition of a steady state vector. That's the definition of, if your probabilities of being in state k satisfy this equation, then one step later, you still have the same probability of being in state k . Two steps later, you still have the same probability of being in state k . So this is called the steady state equation. And a solution to that is called a steady state vector. And that satisfies this.

In matrix terms, if you write this out, what does it say? It says the limit as n approaches infinity of P^n is equal to the column vector, e of all 1s. The transpose here means it's a column vector. So you have a column vector times a row vector. Now, you know if you have a row vector times a column vector, that just gives you a number. If you have a column vector times a row vector, what happens? Well, for each element of the column, you get this whole row. And for the next element of the column, you get the whole row down beneath it multiplied by the element of the column, and so forth, day on.

So a column vector times a row vector is, in fact, a whole matrix. It's a j by j matrix. And since e is all 1s, what that matrix is is a matrix where every row is a steady state vector π . So we're saying not only does this π that we're talking about satisfy this steady state equation, but more important, it's this limiting vector here. And as n goes to infinity, you in fact do forget where you were. And the entire matrix of where

you are at time n , given where you were at time 0, goes to just this fixed vector π .

So this is a column vector, and π is a row vector then. The same result almost holds for ergodic unit chains. What's an ergodic unit chain? An ergodic unit chain is an ergodic set of states plus a whole bunch of transient states. Doesn't matter whether the transient states are one class of transient states or whether it's multiple classes of transient states. It's just transient states. And there's one recurrent class. And we're assuming here that it's recurrent. So you can almost see intuitively that if you start out in any one of these transient states, you bum around through the transient states for a while. And eventually, you flop off into the recurrent class. And once you're in the recurrent class, there's no return. So you stay there forever.

Now, that's something that has to be proven. And it's proven in the notes. It was probably proven last time. But anyway, what happens then is that the sole difference between ergodic unit chains and just having a completely ergodic Markov chain is that the steady state factor is now positive for all ergodic states and it's 0 for all transient states. And aside from that, you still get the same behavior still. As n gets large, you go to the steady state vector, which is the steady state vector of the ergodic chain.

If you're doing this stuff by hand, how do you do it? Well, you start out just with the ergodic class. I mean, you might as well ignore everything else, because you know that eventually you're in that ergodic class. And you find the steady state vector in that ergodic class, and that's the steady state vector you're going to wind up with. This is one advantage of understanding what you're doing, because if you don't understand what you're doing and you're just using computer programs, then you never have any idea what's ergodic, what's not ergodic or anything else. You just plug it, you grind away, you get some answer and say, ah, I'll publish a paper. And you put down exactly what the computer says, but you have no interpretation of it at all.

So the other way of looking at this is, when you have a bunch of transient states, and you also have an ergodic class, you can represent a matrix if the recurrent

states are at the end of the chain and the transient states are at the beginning of the chain. This matrix here is the matrix of transition probabilities within the recurrent class. These are the probabilities for going from the transient states to the recurrent class. And once you get over here, the only place to go is down here. And the transient class is just a t by t class. And the recurrent class is just a j minus t by j minus t matrix.

So the idea is that each transient state eventually has a transition to a recurrent state, and the class of recurrent states leads to steady state as before. So that really, all that analysis of ergodic unit chains, if you look at it intuitively, it's all obvious. Now, as in much of mathematics, knowing that something is obvious does not relieve you of the need to prove it, because sometimes you find that something that looks obvious is true most of the time but not all of the time. And that's the purpose of doing these things.

There's another way to express this eigenvalue, eigenvector equation we have here. And that is that the transition matrix minus λ times the identity matrix times the column vector v is equal to 0. That's the same as the equation $Pv = \lambda v$. That's the same as a right eigenvector. Well, this is the equation for an eigenvalue λ . This is an equation for an arbitrary eigenvalue λ . But $Pv = \lambda v$ is the same as $(P - \lambda I)v = 0$.

Why do we even bother to say something so obvious? Well, because when you look at linear algebra, how many of you have never studied any linear algebra at all or have only studied completely mathematical linear algebra, where you never deal with n -tuples as vectors or matrices or any things like this? Is there anyone? If you don't have this background, pick up-- what's his name?

AUDIENCE: Strang.

PROFESSOR: Strang. Strang's book. It's a remarkably simple-minded book which says everything as clearly as it can be stated. And it tells you everything you have to know. And it does it in a very straightforward way. So I highly recommend it to get any of the background that you might need. Most of you, I'm sure, are very familiar with these

things. So I'm just reminding you of then. Now, a square matrix is singular if there's a vector v , such that a times v is equal to 0. That's just a definition as a singularity.

Now, λ is an eigenvalue of a matrix p if and only if p minus λ times i is singular. In other words, if there's some v for which p minus λ times i times v is equal to 0, that's what this says. You put p minus λ times i in for a , and it says it's singular if there's some v for which this matrix-- this matrix is singular if there's some v such that p minus λ times i times v is equal to 0.

So let a_1 to a_m be the columns of a . Then a is going to be singular if a_1 to a_m are linearly dependent. In other words, if there's some set of coefficients you can attach to a_1 times v_1 plus a_2 times v_2 , plus up to a_m times v_m such that that sum is equal to 0, that means that a_1 to a_m are linearly dependent. It also means that the matrix a times that v is equal to 0. So those two things say the same thing again. So the square matrix, a , is singular if and only if the rows of a are linearly independent. We set columns here. Here, we're doing the same thing for rows. It still holds true. And one new thing, if and only if the determinant of a is equal to 0.

One of the nice things about determinants is that determinants are 0 if the matrix is singular, if and only if the matrix is singular. So the summary of all of this for a matrix which is a transition matrix-- namely, a stochastic matrix-- is λ , is an eigenvalue of p , if and only if p minus λ times i is singular, if and only if the determinant of p minus λ times i is equal to 0, if and only if p times some vector v equals λ times v , and if and only if u times p equals λ times u for some u . Yes?

AUDIENCE: The second to last statement is actually linearly independent, you said? The second to last. Square matrix a . No, above that.

PROFESSOR: Oh, above that. A square matrix a is singular if and only if the rows of a are linearly dependent, yes.

AUDIENCE: Dependent.

PROFESSOR: Dependent, yes. In other words, if there's some vector v such that a times v is equal to 0, that means that those columns are linearly dependent. So we need all of those

relationships. It says for every stochastic matrix-- oh, now this is something new. For every stochastic matrix, P times e is equal to e . Obviously, because if you sum up the sum of P_{ij} over j is equal to 1. P_{ij} is the probability, given that you start in state i , that in the next step, you'll be in state j . You have to be somewhere in the next step. So if you sum these quantities up, you have to get 1, which says you have to be some place.

So that's all this is saying. That's true for every finite-state Markov chain in the world, no matter how ugly it is, how many sets of recurrent states it has, how much periodicity it has. A complete generality, P times e is equal to e . So $\lambda = 1$ is always an eigenvalue of a stochastic matrix, and e is always a right eigenvector. Well, from what we've just said, that means there has to be a left eigenvector also. So there has to be some π such that πP is equal to π .

So suddenly, we find there's also a left eigenvector. What we haven't shown yet is that that π that satisfies this equation is a probability vector. Namely, we haven't shown that all the components of π are greater than or equal to 0. We still have to do that. And in fact, that's not completely trivial. If we can find such a vector that is a probability vector, the components sum to 1 and they're not negative, then this is the equation for a steady state vector. So what we don't know yet is whether a steady state vector exists. We do know that a left eigenvector exists.

We're going to show later that there is a steady state vector π . In other words, a non-negative vector which sums to 1 for all finite-state Markov chains. In other words, no matter how messy it is, just like e , the column vector of all 1s is always a right eigenvector of eigenvalue 1. There is always a non-negative vector π whose components sum to 1, which is a left eigenvector with eigenvalue 1. So these two relationships hold everywhere.

Incidentally, the notes at one point claim to have shown this. And the notes really don't show it. I'm going to show it to you today. I'm sorry for that. It's something I've known for so long that I find it hard to say is this true or not. Of course it's true. But it does have to be shown, and I will show it to you later on. Chapter three of the notes

is largely rewritten this year. And it has a few more typos in it than most of the other chapters. And a few of the typos are fairly important. I'll try to point some of them out as we go. But I'm sure I haven't caught them all yet.

Now, what is the determinant of an M by M matrix? It's this very simple-looking but rather messy formula, which says the determinant of a square matrix A is the sum over all partitions-- and then there's a plus minus here, which I'll talk about later-- of the product from i equals 1 to M . M is the number of states of A sub i . This is the component of the ij position. And we're taking A sub i . And then the partition that we're dealing with, μ sub i .

So what we're doing is taking a matrix with all sorts of terms in it-- A_{11} up to A_{1j} on to A_{j1} up to A sub jj . And these partitions we're talking about are ways of selecting one element from each row and one element from each column. Namely, that first sum there is talking about one element from each row. And then when we're talking about a permutation here, we're doing something like, for this row, we're looking at, say, this element. For this row, we might be looking at this element. For this row, we might be looking at this element, and so forth down, until finally, we're looking at some element down here.

Now, we've picked out every column and every row in doing this, but we only have one element in each row and one element in each column. If you've studied linear algebra and you're at all interested in computation, the first thing that everybody tells you is that this is a god-awful way to ever compute a determinant, because the number of permutations grows very, very fast with the size of the matrix. And therefore you don't want to use this formula very often. It's a very useful formula conceptually, though, because if we look at the determinant of p minus λ i , if we want to ask the question, how many eigenvalues does this transition matrix have? well, the number of eigenvalues it has is the number of values of λ such that the determinant of p minus λ i is 0.

Now, how many such values are there? Well, you look the matrix for that, and you get A_{11} minus λ A_{12} and A_{22} minus λ A_{jj} minus λ . And none of

the other elements have λ in it. So when you're looking at this formula for finding the determinant, one of the partitions is this partition, which is a polynomial of degree j in λ . All of the others are polynomials of degree less than j in λ . And therefore this whole bloody mess here is a polynomial of degree j in λ .

So the equation, determinant of p minus λ i , which is a polynomial of degree j in λ , equals 0. How many roots does it have? Well, the fundamental theorem of algebra says that a polynomial of degree j , of complex numbers-- and real is a special case of complex-- that it has exactly j roots. So there are exactly, in this case, M -- excuse me, I've been calling it j sometimes and M sometimes. This equation here has exactly M roots to it. And since it has exactly M roots, that's the number of eigenvalues there are.

There's one flaw in that argument. And that is, some of the roots might be repeated. Say you have M roots altogether. Some of them appear more than one time, so you'll have roots of multiplicity, something or other. And when you add up the multiplicities of each of the distinct eigenvalues, you get capital M , which is the number of states. So the number of different eigenvalues is less than or equal to M . And the number of distinct eigenvalues times the multiplicity of each eigenvalue is equal to M .

That's a simple, straightforward fact. And it's worth remembering. So there are M roots to the equation. Determinant p minus λ i equals 0. And therefore there are M eigenvalues of p . And therefore you might think that there are M eigenvectors. That, unfortunately, is not true necessarily. That's one of the really-- it's probably the only really ugly thing in linear algebra. I mean, linear algebra is a beautiful theory.

I mean, it's like Poisson's stochastic processes. Everything that can be true is true. And if something isn't true, there's a simple counter-example of why it can't be true. This thing is just a bloody mess. But unfortunately, if you have M states in a finite-state Markov chain, you might not have M different eigenvectors. And that's

unfortunate, but we will forget about that for as long as we can, and we'll finally come back to it towards the end.

AUDIENCE: [INAUDIBLE]?

PROFESSOR: What?

AUDIENCE: Why would we care about all the eigenvectors if we are only concerned with the ones that [INAUDIBLE]?

PROFESSOR: Well, because we're interested in the other ones because that tells us how fast p to the M converges to what it should be. I mean, all those other eigenvalues, as we'll see, are the error terms in p to the M as it approaches this asymptotic value. And therefore we want to know what those eigenvalues are. At least we want to know what the second-biggest eigenvalue is. Now, let's look at just a case of two states. Most of the things that can happen will happen with two states, except for this ugly thing that I told you about that can't happen with two states. And therefore two states is a good thing to look at, because with two states, you can calculate everything very easily and you don't have to use any linear algebra.

So if we look at a Markov chain with two states, P_{ij} is this set of transition probabilities. The left eigenvector equation is $\pi_1 P_{11} + \pi_2 P_{21} = \lambda \pi_1$. And so this is writing out what we said before. The vector π times the matrix P is equal to λ times the vector π . That covers both of these equations. Since M is only 2, we only have to write things out twice. Same thing for the right eigenvector equation. That's this.

The determinant of $P - \lambda I$, if we use this formula that we talked about here, you put $A_{11} - \lambda$, $A_{22} - \lambda$. Well, then you're done. So all you need is $(P_{11} - \lambda)(P_{22} - \lambda) - P_{12}P_{21}$. That's this permutation there. And then you have an odd permutation, $A_{12}A_{21}$. How do you know which permutations are even and which permutations are odd?

It's how many flips you have to do. But to see that that's consistent, you really have to look at Strang or some book on linear algebra, because it's not relevant here. But

anyway, that determinant is equal to this quantity here. That's a polynomial of degree 2 in λ . If you solve it, you find out that one solution is $\lambda = 1$. The other solution is $\lambda = 1 - P_{12} - P_{21}$.

Now, there are a bunch of cases to look at here. If the off-diagonal transition probabilities are both 0, what does that mean? It means if you start in state 0, you stay there. If you start in state 1, you stay there forever. If you start in state 2, you stay there forever. That's a very boring Markov chain, but it's not very nice for the theory. So we're going to leave that case out for the time being. But anyway, if you have that case, then the chain has two recurrent classes. $\lambda = 1$, has multiplicity 2. You have two eigenvalues of algebraic multiplicity 2. I mean, it's just one number, but it appears twice in this determinant equation. And it also appears twice in the sense that you have two recurrent classes.

And you will find that there are two linearly independent left eigenvectors, two linearly independent right eigenvectors. And how do you find those? You use your common sense and you say, well, if you start in state 1, you're always there. If you start in state 2, you're always there. Why do I even look at these two states? This is a crazy thing where wherever I start, I stay there and I only look at state 1 or state 2. It's scarcely even a Markov chain.

If P_{12} and P_{21} are both 1, what it means is you can never go from state 1 to state 1. You always go from state 1 to state 2. And you always go from state 2 to state 1. It means you have a two-state periodic chain. And that's the other crazy case. The other case is not very interesting. There's nothing stochastic about it at all. So the chain is periodic. And if you look at this equation here, the second eigenvalue is equal to minus 1.

I might as well tell you that, in general, if you have a periodic Markov chain, just one recurrent class and it's periodic, a period d , then the eigenvalues turn out to be the uniformly spaced eigenvalues around the unit circle. One is one of the eigenvalues. We've already seen that. And the other $d - 1$ eigenvalues are those uniformly spaced around the unit circle. So they add up to 360 degrees when you get all done

with it.

So that's an easy case. Proving that is tedious. It's done in the notes. It's not even done in the notes. It's done in one of the exercises. And you can do it if you choose. So let's look at these eigenvector equations and the eigenvalue equations. Incidentally, if you don't know what the eigenvalues are, is this a linear set of equations? No, it's a nonlinear set of equations. This is a nonlinear set of equations in π_1 , π_2 , and λ .

How do you solve non-linear equations like that? Well, if you have much sense, you first find out what λ is and then you solve linear equations. And you can always do that. We've said that these solutions for λ , there can only be M of them. And you can find them by solving this polynomial equation. Then you can solve the linear equation by finding the eigenvectors. There are packages to do all of these things, so there's nothing you should waste time on doing here. It's just knowing what the results are that's important.

From now on, I'm going to assume that P_{12} or P_{21} are greater than 0. In other words, I'm going to assume that we don't have the periodic case and we don't have the case where you have two classes of states. In other words, I'm going to assume that our Markov chain is actually ergodic. That's the assumption that I'm making here. If you then solve these equations using $\lambda = 1$, you'll find out that π_1 is the component sum to 1. First component is P_{21} over the sum. Second component is P_{12} over the sum. Not very interesting.

Why is the steady state probability weighted towards the largest of these transition probabilities? If P_{21} is bigger than P_{12} , how do you know intuitively that you're going to be in state 1 more than you're in state 2? Is this intuitively obvious to-- yeah?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Because you make more transistors from 2 to 1. Well, actually you don't make more transitions from 2 to 1. You make exactly the same number of transitions, but since

the probability is higher, it means you have to be in state 1 more of the time. Good. So these are the two. And this is the left eigenvector for the second eigenvalue-- namely, the smaller eigenvalue. Now, if you look at these equations, you'll notice that the vector π_i , the left i -th eigenvector, multiplied by the right j -th eigenvector, is always equal to δ_{ij} . In other words, the left eigenvectors are orthogonal to the right eigenvectors.

I mean, you can see this just by multiplying it out. You multiply π_i times ν_j , and what do you get? You get this plus this, which is 1. δ_{ij} means there's something which is 1 when i is equal j and 0 when i is unequal to j . You take this and you multiply it by this, and what do you get? You get $\pi_i \nu_j$ times $\nu_j \pi_i$ over the square. Minus $\pi_i \nu_j$ times $\nu_j \pi_i$, it's 0. Same thing here. $1 - 1$, that vector times this vector, is 0 again. So the cross-terms are 0. The diagonal terms are 1. That's the way it is.

So let's move on with this. These right eigenvector equations, you can write them in matrix form. I'm doing this slowly. I hope I'm not boring those who have done a lot of linear algebra too much. But they won't go on forever, and it gets us to where we want to go. So if you take these two equations and you write them in matrix form, what you get is $P \nu$, where ν is a matrix whose columns are the vector ν_1 and the vector ν_2 . And capital Λ is the diagonal matrix of the eigenvalues. If you multiply P times the first column of ν , and then you look at the first column of this matrix, what you get-- yes, that's exactly the right way to do it.

And if you're not doing that, you're probably not understanding it. But if you just think of ordinary matrix vector multiplication, this all works out. Because of this orthogonality relationship, we see that the matrix whose rows are the left eigenvectors times the matrix whose columns are the right eigenvectors, that's equal to I . Namely, it's equal to the identity matrix. That's what this orthogonality relationship means. This means that this matrix is the inverse of this matrix. This proves that ν is invertible. And in fact, we've done this just for m equals 2. But in fact, this proof is general and holds for arbitrary Markov chains if the eigenvectors span the space. And we'll see that later.

We're doing this for m equals 2 now, so we have to proceed when we have an arbitrary Markov chain. u is invertible. u^{-1} has π_1 and π_2 as rows. And thus P is going to be equal to-- I guess we should-- oh, we set it up here. P times u is equal to u times λ . We've shown here that u is invertible, therefore we can multiply this equation by u^{-1} . And we get the transition matrix P is equal to u times the diagonal matrix λ times the matrix u^{-1} .

What happens if we try to find P^2 ? Well, it's u times λ times u^{-1} times u times λ times u^{-1} . One of the nice things about matrices is you can multiply them, if you don't worry about the details, almost like numbers. Times u times λ times u^{-1} times u times λ times u^{-1} . Except you don't have commutativity. That's the only thing that you don't have. But anyway, you have u times λ times u^{-1} times u times λ times u^{-1} . This and this turn out to be the identity matrix, so you have u times λ times λ , which is λ^2 , times u^{-1} . You still have this diagonal matrix here, but the eigenvalues have all been doubled.

If you keep doing that repeatedly, you find out that P^n -- namely, this long-term transition matrix, which is the thing we're interested in-- is the matrix u times this diagonal matrix, λ^n , times u^{-1} . Equation 329 in the text has a typo, and it should be this. It's given as u^{-1} times λ^n times u , which is not at all right. That's probably the worst typo, because if you try to say something from that, you'll get very confused. You can solve one in general if all the M eigenvalues are distinct as easily as for M equals 2. This is still valid so long as the eigenvectors span the space.

So now the thing we want to do is relatively simple. This λ^n is a diagonal matrix. I can represent it as the sum of M different matrices. And each of those matrices has only one diagonal element, non-0. In other words, for the case here, what we're doing is taking λ_1^n , 0 to the n , 0 λ_2^n to the n , and representing this as λ_1^n , 0, 0, 0, plus 0, 0, 0 λ_2^n to the n .

So we have those trivial matrices with u on the left side and u^{-1} on the

right side. And we think of how to multiply the matrix u , which is a matrix whose columns are the eigenvectors, times this matrix with only one non-0 element, times the matrix here, whose elements are the left eigenvectors. And how do you do that? Well, if you do this for a while, and you think of what this one element here times a matrix whose rows are eigenvectors does, this non-0 term in here picks out the appropriate row here. And this non-0 element picks out the appropriate column here.

So what that gives you is p to the n is equal to the sum over the number of states in the Markov chain times λ sub i -- the i -th value to the n th power-- times ν to the i times π to the i . π to the i is the i -th eigenvector of p . ν to the i is the i -th right eigenvector of p . They have nothing to do with n . The only thing that n affects is this eigenvalue here. And what this is saying is that p to the n is just the sum of eigenvalues which are, if λ is bigger than 1, this is exploding. If λ is less than 1, it's going to 0. And if λ is equal to 1, it's staying constant. If λ is complex but has magnitude 1, then it's just gradually rotating around and not doing much of interest at all, but it's going away.

So that's what this equation means. It says that we've converted the problem of finding the n th power of p just to this problem of finding the n th power of these eigenvalues. So we've made some real progress.

AUDIENCE: Professor, what is ν_i right here?

PROFESSOR: What?

AUDIENCE: What is ν_i ?

PROFESSOR: ν_i is the i -th of the right eigenvectors of the matrix p .

AUDIENCE: And π_i ?

PROFESSOR: And π_i is the i -th left eigenvector. And what we've shown is that these are orthogonal to each other, orthonormal.

AUDIENCE: Can you please say again what happens when λ is complex?

PROFESSOR: What?

AUDIENCE: When λ is complex, what exactly happens?

PROFESSOR: Oh, if λ is complex and the magnitude is less than 1, it just dies away. If the magnitude is bigger than 1, it explodes, which will be very strange. And we'll see that can't happen. And if the magnitude is 1, as you take powers of a complex number of magnitude 1, I mean, it starts out here, it goes here, then here. I mean, it just rotates around in some crazy way. But it maintains its magnitude as being equal to 1 all the time.

So this is just repeating what we had before. These are the eigenvectors. If you calculate this very quickly using this and this, and if you recognize that the right eigenvector, ν_2 , is the first part of it is π_2 , the second part of it minus π_1 , where π_1 is just this first eigenvector here. So if you do this multiplication, you find that ν_1 -- oh, I thought I had all of these things out. This should be ν_1 .

The first right eigenvector times the first left eigenvector. Oh, but this is all right, because I'm saying the first left eigenvector is a steady state vector, which is the thing we're interested in. That's $\pi_1, \pi_2, \pi_1, \pi_2$, where π_1 is this and π_2 is this. ν_2 times π_2 is just this. So when we calculate P^n , we get π_1 plus π_2 times this eigenvalue to the n th power. π_1 minus π_1, λ_2 to the n th power. π_2 and π_2 is what we get for the main eigenvalue. This is what we get for the little eigenvalue. This little eigenvalue here is $1 - P_{12} - P_{21}$, which has magnitude less than 1, unless we either have the situation where P_{12} is equal to P_{21} is equal to 0, or both of them are 1.

So these are the terms that go to 0. This solution is exact. There were no approximations in here. Before, when we analyzed what happened to P to the n , we saw that we converged, but we didn't really see how fast we converged. Now we know how fast we converge. The rate of convergence is the value of this second eigenvalue here. And that's a pretty general result. You converged like the second-largest eigenvalue. And we'll see how that works out.

Now, let's go on to the case where you have an arbitrary number of states. We've almost solved that already, because as we were looking at the case with two states, we were doing most of the things in general. If you have an n state Markov chain, a determinant of $P - \lambda I$ is a polynomial of degree M in λ . That was what we said a while ago. It has M roots, eigenvalues. And here, we're going to assume that those roots are all distinct. So we don't have to worry about what happens with repeated roots. Each eigenvalue λ_i -- there are M of them now -- has a right eigenvector, ν_i , and a left eigenvector, π_i . And we have seen that -- well, we haven't seen it yet. We're going to show it in a second.

$\pi_i \nu_j$ is equal to δ_{ij} for each i, j unequal to i . If you scale either this or that, when you saw this eigenvector equation, you have a π on both sides or a ν on both sides, and you have a scale factor which can't be determined from the eigenvector equation. So you have to choose that scaling factor somehow. If we choose the scaling factor appropriately, we get π_i , the i -th left eigenvector, times the i -th right eigenvector. This is just a number now. It's that times that. We can scale things, so that's equal to 1.

Then as before, let U be the matrix with columns ν_1 to ν_M , and let V have the rows, π_1 to π_M . Because of this orthogonality relationship we've set up, VU is equal to I . So again, the left eigenvector rows forms a matrix which is the inverse of the right eigenvector columns. So that says V is equal to U^{-1} . Thus the eigenvector is ν , the first right eigenvector up to the n th right eigenvector, these are linearly independent. And they span M space.

That's a very peculiar thing we've done. We've said we have all these M right eigenvectors. We don't know anything about them, but what we do know is we also have M left eigenvectors. And the left eigenvectors, as we're going to show in just a second, are orthogonal to the right eigenvectors. And therefore, when we look at these two matrices, we can multiply them and get the identity matrix. And that means that the right eigenvectors have to be -- when we look at the matrix of the right eigenvectors, is non-singular.

Very, very peculiar argument. I mean, we find out that those right eigenvectors span the space, not by looking at the right eigenvectors, but by looking at how they relate to the left eigenvectors. But anyway, that's perfectly all right. And so long as we can show that we can satisfy this orthogonality condition, then in fact all this works out. v is equal to u to the minus 1. These eigenvectors are linearly independent and they span M space. Same here. And putting these equations together, P times u equals u times λ . This is exactly what we did before. Post-multiplying by u to the minus 1, we get P equals u times λ times u to the minus 1. P to the n is then u times λ to the n times u to the minus 1.

All this stuff about convergence is all revolving down to simply the question of what happens to these eigenvalues. I mean, there's a mess first, finding out what all these right eigenvectors are and what all these left eigenvectors are. But once you do that, P to the n is just looking at this quantity, breaking up λ to the n the way we did before. P to the n is just this sum here. Now, each row of P sums to 1, so e is a right eigenvector of eigenvalue 1. So we have a theorem that says the left eigenvector π of eigenvalue 1 is a steady state vector if it's normalized to π times e equals 1.

So we almost did that before, but now we want to be a little more careful about it. Oh, excuse me. The theorem is that the left eigenvector π is a steady state vector if it's normalized in this way. In other words, we know that there is a left eigenvector π , which has eigenvalue 1, because there's a right eigenvector. If there's a right eigenvector, there has to be a left eigenvector. What we don't know is that π actually has non-negative terms. So that's the thing we want to show.

The proof is, there must be a left eigenvector π for eigenvalue 1. We already know that. For every j , π_j is equal to the sum over k times π_k times p_{kj} . We don't know whether these are complex or real. We don't know whether they're positive or negative, if they're real. But we do know that since they satisfy this eigenvector equation, they satisfy this equation. If I take the magnitudes of all of these things, what do I get?

The magnitude on this side is $\sum_j |\pi_j|$ magnitude. This is less than or equal to the sum of the magnitudes of these terms. If you take two complex numbers and you add them up, you get something which, in magnitude, is less than or equal to the sum of the magnitudes. It might sound strange, but if you look in the complex plane - imaginary, real - and you look at one complex number, and you add it to another complex number, this distance here is less than or equal to this magnitude plus this magnitude. That's all that equation is saying.

And this is equal to this distance plus this distance if and only if each of these components of the eigenvector that we're talking about, if and only if those components are all heading off in the same direction in the complex plane. Now what do we do? Well, you look at this for a while and you say, OK, what happens if I sum this inequality over j ? Well, if I sum this over j , I get one. And therefore when I sum both sides over j , the sum over j of the magnitudes of these eigenvector components is less than or equal to the sum over k of the magnitude. This is the same as this. This j is just a dummy index of summation. This is a dummy index of summation.

Obviously, this is less than or equal to this. But what's interesting here is that this is equal to this. And the only way this can be equal to this is if every one of these things are satisfied with equality. If any one of these are satisfied with inequality, then when you add them all up, this will be satisfied with inequality also, which is impossible. So all of these are satisfied with equality, which says that the magnitude of $\sum_j \pi_j$, the vector whose elements are the magnitudes of this thing we started with, in fact form a steady state vector if we normalize them to 1.

It says these magnitudes satisfy the steady state equation. These magnitudes are real and they're positive. So when we normalize them to sum to 1, we have a steady state vector. And therefore the left eigenvector π of eigenvalue 1 is a steady state vector if it's normalized to $\pi \cdot e = 1$, which is the way we want to normalize them. So there always is a steady state vector for every finite-state Markov chain. So this is a non-negative vector satisfying a steady state vector equation. And normalizing it, we have a steady state vector. So we've demonstrated

the existence of a left eigenvector which is a steady state vector.

Another theorem is that every eigenvalue satisfies $|\lambda| < 1$, magnitude of the eigenvalue is less than or equal to 1. This, again, is sort of obvious, because if you have an eigenvalue which is bigger than 1 and you start taking powers of it, it starts marching off to infinity. Now, you might say, maybe something else is balancing that. But since you only have a finite number of these things, that sounds pretty weird. And in fact, it is. So the proof of this is, we want to assume that π_i is the i -th of these eigenvectors of P . Its eigenvalue is λ^i . It also is a left eigenvector of P to the n with eigenvalue λ^n . That's what we've shown before.

I mean, you can multiply this matrix P , and all you're doing is just taking powers of the eigenvalue. So if we start out with λ^n , let's forget about the i 's, because we're just looking at a fixed i now. λ^n times the j -th component of π_i is equal to the sum over i of the i -th component of π_i times P_{ij} to the n , for all j .

Now I take the magnitude of everything is before. The magnitude of this is, again, less than or equal to the magnitude of this. I want to let β be the largest of these quantities. And when I put that maximizing j in here, λ^n is less than or equal to the sum over i of β times P_{ij} to the n . I can upper-bound these by β . So I wind up with λ^n is less than or equal to the sum over i of β times P_{ij} to the n .

I don't know what these powers are, but they're certainly less than or equal to 1. So λ^n is less than or equal to n . That's what this said. When you take this magnitude of the i -th eigenvalue, it's less than or equal to this number n . Now, if this number were larger than 1, if it was 1 plus 10^{-6} , and you multiplied it by a large enough number n , that this would grow to be arbitrarily large. It can't grow to be arbitrarily large, therefore the magnitude of λ^n has to be less than or equal to 1. Tedious proof, but unfortunately, the notes just assume this.

Maybe I had some good, simple reason for it before. I don't have any now, so I

have to go through a proof. Anyway, these two theorems, if you look at them, are valid for all finite-state Markov chains. There was no place that we used the fact that we had anything with distinct eigenvalues or anything. But now when we had distinct eigenvalues, we have the n th power of P is the sum here again over right eigenvectors times left eigenvectors. When you take a right eigenvector, which is a column vector, times a left eigenvector, which is a row vector, you get an M by M matrix. I don't know what that matrix is, but it's a matrix. It's a fixed matrix independent of n . And the only thing that's varying with n is these eigenvalues.

These quantities are less than or equal to 1. So if the chain is an ergodic unit chain, we've already seen that one eigenvalue is 1, and the rest of the eigenvalues are strictly less than 1 in magnitude. We saw that by showing that for an ergodic unit chain, P to the n converged. So the rate at which P to the n approaches e times π_i is going to be determined by the second-largest eigenvalue in here. And that second-largest eigenvalue is going to be less than 1, strictly less than 1. We don't know what it is. Before, we knew this convergence here for an ergodic unit chain is exponential. Now we know that it's exponential and we know exactly how fast it goes, because the speed of convergence is just the second-largest eigenvalue.

If you want to know how fast P to the n approaches e times the steady state vector π_i , all you have to do is find that second-largest eigenvalue, and that tells you how fast the convergence is, except for calculating these things, which are just fixed. If P is a periodic unit chain with period d , then if you read the notes-- you should read the notes-- there are d eigenvalues equally spaced around the unit circle. P to the n doesn't converge. The only thing you can say here is, what happens if you look at P to the d -th power? And you can imagine what happens if you look at P to the d -th power without doing any analysis.

I mean, we know that what happens in a periodic chain is that you rotate from one set of states to another set of states to another set of states to another set of states, and then back to the set of states you started with. And you keep rotating around. Now, there are d sets of states going around here. What happens if I take P to the d ? P to the d is looking at the d -step transitions. So it's looking at, if you start here,

after d steps, you're back here again, after d steps, you're back here again. So the matrix, P to the d , is in fact the matrix of d ergodic subclasses.

And for each one of them, whatever subclass you start in, you stay in that subclass forever. So the analysis of a periodic unit chain, really the classy way to do it is to look at P to the d and see what happens there. And you see that you get convergence within each subclass, but you just keep rotating among subclasses. So there's nothing very fancy going on there. You just rotate from one subclass to another. And that's the way it is. And P to the n doesn't converge. But P to the d times n does converge.

Now, let's look at the next-most complicated state. Suppose we have M states and we have M independent eigenvectors. OK, remember I told you that there was a very ugly thing in linear algebra that said, when you had an eigenvalue of multiplicity k , you might not have k linearly independent eigenvectors. You might have a smaller number of them. We'll look at an example of that later. But here, I'm saying, let's forget about that case, because it's ugly. Let's assume that whatever multiplicity each of these eigenvalues has, if you have an eigenvalue with multiplicity k , then you have k linearly independent right eigenvectors and k linearly independent left eigenvectors to correspond to that.

And then when you add up all of the eigenvectors, you have M linearly independent eigenvectors. And what happens when you have M linearly independent vectors in a space of dimension M ? If you have M linearly independent vectors in a space of dimension N , you expand the whole space, which says that the vector of these eigenvectors is in fact non-singular, which says, again, we can do all of the stuff we did before. There's a little bit of a trick in showing that the left eigenvectors and the right eigenvectors can be made orthogonal. But aside from that, P to the n is again equal to the same form. And what this form says is, if all of the eigenvalues except one are less than 1, then you're again going to approach steady state.

What does that mean? Suppose I have more than one ergodic chain, more than one ergodic class, or suppose I have a periodic class or something else. Is it

possible to have one eigenvalue equal to 1 and all the other eigenvalues be smaller? If there's one eigenvalue that's equal to 1, according to this formula here, eventually P^n converges to that one value equal to 1. And right eigenvector can be taken as e .

Left eigenvector can be taken as a steady state vector π . And we have the case of convergence. Can you have convergence to all the rows being the same if you have multiple ergodic classes? No. If you have multiple ergodic classes and you start out in one class, you stay there. You can't get out of it. If you have a periodic class and you start out in that periodic class, you can't have convergence there.

So in this situation here, where all the eigenvalues are distinct, you can only have one eigenvalue equal to 1. Here, when we're going to this more general case, we might have more than one eigenvalue equal to 1. But if in fact we only have one eigenvalue equal to 1, and all the others are strictly smaller in magnitude, then in fact you're just talking about this case of an ergodic unit chain again. It's the only place you can be.

So let's look at an example of this. Suppose you have a Markov chain which has l ergodic sets of states. You have one set of states. So we have one set of states over here, which will all go back and forth to each other. Then another set of states over here. Let's let l equal 2 in this case. So what happens in this situation? We'll have to work quickly before it gets up. Anybody with any sense, faced with a Markov chain like this, would say if we start here, we're going to stay here, if we start here, we're going to stay here. Let's just analyze this first. And then after we're done analyzing this, we'll analyze this. And then we'll put the whole thing together.

And what we will find is a transition matrix which looks like this. And if you're here, you stay here. If you're here, you stay here. We can find the eigenvalues and eigenvectors of this. We can find the eigenvalues and eigenvectors of this. If you look at this crazy formula for finding determinants, what you're stuck with is permutations within here times permutations within here. So the eigenvalues that you wind up with are products of the two eigenvalues. Or any eigenvalue here is an

eigenvalue of the whole thing. Any eigenvalue here is an eigenvalue of the whole thing. And we just look at the sum of the number of eigenvalues here and the number there.

So we have a very boring case here. Each ergodic set has an eigenvalue equal to 1, has a right eigenvector equal to 1. When the steps of that state and 0 elsewhere. There's also a steady state vector on that set of states. We've already seen that. So P to the n converges to a block diagonal matrix, where for each ergodic set, the rows within that set are the same. So P to the n then is π_1, π_1 . And then here, we have π_2, π_2, π_2 . So that's all that can happen here. This is limit.

So one message of this is that, after you understand ergodic unit chains, you understand almost everything. You still have to worry about periodic unit chains. But you just take a power of them, and then you have ergodic sets of states. one final thing. Good, I have five minutes to talk about this. I don't want any more time to talk about it, because I'll get terribly confused if I do. And it's a topic which, if you want to read more about it, read about it in Strang. He obviously doesn't like the topic either. Nobody likes the topic. Strang at least was driven to say something clear about it. Most people don't even bother to say something clear about it.

There's a theorem, due to, I guess, Jordan, because it's called a Jordan form. And what Jordan said is, in the nice cases we talked about, you have this decomposition of the transition matrix in P into a matrix here whose columns are the right eigenvectors times a matrix here, which is a diagonal matrix with the eigenvalues along it. And this, finally, is a matrix which is the inverse of this, and, which properly normalized, is the left eigenvectors of P . And you can replace this form by what's called a Jordan form, where P is equal to some matrix u times the Jordan form matrix J times the inverse of u .

Now, u is no longer the right eigenvectors. It can't be the right eigenvectors, because when we needed Jordan form, we don't have enough right eigenvectors to span the space. So it has to be something else. And like everyone else, we say, I don't care what that matrix is. Jordan proved that there is such a matrix, and that's

all we want to know. The important thing is that this matrix J in here is as close as you can get it. It's a matrix, which along the main diagonal, has all the eigenvalues with their appropriate multiplicity. Namely, λ_1 is an eigenvalue with multiplicity 2. λ_2 is an eigenvalue of multiplicity 3. And in this situation, you have two eigenvectors here, so nothing appears up there.

With this multiplicity 3 eigenvalue, there are only two linearly independent eigenvectors. And therefore Jordan says, why don't we stick a 1 in here and then solve everything else? And his theorem says, if you do that, it in fact works. So every time-- well, the eigenvalue is on the main diagonal, the ones on the next diagonal up, the only place would be anything non-0 is on the main diagonal in this form, and on the next diagonal up, where you occasionally have a 1. And the 1 is to replace the need for deficient eigenvectors. So every time you have a deficient eigenvector, you have some 1 appearing there. And then there's a way to solve for u . And I don't have any idea what it is, and I don't care.

But if you get interested in it, I think that's wonderful. But please don't tell me about it. Nice example of this is this matrix here. What happens if you try to take the determinant of $P - \lambda I$? Well, you have $1/2 - \lambda$, $1/2 - \lambda$, $1 - \lambda$. What are all the permutations here that you can take? There's the permutation of the main diagonal itself. If I try to include that element, there's nothing I can do but have some element down here. And all these elements are 0. So those elements don't contribute to a determinant at all. So I have one eigenvalue which is equal to 1. I have two values at multiplicity 2, eigenvalue which is $1/2$.

If you try to find the eigenvector here, you find there is only one. So in fact, this corresponds to a Jordan form, where you have $1/2$, 1, and a 0, and a 1 here, and 0 everywhere else. And now if I want to find P^{-n} , I have u times this J times u to the minus 1 times u . All the u 's in the middle cancel out, so I wind up eventually with u times J to the n th power times u to the minus 1.

What is J to the n th power? What happens if I multiply this matrix by itself n times?

Well, it turns out that what happens is that this main diagonal here, you wind up with a $1/4$ and then $1/8$ and so forth. This term here, it goes down exponential. Well, if you multiply this by itself, eventually, you can see what's going on here more easily if you draw the Markov chain for it. You have state 1, state 2, and state 3. State 1, there's a transition $1/2$ and a transition $1/2$. State 2, there's a transition $1/2$ and a transition $1/2$. And state 3, you just stay there.

So the amount of time that it takes you to get to steady state is the amount of time it takes you-- you start in state 1. You've got to make this transition eventually, and then you've got to make this transition eventually. And the amount of time that it takes you to do that is the sum of the amount of time it takes you to go there, plus the amount of time that it takes to go there. So you have two random variables. One is the time to go here. The other is the time to go here. Both of those are geometrically decreasing random variables. When we convolve those things with each other, what we get is an extra term n . So we get an n times $1/2$ to the n .

So the thing which is different in the Jordan form is, instead of having an eigenvalue to the n th power, you have an eigenvalue times-- if there's only a single one there, there's an n there. If there are two 1s both together, you get an n times n minus 1, and so forth. So worst case, you've got a polynomial to the n th power times an eigenvalue. For all practical purposes, this is still the eigenvalue going down exponentially. So for all practical purposes, what you wind up with is the second-largest eigenvalue still determines how fast you get convergence. Sorry, I took eight minutes talking about the Jordan form. I wanted to take five minutes talking about it. You can read more about it in the notes.