[SQUEAKING]

[RUSTLING]

[CLICKING]

**LAWRENCE GUTH:** Yeah. So we're sort of in the last part of the class now. And I've been reflecting on what to do together. And I asked everybody for some input, a little survey about what topics to do and how much detail to do. And there was no consensus in the survey of what people wanted to do. But I decided I would like to do, for a couple of weeks, talk about some applications of projection theory in other areas and then, the last couple of weeks, try to talk about at least some of the main ideas in the proof of the projection theorem.

Yeah.

- So we recently had been doing Bourgain's projection theorem and the Bourgain-Katz-Tao projection theorem in finite fields-- and so feeling that we really had to work pretty hard for those things. And it was striking that we had to work pretty hard just to get this very small improvement. And so in some of what we'll do in the rest of the class, I will try to show that that small improvement is really valuable.

OK. So yeah. So the plan for the rest of the class is, for about one week, we'll talk about random walks on groups. And then for about one week, we'll talk about homogeneous dynamics. And then at the end, we'll talk about the sharp projection theorem. OK. Cool. And also, in the last few classes, I felt that we were doing some very technical things and sometimes a little bit of a struggle.

So for today, what I want to do is to give, I hope, a very friendly introduction to random walks on groups and then next week, on Tuesday, to explain how projection theory helps to figure out some interesting things. OK. So I'll talk about. random walks on groups.

All right. So suppose that G is a finite group. And then mu is a probability measure on G. So I'll think of it as a function on G which is non-negative and sums to 1.

OK. So if you have a probability measure like this, you can do a random walk where, at each step of the random walk, you pick an element of G by this probability, and then you move by that element. So here's how the random walk works. So you start somewhere, say at g0. Then you sample a step h from mu. And the next guy, g1, is g0 times h.

So that's one step of the random walk. And then you can keep doing this. OK. And so if you do take k steps of the random walk, then there will be some probability distribution of where you end up after k steps. And one main question about random walks is, how evenly distributed is that probability distribution?

So question-- how evenly distributed is the random walk after k steps? And we'll make this more precise in a moment.

All right. So it is helpful to describe this random walk using convolution. OK. So convolution on a group looks like this. We have two functions, f1 and f2, on this group.

And then if I convolve them with each other and I evaluate that at g, that's the sum g1 times g2 equals g f of g1, f1 of g1, f2 of g2. That's the sum over all the g1 and g2 that multiply to g of this. OK. And then for mu, we can make an operator. And T mu of f is just f convolved with mu.

And this operator basically describes this random walk. So let's say how they're related. So suppose I were to take T mu and apply it to the delta function at g0 and evaluate that at some g. OK. So let's see. What's a good way to write this?

So that's the sum of-- actually, let me look at how I wrote it in my notes. I tried writing this a few ways, and then I decided there was a good way to write it. Yeah. Let's write it this way. So suppose I want to evaluate this at g0 times h. So that's the sum over g1 times g2 equals g0 times h of delta function at g0 of g1 times mu of g2.

All right. So the only way that this sum can be not 0 is if g1 is g0. So there's only one term in this sum. That's this g0 times g2 is g0 times h of mu of g2. There's only one term, and it's the term where g2 is h.

OK. So this thing is the probability distribution of one step of the random walk. The probability if I started at g0 that I end up at g0 times h is mu of h. That's the definition of my random walk. So this operator is performing one step of the random walk. And therefore, if I take T mu to the k of delta g0, that's like k steps of the random walk, starting at g0.

OK. Actually, let me say this-- I could say this better. So this is the probability that our random walk is at g after k steps.

OK. Now we can take this question, how evenly distributed is the random walk? We can write that as a precise question about this. I'll call it question 1. So I'd like to estimate T mu to the k of delta g0. So this is the actual probability distribution. And if the probability was completely even over the whole group, it would be 1 over the size of the group everywhere. So I'm going to see how close it is to that by measuring the norm of this difference.

So I'd like to estimate that. OK. And L2 is helpful but also perhaps a little arbitrary. You could also put other norms, so could put or L infinity or maybe some other norm. OK. So this is the kind of question about random walks on groups that we are going to try to understand.

All right. So to understand this, it is helpful to notice that this is a linear operator. This T mu is a linear operator. It's basically a matrix. And it's helpful to think about that in terms of matrix ideas, like the singular values and singular vectors of this matrix.

So T mu is a linear operator. And we're going to study the singular values and singular vectors of T. OK. So there is one simple singular value or eigenvalue, I guess, that T mu of 1 is 1. Yeah?

**AUDIENCE:** Are singular values and single vectors just the same as eigenvalues and eigenvectors [INAUDIBLE]?

**LAWRENCE GUTH:** They are closely related. So let me do a quick review of singular values and eigenvalues. Right. If you have a matrix and you think about what it does to the unit ball, it will turn the unit ball into an ellipsoid. And the principal axes of this ellipsoid are the singular values, sigma 1, sigma 2.

OK. So if we look at the preimages of these things, those will be the singular vectors-- well, depending on the matrix, they could be pointing any which way. But maybe that's v2 and that is v1.

OK. And they will always be-- so we'll have-- so the Vi, or orthonormal, and the size of MVi is sigma i. And the MVi are also orthonormal also orthogonal. Those are the singular vectors.

OK. Now, they are not generally eigenvectors, in the sense that MVi will not usually be a multiple of Vi. But they are closely related to them. The Vi are the eigenvectors of M M star-- or M star M, I guess. And the sigma i squared are the eigenvalues of M star M equal to lamda i M M star.

So in particular, this biggest singular vector measures the operator norm of the matrix, which is the-- so these are ordered like this. And Mv is at most sigma 1 v.

OK. So the reason we're talking about singular values is, we're going to care about stuff like this. And this is naturally associated with singular values if M is not symmetric. It's not necessarily the same as talking about eigenvalues.

OK. So if we take this T mu of the function 1, the constant function, we will just get 1. Because this T mu is like an averaging operator. We're averaging T-- so T mu of f at g is the average value of g at some other places around g that are dictated by mu. So T mu of 1 is 1.

So that implies that the-- OK. So I'll do a little lemma. The largest singular value-- yeah, OK. So little lemma is that. T mu of f in L2 is at most f in L2. And this is sharp because of this example. It could be equal.

OK. So proving this will help us think a little bit about what convolution means. So T mu of f is f convolved with mu-- g here. So that's the sum g1 g2 equals g f of g1 mu of g2. And I'd like to think of this as a sum over g2. So it's the sum over g2. Because I'm thinking of it that way, I can bring the mu of g2 out front. And then what I have left is f of g1, which is g times g2 inverse.

And I want to think of this as a linear combination of a bunch of different functions of g. OK. Let's give them a name. So this function of g is sort of f. But we first did this thing. And I want you to think of it as taking f and then shifting it using the group. So it's the right translation by g2 of f of g is defined to be this thing here. So we could plug that in.

And this shifting operation doesn't change the norm of g. So Rg2 of f in L2 is the same as f in L2. Ah, I've been mentioning L2. Yeah, I should say-- OK. Let me pause for a second. I should have said something before.

So when we have these functions on g, we're going to measure their norm. And if I have f L2 of g, this is going to be sum g of g? F of g squared to the 1/2. Or I could take inner product. The inner product of f1 and f2 is the sum g of G f1 of g, f2 of g to that. So the functions on G are a Hilbert space with this natural measure.

OK. So then if I take a function and I translate it or do this right translation, that just permutes the value so it doesn't change this thing. OK. So now if we look at our T mu of f, I have this-- yeah.

So actually, if I'm interested in the norm of T mu of f, this is the norm of the sum on g2 mu of g2 R sub g2 of f. It's a linear combination of these functions, which are all translates of f. And L2 is a norm. It obeys the triangle inequality.

So this is less than the sum on g2, mu of g2, Rg2 of f in L2. All of these have the same norm, which is the L2 norm of f And then this adds up to 1. So this is just the L2 norm of f.

OK. So this is a basic fact about averaging operators. And to understand whether the random walk is mixing things up, we have to care about how it behaves on the space orthogonal to 1. And the first observation is that it preserves the space orthogonal to 1. So let's give that space a name, say L2 of G 0 is the set of f and L2 of g so that some g in G of f of g is 0.

This is saying that f is orthogonal to the constant function. OK. So another simple lemma is that our operator T mu actually takes this space to itself. So it takes the constant function to itself, and it takes the perpendicular space to itself, which will imply that the constant function is one of the singular vectors of this operator.

OK, so proof-- so I want to understand the sum g in G T mu f of g. I'm going to imagine that the sum of f of g is 0, and I want to see if that's still true. All right. So that's the sum g in G. Sum g1 times g2 equals g, f of g1 mu of g2. And if you think about what we're summing over, we're just summing over g1 and g2.

So this is the sum over g1 and g2, of f of g1, mu of G2. And now that factors. So this is 1. And if f was in L2 of G 0, this would be 0. This is 0. So if f is in L2 of G 0, then so T mu of f will be also.

OK, cool. All right. So we've seen that T mu of 1 is 1. And T mu takes L2 of G 0 to itself.

So this is the largest singular value. So we've seen it. And then here, there's something else. So let's say sigma 1 of mu-- so this is 1. I'll call it sigma 0 of T mu. Somehow, in this setting, the 1 is trivial. And everything else will depend on the trace of mu. So sigma 1 of T mu is the largest singular value of this part.

OK. All right. So question 2 is going to be to estimate sigma 1 of t. And these two questions are related to each other. This thing helps to understand how-- it helps a lot to understand how this operator behaves and to answer question 1. So let me put up a simple proposition, how question 2 is related to question 1.

So proposition-- if you look at that thing, T mu to the k delta function of g0 minus 1 over G in L2, this is bounded by sigma 1 of T mu to the k. OK, so proof-- so we want to see what T mu to the k does to a delta function at a point.

Let's take this delta function at a point. And T mu splits up. The constant function is an eigenfunction and then the things with mean 0. So let's split up this delta function. So the constant part of it is like this. And then we have this.

So the reason we split it up that way is that this part, which in the beginning of the class, we would have called the high part of this-- so I guess I'll do that. That's in L2 of G 0. The 0 part. OK. Now what happens when we apply this operator? Well, this operator takes constants to constants. It's the identity on the constant part.

And so if I take T mu to the k delta g0, this part will stay 1 over G inside. I'll take the left-hand side. And then if I subtract that, it will disappear. So what I'll be left with is T mu to the k on this high part of the delta g0.

OK. How big is that? Well, this part lives in this space. So we can bound the size of this thing using the singular vector. So we get T mu to the k of delta g0 high to-- that's bounded by sigma 1 of T mu to k times delta g0 high L2.

OK. And now it's also now straightforward to check that this is smaller than 1. OK. So it's an application. If we were able to arrange that sigma 1 of T mu to the k was less than 1 over 100 times the size of G, then, well, this L2 norm would be less than 1 over 100 times the size of G.

And so in particular, the L infinity norm would be less than 1 over 100 times the size of G. So then I would say that for every G, T mu k delta g0 of g minus 1 over the size of g would be smaller than 1 over 100 times the size of g. That would tell us, say, that this function is pretty close to this constant function, within 1%. So that random walk would now have mixed pretty thoroughly.

OK, great. So if we can get our hands on sigma 1 of T mu, then we can figure out how long it takes this random walk to mix evenly. Cool. OK. So now let me tell you an interesting example of a particular group and a particular mu where there's a cool estimate for this.

So we are going to mostly focus-- or maybe only focus-- on the group SL2 of Fp-- and could pick many different measures for a random walk. But one nice one which has been studied is to take the set A to be-- so, first of all, if A is a subset of g, there's a natural measure on A, which is the uniform measure, so measure on A, which is 1 over the cardinality of A times the sum a in A of delta A.

OK. So there's a nice set, A, inside of here, which was studied by Selberg, A. Selberg. Well, OK, which is relevant to a theorem of Selberg. He set his theorem differently. OK. So there are four elements in this set. And these are our generators of the group SL2 Fp. There are many ways you could choose generators for this group, but this is probably the simplest way.

OK. And now I can state a theorem. So theorem 1 exists a positive c-- I think 1 over 100 should be fine-- so that for every p prime, if you take A Selberg and SL2 of Fp. You make a random walk. So in this setting, let me write TA for T sub mu sub A.

Then sigma 1 of TA is less than 1 minus c, so uniformly for all primes. So this is, say, 99/100. Then you can just read off high powers of this or bounded by something. And you can plug that into the discussion over here. So it says that that random walk mixes extremely fast. Yeah?

**AUDIENCE:** How does sigma 1 behave as you iterate over the primes? Does it converge to sub 1 minus c or oscillate a lot ?

**LAWRENCE GUTH:** So the question is, how does sigma 1 of to behave as the prime varies? I do not know the answer to that question. So let me say that this is a corollary of a theorem of Selberg from the late '50s.

And Selberg's actual theorem is about the first eigenvalue of the Laplacian on a hyperbolic surface, which is associated to this group. And you can go back and forth. With modern technology, it's considered not very difficult to go back and forth between a bound for the eigenvalue of the surface and a bound here. But when you do go back and forth, you will screw up the constants.

Selberg's theorem was that the first eigenvalue of that surface was at least 3/16. And that's probably not the right answer. And he conjectured that it should be a quarter. So for his actual problem, there is a beautiful conjecture about what happens as p goes to infinity, which is that the eigenvalues should converge to a quarter. For this, I'm not sure.

OK, cool. Cool. OK. So bounds like this are also related to expander graphs, isoperimetric inequalities in graphs. And I thought I would also show that. And so it shows that-- well, OK, so let me explain that. So if you have a set of generators inside of a group, we can use that to make a graph called a Cayley graph.

And this mixing inequality can be thought of or implies some interesting geometric features of the graph.

So expansion in graphs-- OK. So if G is a finite group And A is a set of generators-- and i guess that to have a graph, maybe we should say that is a symmetric set of generators. I don't know if that's important. Yeah, let's do it. Then we'll have a Cayley graph, Cayley graph of G, A-- so is a graph.

The vertices are g in G. And a pair, g1, g2, is an edge if and only if g2 is g1 times a or vice versa. Yeah, so I guess we should have it symmetric. g2 equals g1 times a, where a is in A. OK. And the random walk we've been imagining is basically the random walk on this graph. You start at a vertex, and you move randomly to a neighboring along the edges. Yeah?

**AUDIENCE:** How do you identify SL2 of a finite field with a hyperbolic surface?

**LAWRENCE GUTH:** Yeah, I will-- OK. We'll talk about the connection to hyperbolic geometry a little later when we talk about the proof of this theorem. Let's do it then. Yeah. OK. So that's the Cayley graph.

And in this graph, if I have subsets S and T inside of the vertices, E of S and T, the edges from S to T, that's going to be the set of g1, g2 in S cross T so that g1, g2 is an edge-- so the set of edges from S to T.

OK. So proposition-- if we have C G, A, as above, then the number of edges from S to S complement-- so S is going to be a subset of the vertices. The number of these edges is at least 1 minus sigma 1 of t A times A times S times S complement over G.

OK. So let's digest this a little bit. So what does this expression mean? Well, this is how many points there are in the set, S. From each of those points, there are A edges coming out. So this is how many edges there are that are leaving points of S. How many of them do we think end up in S complement? Well, we don't really know. But as a reference point, if everything was random, the fraction of edges that ended up in S complement might be S complement over G.

So this expression here would be the typical number of edges from S to S complement if we had a random graph where each vertex was in this many edges. OK. And then we're going to say that the actual number of edges is at least some fraction of that. If sigma 1 of TA were 1, this would be a trivial inequality, which is bigger than 0. But if sigma 1 is smaller than 1, then this is a nontrivial inequality. And it says that the number of edges is a definite fraction of what you would see in the random model.

OK. So let's prove this or at least give a proof sketch. So the key point is that we can describe this quantity using this operator TA. And then it becomes not so surprising that the singular value of TA is relevant. How do you do it? The number of edges from S to S complement is the inner product, is the inner product of TA of the characteristic function of S with the characteristic function of S complement.

And there's a factor of A in front. This is a normalization if TA had been just-- TA is an averaging operator. But we didn't really want to average, we wanted to count edges. So there's an extra factor. OK. Does this look plausible to people? Should we write this? Let's write this out a little bit.

So TA 1S of g is 1 over A sum of a in A of 1S of g a inverse. So that's the definition of the convolution.

So then when we take this inner product, we're going to have the sum g in G 1 over A sum a in A 1S of ga inverse 1S complement of g. So we'll be counting g's that are in S complement but g a inverses in S. And that's an edge that starts in S and goes to S complement.

So S, S complement. Over here, we have g. Over here, we have g a inverse. And those are connected by an edge. That's the thing that we're counting. OK. This double sum here means going through all the edges. But we're only counting an edge if it goes from S to S complement. And we have this, which we didn't really want, which is why we multiply by this over here.

OK. Great. OK. So now remember that this operator sort of splits. It takes the constant function to the constant functions, and it takes the mean 0 functions to the mean 0 functions. So we should split everybody in sight into a constant part and a mean 0 part. So we split. 1S is S over G-- that's the constant part-- plus 1S minus S over G. So this-- so I'll call it 1S high. And it lives in L2 of G0.

OK. Do the same thing for S complement. So now the number of edges from S to S complement, that is A times-- OK. We'll have TA of the constant parts. So that's S, G, S complement G plus A times the high-frequency parts, the mean 0 parts TA of 1S high 1S complement high.

OK. So let's call this 0. And this is the high part. Or let's call this the constant part, and this is the high part. So the rest of it is just a computation. And I won't do the algebra carefully. But if you do the constant part, you get this A S S complement over G. The constant part exactly tracks what would happen if all the edges were random.

And then for the high part-- of course, we won't compute it exactly because we have to know what the operator TA is. But we can bound it by the norm of TA, the sigma 1 of TA, times 1S high in L2 times 1S complement high in L2.

OK. And then those things are straightforward to compute, those L2 norms. And if you compute them, you get that this is A-- I'll put the sigma 1 out front. And then it's the same thing-- A S S complement over G. OK. So the true answer is the constant term plus the high term, which is like an error term, and whose size is at most this. And if sigma 1 is strictly less than 1, then the error term is strictly smaller than the constant term. So something will be left over. And that's the inequality.

OK. So let me make a little remark. If S is smaller than G over 2, then S complement is bigger than G over 2. So this is more or less without loss of generality if you switch the roles of S and S complement. But it means S complement is at least G over 2, which means that this whole fraction has order of 1. So then we get that the number of edges from S to S complement is at least 1/2 sigma 1 of TA A times S.

So it's at least a definite fraction of all of the possible edges that it could be leaving us. OK. So this property here is called being an expander graph. Because if you take a set that's less than half of the graph and then you add to it all of the edges that are leaving it, then it will get significantly bigger. It will expand. So it's called an expander graph.

OK. Cool. So this proposition shows that if this sigma 1 is strictly less than 1-- it's something like in Selberg's theorem-- then the edges in the graph behave in this way, that it's an expander graph. OK, cool. OK. So Selberg did not state his theorem in this way. So this is slightly ahistorical.

But Selberg's theorem plus things that are not very difficult in hindsight show that this Cayley graph of SL2 Fp is an expander graph. And it is in a sense the first-- Selberg didn't quite say this, but in a sense, this was the first proof that any graph was an expander graph. Expander graphs were officially defined in 1970 by Pinsker and something pretty close by Kolmogorov and Barzdin in the '60s. And all of those people realized that random graphs were expander graphs.

Yeah. So maybe many people have seen, but it's really a remarkable thing. So just for comparison, like a simple graph that we might think of that we can easily visualize would be an n-by-n grid. And that kind of graph is not an expander at all. So an n-by-n grid has n squared vertices. But you can cut it in half with a vertical line, and you only cut n edges. So the number of edges-- so S and S complement are each about half of the graph. The number of edges between them is much smaller than S.

So these are very large graphs. But the number of edges it takes to cut it in half is comparable to all of the edges. And most graphs are like that. But before people realized it, it was surprising and counterintuitive that any graphs were like that. And it's a very significant fact in many domains of math and science.

OK. Let me pause there. And questions about Selberg's theorem and this setup, mixing and expansion and--

**AUDIENCE:** So the point of A being generators is just so that it's connected?

**LAWRENCE GUTH:** Yes, that's right. So the question is, what is the point of A being generators? So if A did not generate the group, then this graph would be disconnected, and then it wouldn't mix at all. So make a remark-- suppose that A is in a finite group and that H-- so the subgroup generated by A is H, which is a proper subgroup.

OK. So then if we take TA and we apply it to the characteristic function of this subgroup, since all of these guys are in the subgroup, we would just stay in the subgroup. So this would be 1H. OK. And that implies, with a teeny bit of work, that sigma 1 of TA would be 1.

OK. One thing we could see is that no matter how many times we repeat this operation, we will never mix over the whole group because we'll never get out of H. And so based on what we've seen before, this should be 1. Yeah?

**AUDIENCE:** Does it make A big-- does it just make the graph more likely to be expander?

**LAWRENCE GUTH:** Yeah. So the question is, if we make A big, is the graph more likely to be an expander? Perhaps. So we could take A big, and then it would-- so we're allowed to take A big. And we will talk in a little bit about what we can say. If we just know that A is big, what does that tell us about sigma of TA? We'll talk about that.

All right. So another remark I wanted to make is that the first proof of this theorem was very difficult. So the first proof depended on-- so it had some interesting analysis related to hyperbolic geometry. But then it depended on the Riemann hypothesis for curves over finite fields. So that's a beautiful and important theorem. But if you wanted to include the whole proof of that, then the proof of this theorem is effectively several hundred pages long and has a lot of algebraic geometry.

And in around the early 1990s, Sarnak and Xue gave a much simpler proof, which is the proof that we'll talk about. OK. So in the rest of the class, I will start to describe the proof by Sarnak and Xue and introduce the first key idea.

All right. OK. So the first idea, this idea has to do with the representation theory of the group SL2 Fp. So proposition-- OK. So if we have a representation-- so a representation is a group homomorphism to-- it's a unitary representation-- so to the unitary group in d dimensions.

So this is a representation, meaning a group homomorphism. OK. And so there's the trivial representation, where everything goes to the identity. But let me say it is a nontrivial representation. Then actually, d needs to be quite big. d is at least p plus 1 over 2.

OK. So before we prove this theorem, let's try to put it in context by trying to think about group representations. Where would we find a group representation? OK. Well, SL2 Fp acts on itself. So it acts on the space that we started with L2 of G. So any group G acts on L2 of G by the action that we wrote down. You take a function, and you shift the space under it.

OK. That action is not trivial. Only the constant functions are fixed. So that's a nontrivial action. And the dimension of this space is the cardinality of G. So for this group, that would be about p cubed. So cardinality of SL2 Fp sub f p cubed.

OK. What else? Well, if you have a subgroup of G-- H is a subgroup of G-- then G acts on G mod H. And so it acts on this as a set. And so it acts on L2 of G mod H by a unitary representation.

And so the dimension of L2 of G mod H is the cardinality of G mod H, which is the cardinality of G divided by the cardinality of H. OK. So let's think of some subgroups of SL2 Fp.

I think of-- the diagonal subgroup looks like this. And the size of the diagonal subgroup is around p. There's the unipotent subgroup, looks like this. Size of the unipotent subgroup is p. And you could mix these together. So there's the upper triangular subgroup. But we already used the letter u. So I'm going to call this B for Borel.

So I'll write it a a inverse t. That's a subgroup. And the size of the upper triangular subgroup is around p squared. So that's the biggest one that we have come up with here. And if you plug this in here-- so if you take SL2 Fp modulo rho, that's around p. And so this is the right order of magnitude.

OK. And this is an interesting fact about this group, which is extremely different from abelian groups. So for an abelian group, every irreducible representation is one-dimensional. So there are lots and lots of representations where d is 1. But for SL2 Fp. Life is very different. All of them have very large dimension.

OK. So let's prove this theorem. So I had heard about this theorem years ago. And I always assumed that to prove this theorem, one had to study a lot of representation theory. And I was a little intimidated by it. But then I read the proof as I was getting ready to teach this class. And actually, it doesn't require anything that's outside of a good undergraduate algebra curriculum. And it's quite elegant.

All right. So here's the proof. So we're going to focus on this matrix for a little while and try to figure out what the representation does with this matrix. So let's suppose for now that rho of this matrix is not just the identity.

Actually, let me say it in a different order. So u is that matrix and v is this matrix, which is very similar to u. And we'll make the remark that-- so exercise-- at this point, I won't include the details. u and v generate SL2 of Fp. OK. If you play around a little bit, you can see this.

All right. So therefore, if we have a nontrivial representation, these cannot both go to the identity. So rho is nontrivial. So either rho of u or row of v is not the identity. And u and v are so similar to each other, it doesn't really matter which one. So I'm going to say, essentially without loss of generality, rho of u is not i lamda

All right. Now let's think about rho of u. There's something quite special about this matrix, which is that it is conjugate to powers of itself. Let me write that down.

So suppose take a diagonal matrix, a, a inverse. And I put this matrix in the middle. And then I want to conjugate it. So now I put on the other side the inverse. So this is the conjugate of my matrix u. OK. So I multiply this by this. I multiply the top row by a and the bottom row by a inverse. So I get a a 0 a inverse. And then I still have this guy.

Now, when I multiply these matrices together, I multiply the left column by a inverse and the right column by a. The diagonals cancel. But I get 1 0 1 a squared.

OK, so see that u is conjugate tp u to the power a squared. And this is for any a in F we saw.

So that is a special and really interesting feature of this element of my group. And so now I have a representation. And rho of u is conjugate to rho of u of a squared, which is rho of u to the a squared. So now I have a matrix. I have a unitary matrix. And it is conjugate to a big power of itself.

OK. In the world of matrices, it's not so easy to get that to happen. So this group has an interesting property that has an element which is conjugate to many different powers of itself. And in matrices, it's not so easy to find a matrix which is conjugate to many different powers of itself. Why? Because of how the eigenvalues behave. OK. So now let's think about eigenvalues.

OK. So rho of u is in U of d is a matrix. And if rho of u is conjugate to rho of u to the a squared, then that tells us that the eigenvalues, the spectrum of rho of u is the same as the eigenvalues of rho of u to the a squared.

Two conjugate matrices have the same eigenvalues. But if you take a matrix and raise it to a power, the eigenvalues are all raised to a power. So this is the eigenvalues of rho of u to the a squared. OK. So for instance, you could take a to be 2. And it tells if you have an eigenvalue of rho of u, you raise it to the fourth power, that will also be an eigenvalue of rho of u. And you can keep doing that. And it's not just 4. We have many choices for a squared.

OK. So let's think about what the eigenvalues of rho of u may be like-- so eigenvalues of rho of u. So this is an amazing fact about them. And there's a more pedestrian fact that we should also observe. We know that u to the p is the identity. And therefore, rho of u to the p is the identity. So then if lambda is an eigenvalue of rho of u, then we get that lambda to the p is 1.

OK. So the eigenvalues of rho of u, they're all p-th roots of unity. So we get the following. If lambda is an eigenvalue of rho of u, then lambda to the p is 1 and lambda to the a squared is also an eigenvalue of rho of u.

OK. So since lambda to the p is 1, we only care about a squared mod p. So we could put mod p there. And there are p minus 1 over 2 quadratic residues mod p. So this a squared mod p, it could be any quadratic residue mod p. There are p minus 1 over 2 of them.

And if you take a p-th root of unity and you raise it to a power that's between 1 and p, you'll change it. So lambda is a primitive. So if lambda is not 1, then lambda is a primitive p-th root of unity. And that tells us that the lambda to the a squared mod p are all distinct for the p minus 1 over 2 quadratic residues.

OK. So we now have seen that there are an awful lot of different eigenvalues, which can only happen if the matrix is in a large dimension. So that gives us that d is at least p plus 1 over 2. OK. So how do we know that rho of you has an eigenvalue that's not 1?

Well, if all the eigenvalues were 1, then rho of u is conjugate to a matrix with 1's on the diagonals and some upper triangular stuff. And if you raise that to a power, you'll never get the identity. There's an exercise. So there's a little something more to

exercise-- if the eigenvalues of rho of u are exactly the set 1 and rho of u to the p equals 1, then rho of u is the identity. I'll leave the matrix details. Cool.

OK. So all the representations of SL2 are really big. Why does that matter? Let's do another proposition. So proposition 2-- you take sigma 1 of T mu.

So mu is a probability measure on SL2 Fp.

Then sigma 1 of this guy, if you square it and you multiply by d plus 1 over 2, that is at most the size of SL2 Fp times mu L2 squared. Simplifying the algebra a little bit, sigma 1 of T mu squared is smaller than around 1 over p squared times mu L2 squared.

OK. All right. So we come back to the question from earlier in the class. How are singular values related to eigenvalues? So sigma i of T mu squared is lambda i of T mu T mu star, i-th eigenvalue of this matrix. And the key observation is that the eigenspaces of this matrix have a G action on them. And except for the constant functions, the eigenspaces are not trivial.

There's the key. Eigenspaces of T mu T mu star have the G action. And except for the constant functions-- so the constant functions are an eigenspace of T mu, T mu star. But all the other ones, it's a nontrivial action.

Let's suppose that for a moment and see why it's helpful. And then we can think more carefully about what exactly are the eigenspaces. All right. OK. So that tells us that each sigma i has a multiplicity which is at least d plus 1 over 2.

And we really care about the biggest singular vector. There's not just one biggest singular vector, but there's a whole singular subspace of a big dimension. And all of those vectors have this large singular value. OK. So now another fact from linear algebra is that if you add up the singular values squared of the matrix with multiplicity, that is the Hilbert-Schmidt norm of the matrix. So it's the sum of the entries of the matrix.

So this is a matrix where the rows and columns are indexed by our group, g. So it has an entry, g1, g2. And it's the sum of the entries squared. Does this look familiar to people from linear algebra?

OK. Now, the entries of our matrix, we can read off. It's the sum g1 g2 mu of something like g1 g2 inverse. This might not be exactly right, but it's the right idea. That's the entry. And then we're squaring it. And so that's the size of the group times the sum of mu of g squared.

OK. So that's the right-hand side. And over here, we're counting with multiplicity. And the multiplicity is always large. So this is at least d plus 1 over 2 times sigma 1 of T mu squared.

So there's an important difference between the sigma 1 and sigma 0. The sigma 0 went with the constant functions. That lives in a space of dimension 1. But anybody else lives in a big dimensional space. And so that gets counted many times. And that changes the playing field.

**AUDIENCE:** What is d here?

**LAWRENCE GUTH:** What is d? Sorry, d is p. Sorry. Thank you. Is it right up there? Yeah. So d is the dimension of the representation. That's why I put it. But it was just a mistake. It's supposed to be p plus 1 over 2. So what needs to be spelled out more carefully? I think we should just spell out a little more carefully why the eigenspaces have this G action.

OK. So first of all, what is T mu star? Let's say mu star of G is just mu of G inverse. And if you write down the definition of the adjoint and play around a little bit, you'll see that T mu star is just T of mu new star. So T mu T mu star of f is just you take f. You convolve it with mu star. And then you convolve it with mu. So it's T of mu convolved with mu star.

OK. So our operator is still just a convolution operator, just convolving by something slightly different. And now we want to see that convolution commutes with some symmetries of g. OK. So our convolution has been defined to be on the right. And so I think it's supposed to commute with the symmetries of f on the left.

So let's say the left action of g on f with h is f of g inverse h. I want to say that convolution commutes with this action. So I want to claim that for any measure in nu, If I take T nu of left action g of f, this is left action g T nu of f.

OK. Cross my finger that I wrote everything correctly so this is true. T nu of left action g of f at some-- let's say I evaluate that at x. So I'm going to convolve this function with nu. So it's Lgf convolved with nu at x. So it's the sum x1 x2 equals x Lg f of x1 nu of x2.

So that's the sum x1 x2 equals x f of g inverse x1 nu of x2. All right. So the product of these two guys is g inverse x. So really, the product x1 tilde times x2 tilde is g inverse x f of x1 tilde nu of x2 tilde.

So that is T nu of f evaluated at g inverse x. So that's Lg of T nu of f of x. Phew. You probably can see that I'm not a very algebraically minded person. I think probably a more highbrow proof of this says something about acting on the left commutes with acting on the right.

Anyway, OK. So the group acts on functions, and that action commutes with this. So if you have an eigenspace of this, the group will act on it. All right. So if T nu of f is lambda f, then T nu of Lgf will be Lg T nu of f, which is Lg of lambda f, which is lambda Lg of f.

So g acts on V lambda, the eigenspace. OK. So that fleshes out this. OK. Cool. So let's do a quick application of this to see what we have learned about mixing.

So remember that we already know that this thing is bounded by 1. And it's not at all clear whether this is bigger or smaller than 1. It depends upon mu. If this is bigger than 1, then this is kind of useless. But if this is smaller than 1, then it gives us some new information. OK.

**AUDIENCE:** One question about proposition.

**LAWRENCE GUTH:** Yeah?

**AUDIENCE:** Does it require the existence of the representation of degree p plus 1 over 2?

**LAWRENCE GUTH:** Not exactly. This operator here has eigenspaces. And we know that the group acts on them. So therefore, there are representations of this group. OK, so application. So suppose A is a subset of SL2 Fp. It doesn't have to be A Selberg, just any subset.

And remember that mu of A is the uniform measure on A. OK. So what is mu of A L2 squared? Well, we square the amplitude of 1 over a squared. And then we add it up. And we'll get a contribution from A points. So we get 1 over the cardinality of A.

Right. So if we just plug in our estimate, proposition 2 tells us that sigma 1 of TA squared is bounded by p squared 1 over A. Whoops. I did that wrong. That's not good. This is about p cubed. This is about p. So it's bounded by p squared over A.

So we can see that if A is significantly bigger than p squared, then sigma 1 of TA is less than 1. So if A is 10 p squared-- or, even better, p to the 2 plus epsilon-- than sigma 1 is quite a bit less than 1. And so this random walk will mix quickly.

In particular, corollary of that is that if H is a proper subgroup of SL2 Fp-- so we saw earlier, if you take a proper subgroup and you walk, you just stay in the subgroup, it won't mix at all. Sigma 1 would be 1. Then H has size which is smaller than around p squared. And we saw earlier that there is a subgroup of size about p squared. So this is sharp.

OK. So this addresses the question that was raised earlier. The set A may be large. If you just know the size of A, can you say whether it mixes or not? This is essentially a sharp answer to that question. If A is significantly bigger than p squared, then the random walk does mix. And if A is smaller than p squared, A could live in a subgroup, and the random walk may not mix.

This whole phenomenon is quite different from what would happen in a commutative group. In a commutative group, you could take very large subsets, and still, nothing might mix, partly because the commutative group might have very large subgroups. But even if it doesn't, it wouldn't work. Yeah. Cool.

OK. So that's all very interesting, but it doesn't look very close to this theorem yet. Because this theorem says that we get mixing if our subset has size larger than p squared. And this is a subset of size 4. So we will have to think some more about how to do this. And we will talk about that next time.