[SQUEAKING]

[RUSTLING]

[CLICKING]

LAWRENCE GUTH: Today we're going to talk about the Szemeredi-Trotter theorem. I mentioned the Szemeredi-Trotter theorem on the first day, and you could think of as a fundamental theorem in projection theory. It's one of the rare theorems that is completely sharp. It's just the right answer for a natural question in projection theory. Yeah. So we're so we're going to talk about it and see-- yeah, so there's some interesting techniques involved in it.

Yeah. There, so that's what we'll do. All right. So let me remind you what it says. So what we talked about on the first day, I called it the Szemeredi-Trotter projection theorem. So the setup was that you have a finite set in the plane and you have a set of directions. And then S equals S of XD is the largest projection of X in any of these directions.

So that's the if. And then the conclusion is that the number of directions is bounded by S squared over X plus 1, which we saw in the first. OK. So this theorem is really a corollary of the theorem that they proved that is more general that I'll state now. And the more general theorem is a theorem about how points and lines intersect in the plane.

OK. So let's say X is a set of points in the plane and L is a set of lines in the plane. The number of incidences between X and L is the number of pairs. So take a point and a line. So that the point is in the line. Let me make a picture. So suppose those are our lines and these are our points.

How many incidences are there? This line has two points in it, so that counts two. And that line has one point and that line has one point, so that's four incidences. You can also think of this as the sum over all the lines in the set of lines of the number of points on that line. This is a way of measuring how much the points and lines are intersecting with each other.

OK. So the general Szemeredi-Trotter theorem says that if you have some points and some lines, the number of incidences between them is bounded by the number of points plus the number of lines plus points to the 2/3 lines to the 2/3. OK. Let's look at some examples to see where this stuff comes from. And the examples will also remind us of the connection with projection theory.

OK. So how could we have X incidences? That's not too hard to do. If you have one line and you put all the points on it, then the number of incidences is around X. How could you have L incidences? That's not too hard to do either. If you have one point and you have all the lines passing through it, then the number of incidences is around the number of lines.

OK. Now how can you get this? This is the most interesting, and it's a fundamentally trickier example than the first. So the third example is one that we have discussed. It's based on an integer grid. So let's say X is an N by N integer grid.

OK. And then we're going to try to draw some lines that each hit many points. And we observed in class on the first day that if you have a line at a rational slope, then it will go through many points. So those are our lines. So we have to make this a little more precise. So first I'll describe the rational slopes. So Q sub M is defined to be the fractions a over b where a and b are integers, and the a and b have size at most M.

OK. Now our set of lines is the set of lines with slope in QM. And they pass through a point of X. OK. This is a good day for colored chalk. So here is a typical line from our set of lines. I pick a rational slope and then I start at any of the points, and I draw a line at that slope. OK.

All right. So now let's try to compute how many points and how many lines and how many incidences this has. All right. So the size of X is N squared. It's easy to see. How many different slopes are we talking about? Well, the number of fractions is around M squared. Some of those are reducible. But it turns out that the number of distinct fractions is still around M squared.

So this is not trivial but not difficult. Maybe I'll put a little something about it on the homework. So now, how many incidences do we have? And we can say that, well, each point is in a line of every one of these slopes. So the number of incidences is the number of points times number of slopes. So that's N squared times M squared.

OK. Now how many different lines do we have? Well, we have M squared directions, so we have to think about how many lines are in each direction. And actually thinking about the number of lines in each direction, really you can think of it in terms of projection theory. So suppose S is one of our directions. Let's say pi sub S, the projection at slope S of X1 and X2 is X2 minus SX1, I guess. The projection whose fibers are lines of slope S.

All right. And so the number of lines in direction S is exactly the size of this projection. OK. And how big is the projection? Well, so let's make a little lemma. So S is in QM and X is as above. Then the size of this projection is bounded by M times N. So, proof. OK. So we have X1 and X2 are integers between 1 and N, and S is a over b. So if I take pi S of the point X1, X2, I have-- what is it?

X2 minus a over b X1, which I'll put over a common denominator. So it's b X2 minus a X1 over b. And this here is an integer of size at most, like, M times N. M times n. So there are at most M times N choices for what this could be up to a constant. OK. So therefore, the number of lines is smaller than the number of slopes times M times N lines per slope. So that's like M cubed N.

OK. These are written as upper bounds. There are also matching lower bounds, so those upper bounds are actually tight. And if you do a little algebra, you will see that the incidence is between X and L. So at least L to the 2/3, X to the 2/3. Just by plugging in that this is L, that's the number of incidences, and that's X. So that's where this thing comes from.

OK. So the two main ingredients, we're checking how many directions there that have slope-- how many rational numbers are there with height up to M and how big is each of these projections. And you can take those same computations and feed it into the projection theory setup here. You would get theorem one. You would get that this is a sharp example for theorem one. OK. I will pause for questions and comments.

OK. So this example, you can see from it that theorem one and theorem two are related to each other. And theorem two is more general because this could be any set of lines. Whereas in theorem one, in the projection theory setup, we would have a set of directions D and the set of lines would be all the lines in direction D that go through a point of X. So that's a somewhat specific set of lines, and this one is more general.

OK. So it's not very difficult to get from here to here, and it's not that interesting. So I'm going to leave it as an exercise on the homework. OK. Now another question that came up on the first day is, are there other examples in projection theory besides this square grid that also are that also are sharp for Szemeredi-Trotter? And there are very few that are known. But since we just did this, I can tell you a bit the other ones, and it's conceivable that the list I'll show you is everything.

OK. So what are the sharp examples for Szemeredi-Trotter? OK. So we've just seen the historically first example, which is a square integer grid. And that goes back to Erdos in the '60s, I think. There's a second example, which is a rather small variation on it, but we know so few examples that it's worth mentioning. It is a rectangular integer grid.

And that was done by allocation. So instead of N by N, just put N1 by N2 and then do a similar analysis. OK. Then there is a third thing that we can do, which is to replace the integers by integers in some other number field. So for example, integers adjoin the square root of 2, or perhaps other numbers.

OK. So this is in a paper by Olivine Silier and myself, and it's actually only currently worked out for some special number fields, but I suspect it works for any number field. If anybody knows more about number fields than me, this could be a nice little project to work it out correctly. Yeah?

AUDIENCE: One question about the rectangular grid. Is there any restriction on the aspect ratio of the grid, like a ratio of 1 to 2?

LAWRENCE GUTH: Yeah, there probably is. Yeah. So it takes a little writing and whatever. You see what happens for each one, each N1 and N2. Yeah. But there are sharp examples that involve grids that are not square. OK. So let me say a little more precisely what this means here. So we would have a set in our ring of things of height up to N.

So that would be the set a1 plus square root of 2, a2 and with the property with-- so with a1 and a2 integers and of size up to N. And our new QM. Maybe QRM. So quotients in this ring size M. This would be quotients of the form a1 plus root 2 a2 over b1 plus root 2 b2, where the ai and bi are integers and they have norm to M.

OK. And now we can do the same thing. So L is the set of lines with slope in QRM, and they go through a point of X. Oh, sorry. So X is just RN by RN. So that's the example. And then if you compute a little bit, you'll see that everything goes in a way that's analogous to the first example. And I guess I erased the key points, which was unwise. But if you think through the first example, we used an integer grid.

What did we use about integers? Basically we just used that integers are closed under taking products and taking sums. That was what we used in our proof. And this is a ring, this is a ring, so these guys are also, in a similar way, closed under taking products and taking sums. So it works the same way. Yeah.

OK. So it's a big mystery in the field to have what's called an inverse theorem for Szemeredi-Trotter. An inverse theorem would say that if you had a set X and L where this was tight and especially where this dominated, then what could you say about X and L? Or as a special case, if you had a set X and a set of directions D where this is tight, what could you say about X and D?

On the one hand, we have a handful of examples where they're tight, and they're based on a lot of algebraic structure. They're based on the ring of integers in some number field. And on the other hand, we have very little in the way of theorems. That's about what we can definitely say about X and L if this is tight. Yeah?

AUDIENCE: From [INAUDIBLE] quadratic number field examples, is it important for it to be a unique factorization domain? Because it seems like if that fails, then you get kind of extra collapsing along objections. But I don't know if that's a problem.

LAWRENCE GUTH: Yeah. So the question is, is it important whether our ring has unique factorization? I don't think it's too important. In our proof over the integers, I don't think we used unique factorization of integers.

But there are some things we would have to check about each number field. Like, we'd have to check that how many quotients are there in this thing of quotients. Yeah. So anyway, so I'll put something about it on the problem set as an optional project for people who want to explore. Yeah?

AUDIENCE: For the original integer grid case, we required lines that have different slopes. Here, do we have some requirements on our choice of ai, bi?

LAWRENCE GUTH: OK. So I think the comment is that in the original case. I said that the a over b should be a reduced fraction.

AUDIENCE: Yeah.

LAWRENCE GUTH: And do we have an analog of that?

AUDIENCE: Yeah.

LAWRENCE GUTH: Yeah. Yeah, we do. Right. So this is the set of all the fractions a over b. And if a given real number-- if a given rational number can be represented in two different ways, it still only counts once in QM. And the same thing is true here.

So the same number here might have more than one representation of this flavor, which makes it not trivial to count the cardinality of this set. And we do need to know the cardinality of this set to correctly analyze things. I guess something like unique factorization could come in there. Before, I said put it in lowest terms, which is maybe not quite as clear what that means if you don't have unique factorization. Yeah?

AUDIENCE: So say in the regime, X is greater than L squared. Can you see that L squared dominates in some way and there's no other examples like the square root quotient?

LAWRENCE GUTH: Yeah. Yeah. Let me say the question back and then I'll make sure I understand. OK. So if you look at this right-hand side, different terms will dominate depending upon the relative sizes of X and L. And there's different regimes. So one regime is that X could be bigger than L squared. So you have just a ton of points and not very many lines. And in that situation, this term is bigger than that term and that term. So in this situation, the X dominates.

So I think the question was, can we describe some structure theorem in this case? Yeah. So I think we can. So in that case, the example is example one here. And so the structure theorem would say that you really have to have-- well, there may not-- let me say that. So in this regime, I don't think there's that much structure. So I have some lines, not very many. And then I have a ton of points and I want to have x incidences.

That means I had better put each of my points on a line. So I put each point wherever I like, not off in the middle of space, but on one of these lines. And then I get this many incidences. So there's not that much of a structure theorem we could ask for here. And there's not that much of a structure theory we could ask for when this dominates. But here, if that one dominates, there are very few examples and there could be a structure theory. Yeah?

AUDIENCE: Sharp example in theorem two has to be sharp example in the projection version of the theorem?

LAWRENCE GUTH: Yeah. So the question is, how do sharp examples in the two things compare? So if you start with a sharp example in the projection theorem, then you can make a set of lines. That's the set of lines that go through X and have directions in D. And that will become a sharp example in the general theorem. Now, if you start with the sharp example in the general theorem, it may not have the special form that makes it relevant for the projection theorem, so you can't go the other way.

AUDIENCE: So you always have to do that. The sharp example, they are like the rectangle with the square grid and the later one, all of them, they are sharp for both cases.

LAWRENCE GUTH: Yeah. So all the sharp examples we mentioned are sharp examples for both cases. That's right. I guess there's one more comment about it, that this setup is invariant under projective transformations. Projective transformations take lines to lines and of course take points to points. And so if you start with a sharp example, maybe coming from a grid, you can apply a projective transformation to it and get a new sharp example for this theorem. But it won't any longer correspond to the projection set.

OK, so the full list of sharp examples for this theorem that we know about is the ones I just wrote on the board, plus you can apply projective transformations. Yeah. In terms of proving an inverse theorem, which is a big open problem, proving an inverse theorem for this would be harder and more general than proving an inverse theorem for the projections. The one for projections, we don't know either, and that would be a good starting point. It might be a little more accessible.

OK. Let's keep going. So the goal for the class is to describe how to prove this theorem right. The first estimate I'd like to show you is a double counting estimate for the incidences, which is in a similar flavor to the double counting estimates we've done earlier in the class. But let's just see how much double counting tells us about this problem. All right. So double counting.

All right. So by double counting we will prove the following estimate. I'll call it lemma one. The incidence is between X and L are bounded by X times L to the 1/2 plus L. All right. So here's the proof. OK. So the incidences are the sum over the lines, and I count how many points are on each line. And now I'm going to Cauchy-Schwarz that. So it's less than that.

OK. So what do I gain by Cauchy-Schwarzing it? Well, if you think about the number of points on this line squared, it's very similar to counting the number of pairs of points on the line. And then we can make use of the fact that for any pair of points, it only lies on one line. All right. So I'm going to bound this by here. OK, so it's not quite true.

Let me say something in parentheses. It is not quite true that the number of points on the line squared is smaller than the number of points on the line choose two. That would be true if the number of points on the line is pretty large. But if there's only one point on the line, then this is 1 and this is 0. So I have to add here. I guess I can just add one.

OK. So now I get two terms. So this one contributes number of points on the line choose two. All I'm doing is changing this to this, and then I still have my L to the 1/2. And then if I put in one here, this gives me the number of lines to the 1/2 which groups with that and I get the number of lines. This matches that. OK. So this here represents lines that contain at least two points, and this represents lines that contain only one point.

All right. And now the key point that that is smaller than that. So what are we counting here? For each line, we're considering pairs of points on that line. But you can also think you could sum over all the pairs of points and then ask how many lines go through that pair. And there are at most this many pairs of points and for each one of them, there's only one line. So that's the double counting.

So this here is the number of X1, X2 and L so that X1 and X2 are on L. And if you count the lines first and then add up over the lines, you get this expression. But if you fix X1 and X2 first, there will only be one line and so you get this. All right. And now this is that. So that's the end of the proof.

OK. So let's make some remarks about this simple proof, what double counting tells us and what double counting doesn't tell us. All right. So first, we only used that two points determine a line. The only fact we used is that through two points, there's only one line. So that means that it works in some more general cases.

So for instance, it works for points and lines over finite fields. Because over finite fields, it's still true that through any two points there's only one line. And generality is good, but this is a blessing and a curse in this case, because the whole Szemeredi-Trotter theorem is not true over finite fields. I'll show you an example in a second. And that means that this proof, which only uses this fact, cannot really be proving the whole Szemeredi-Trotter theorem.

So that's remark B. So the Szemeredi-Trotter Trotter theorem is false over Fq2. Here's the example. X is the entire plane, all the points Fq2, and L is all of the lines. You could take literally all the lines, or you could take all of the lines y equals mx plus b, where m and b are in Fq. It doesn't matter very much which one. OK. So size of X is q squared. The size of L is q squared. And how many incidences do we have? Well, each line has q points.

So incidences between X and L is q times L which is q cubed. OK. So this example is sharp for lemma one. That's equal to XL to the 1/2. That matches that term up there. And it's much bigger than the right-hand side of Szemeredi-Trotter. Right. OK, yeah. So let's do a little algebra.

So we have here an X times L to the 1/2. In Szemeredi-Trotter, we have an X to the 2/3, L to the 2/3. So there are different regimes about who dominates there. But this thing is always bigger than this thing, and it's usually much bigger. So here, X times L to the half is q cubed. X to the 2/3, L to the 2/3 would be q to the 8/3, which is much less than the number of incidences. So the Szemeredi-Trotter bound does not hold in this example.

OK. So that shows something important about the Szemeredi-Trotter theorem, that we cannot prove it just by using the fact that two points determine a line and doing our double counting in a very clever, complicated way. We genuinely need another input. And all of the proofs of Szemeredi-Trotter theorem use somehow the topology of the plane. So see all proofs of Szemeredi-Trotter use the topology of the plane in some way.

And the topology of R2 is kind of different from the topology of Fq2, whatever that might mean. OK. OK. Cool. So I know four different proofs of-- I've seen four different proofs of the Szemeredi-Trotter theorem, and they use topology in some different ways. I'll show you just one. I think it's probably the one-- the proofs are not that different from each other. I think this probably is the one that is the most convenient for thinking about projection theory for balls in R2. Yeah?

AUDIENCE: You're saying Szemeredi-Trotter, but you always mean this one on the lower board, not the--

LAWRENCE GUTH: Yeah. That's right. That's right. Yeah. So we'll prove this theorem. It's not too hard to see that this theorem implies the one on the top. Yeah. OK. Cool. So the proof I'm going to show you is based on cutting the plane into pieces. And it goes back to work of Clarkson, Guibas, Edelsbrunner, Sharir, and Welzl. It's not the very first proof, but I think it's a very nice proof. And then it's a slightly revised version of that that Nets Katz and I worked on.

OK. So it's based on cutting the plane into pieces. So let me say what I mean by that by stating a lemma. All right. So cell decomposition lemma says that if X is a finite subset of the plane and S is a natural number, then there's a way to cut up the plane.

So I think of it as a cell wall and then a bunch of cells. So each one of these is open, and I'll call it a cell. And this set W is closed and I'll call it the wall between the cells. So the image here is that I have some points. So that's X. And then I have something like this. That's W. And this is 01, 02, 03, and 04. OK.

OK. So it has some properties. Number one. Each line intersects at most around S cells. And also, the line can intersect the wall but in at most around S points. OK. And then each cell in the cell decomposition has at most X over S squared points.

OK. Now it's not very obvious yet where all these formulas came from. And it's definitely not obvious how to prove this, but it's pretty easy to prove this in the example that X is approximately like a square grid. So let me just show you that as a motivation for where this might come from. And to help digest what the lemma says.

AUDIENCE: [INAUDIBLE]

LAWRENCE GUTH: Ah. Yeah. So, right. So the comment is I didn't say anything about a set of lines. So what this means is for every line L. And here, this is for every line L. Now if we had a set of lines, all we would really need is to have this for the lines from our set of lines. But in practice, it's actually just true for every line. OK.

All right. So the cell decomposition lemma, in general, is somewhat difficult. We'll talk about the proof. But let me show you an easy case that I hope will motivate the statement. Yeah?

AUDIENCE: You want X to not intersect the wall? It looks like you just put X in the wall.

LAWRENCE GUTH: It looks like we could just put X in the wall. Let's see here. Yeah. So another comment is that X is allowed to be a subset of the wall. And that sounds a little suspicious. But we do get something from that because every line only hits the wall in S points. So if x were a subset of the wall, then every line would only hit X in S points, and that would be a useful tool for us. Also remember that S is a parameter.

So for each S, there's a way of doing this. We're going to eventually-- we'll look at our set of points and our set of lines, and then we'll pick an S that's strategic. So if S is small, this is really quite strong. OK. So here's an example is that X is roughly a square grid. So let me make that more precise. Let's say X is contained in 0 to N squared, and X intersected with a unit ball around any center is small.

And let's say that there are around N squared points in X. So what does this X look like? We have an N by N box. Yeah. Let me imagine cutting it like a checkerboard into unit squares. And this tells me that I have on the order of 1 point of X in each unit square. So this is a class of sets. So I call this is roughly a square grid.

OK. So now how would we cut it up? I cut it up like this. I'll do N over S. So the wall is S vertical lines union with S horizontal lines. OK. So the wall cuts it into S squared pieces, and because my set is distributed so evenly, each of these pieces has about the same size. So that gives me the third bullet.

And if we were to make a line in this picture, well, it can only cross each of these other lines once. So a line will intersect the wall at most 2S times. And to go from one cell to another cell, it has to cross the wall. So only enter about 2S cells. OK. So the bullets are not hard to check. OK. OK.

So this decomposition method is based on the key observation that it's actually helpful to cut up our problem like this, cut up our space like this, and then count the incidences in each cell separately and the incidences on the wall. And in each of those pieces, we can use our double counting argument. But when we add it up at the end, we may get something better than using the double counting argument globally. OK.

All right. So proof of theorem two, assuming this lemma. All right. So we're given X and L, and S is going to be TBD. And we can apply this thing. OK. So let's say that Xi is X intersected with Oi. Then we can say that Xi is less than or equal to less sim X over S squared. That was given to us as a bullet point. And the sum over i of Xi is at most X, because these are disjoint sets.

OK. Now let's say that Li is the set of lines that intersect the cell Oi. OK. Now these are disjoint. Each point is only in one cell. These are not disjoint because a line goes through many cells. So this blue line belongs to Li for this i and this i and this i, but not for this i or this i or that i.

OK. So if you add up over i, the size of Li, that is bounded by S times the size of L, because each line in L belongs to at most S cells. All right. So now let's count our incidences. So there are incidences from each cell. And the incidences related to the i-th cell have points from the i-th cell and they have lines that enter the i-th cell. There also are incidences involving the wall. So add the incidences from X intersected with the wall and L.

OK. Now we apply our double counting lemma to estimate each of these things. So that's less than the sum on i Xi Li to the 1/2 plus Li. And for the wall, well, each line only hits the wall S times. So we have S times the number of lines. OK. Now from this point on, we just do a little bit of algebra. We just have an expression involving these numbers Xi and Li, and we know a few inequalities about Xi and Li. We'll just put them together.

Here, it's reasonable to Cauchy-Schwarz because I know information about Xi and Li separately. So that will disentangle this, sum on i Xi squared to the 1/2 sum on i of Li to the 1/2. This sum on Li, I just read that off, and that can be grouped with that. OK. And now, what do I do with this sum of squares? Well, I know that each Xi is bounded by something, so I can pull that out. So I'll have X over S squared, and I'll have left the sum of Xi.

This sum of Li, well, it's just here. So it's SL to the 1/2. And I copy this. OK. This sum of Xi is bounded by X. And so then I'll do a little algebra here. This is the important term. I'll have an X squared to the 1/2, so X and X. I have an L to the 1/2. And I'll have some powers of S. There will be 1 over S coming from here and an S to the 1/2 coming from there. So altogether I have an S to the negative 1/2. And this I'll copy over. OK. Let's pause there for a second after the algebra.

OK. So let's take stock of the bound that we got and how it compares to the double counting argument. If you compare this to the double counting argument, the terms match up. So the double counting, we have XL to the 1/2, and that first term gets decreased by S to the minus 1/2 when we introduce our cutting and then the second term of L gets increased by a factor of S.

So if originally the first term dominated, then life will get better when we do this decomposition. And if originally the L dominated, then life will not get better when we do this decomposition. Right. OK. So at the end here, we just choose S greater than or equal to 1 to minimize the right-hand side. And if you do a little bit of algebra, you'll get out the Szemeredi-Trotter bound. OK.

I remember spending some time trying to digest why things were getting better when we subdivided. And here was one thought I had about it. If you look at the double counting bound, there are these different regimes. So there's a regime where the left term dominates. And in that regime, the double counting bound is bigger than the truth. And there's a regime where the right term dominates, and you can always have L incidences, so that's not that-- when the right term dominates, it's the truth.

OK. So what are the regimes? The right term that the L dominates if L is bigger than X squared. So you have a lot of lines compared to the number of points, then double counting already gives a sharp bound. OK. So suppose originally that L is not bigger than X squared. So if I just apply double counting out of the box, the XL to the 1/2 would dominate, which I'm not so happy with. Then I try cutting things up into cells. I count the incidences in each cell.

Is life any better? Well, I have S squared cells. So if I look-- and the points are kind of evenly distributed. So if I look inside one cell, the number of points goes down by S squared, which is quite a lot. How about lines? Well, not every line goes through this cell, so the number of lines also goes down. But because each line goes through many cells, each line goes through S cells, the fraction of lines that I would expect to go through this cell is 1 over S of all the lines.

So the number of points goes down a lot, and the number of lines only goes down a little. By doing that, I'm tilting things towards having a lot of lines and few points. And if you think about-- if you track what is the best value of this S, what happens is the S is best when we cut it in such a way that inside of each cell, the number of lines is the number of points squared. And that means the double counting is giving the correct bound inside of each cell. OK. Questions or comments?

OK. So let's talk about how to prove the cell decomposition lemma which uses a little bit of topology. So in our cell decomposition lemma, we want to take our points and cut them up in evenly in some way. And that is reminiscent of a classical thing in topology called the ham sandwich theorem. And the cell decomposition lemma is like a variant of the ham sandwich theorem. So I will first recall the ham sandwich theorem and how to prove that, and then we'll work up to this.

All right. So our starting point in topology is going to be the Borsuk-Ulam theorem. So I'll recall that, I'll state that, and then we'll try to prove everything else from Borsuk-Ulam. All right.

So the Borsuk-Ulam theorem says that if you have a function from the N sphere to RN, which is continuous, and its antipodal. So that means f of negative theta is negative f of theta for every theta in the sphere. Then the conclusion is that 0 is in the image of f, there is some point in the sphere that got mapped to zero. OK.

OK. So as a corollary of that, we're going to do the ham sandwich theorem. So one version of the ham sandwich theorem says that if O1 up to On contained in RN are bounded open sets. Then there exists a hyperplane h that bisects them all. So for instance, if I had two bounded open sets in the plane, there would be a line somewhere that bisects both of these sets.

OK. So proof. So first we need to parameterize our choice of all the hyperplanes. And a good way to think about that is that a hyperplane is related to a linear function, so this is a hyperplane. I also am going to want a hyperplane and one side of it. I'll pay attention to which side is which. So this shaded region would be the region a1, x1 plus an, xn is bigger than b.

So one side of a hyperplane can always be described that way. And we have the freedom to scale. So let's say that theta is a1 up to an comma b. That could be in the N sphere by scaling theta, which doesn't change the hyperplane. OK. So those are my hyperplanes. And now f of theta.

Here, let me say it this way. This guy plus b is bigger than zero. And this thing here, I will call L sub theta of X. OK. So f of theta is-- f sub i of theta is the volume of Oi intersect L theta is bigger than zero. That's the volume of Oi on this side minus the volume of Oi and L sub theta is smaller than zero. So fi of theta is the volume of this part of Oi minus the volume of that part of Oi.

OK. So notice that fi of theta is zero if and only if the hyperplane associated to theta bisects Oi. OK. Also, fi is antipodal. fi of minus theta is negative fi of theta. Because if I replace theta by minus theta, I take the minus of all these coefficients. Then I'm replacing L of theta by negative L of theta. And so that switches who-- so I have the same zero set, the same hyperplane, but now this one will be positive and that one will be negative. So I'll switch these two things, which reverses the sign of fi of theta.

OK. fi is also continuous. So if I adjust theta, then I'll make a small motion of this plane. And that will change these volumes but only continuously because the sets are bounded and open. OK. So finally I'll make f by just putting together all these fn's. And so f is a function from the N sphere to RN, and it's continuous and it's antipodal. So the conclusion is-- so then Borsuk-Ulam tells me that there exists some theta so that f of theta is zero. And that means that h theta bisects all these sets. OK.

All right. Now this by itself is not going to be super helpful for us. We have a set that lives in the plane. So the theorem says, I can find a line that bisects any two things. I'm hoping to take a complicated set and cut it into many even pieces. And so a more relevant warm up is some way of bisecting many sets, not just two.

OK. So if I put more open sets here, then there will not be a hyperplane that bisects all of them. But I could be more flexible about what cut I would allow. So it can't be a hyperplane. It has to be a little bit more complicated. And it turns out to be a reasonable idea to consider cuts by polynomials of some degree. So there's a natural generalization of this. You could do really any class of functions, but for this application, polynomials are useful. It's called the polynomial ham sandwich theorem. All right.

All right. So here's a more general version polynomial ham sandwich. OK. So let's say polynomials of degree D in RN that is the set of polynomials-- so polynomials in N variables over R so that the degree is at most D. And so this is a vector space and it has dimension approximately D to the N if the dimension N is fixed and the degree is going to infinity. The dimension is not hard to figure out. You count how many monomials are there of degree at most D.

OK. So the proposition says that if capital n is strictly less than the dimension of this space, then there exists a polynomial p, the space of polynomials, which is nonzero, and the polynomial bisects each of the open sets. So let's say fi of p is defined to be the volume of Oi intersect p is positive minus the volume of Oi intersect p is negative, and fi of p equals 0 for all i from 1 up to n.

So that's the statement of the proposition. Instead of using a hyperplane, we use a degree D zero set of a degree D polynomial to bisect our sets. And then we can bisect many more sets. This is exactly how many. As D increases, we have more degrees of freedom, bigger dimension, and we can bisect more things. So at the level of intuition, you can think of this in terms of degrees of freedom.

So the dimension of this space tells us how many degrees of freedom we have to play with. Each one of these is a constraint. And if the number of degrees of freedom is at least the number of constraints, then the theorem says we can actually find a solution. OK. The proof is essentially the same as this.

So proof. What do I need to check? Well, fi. So I just need to check that fi are continuous and fi of negative theta is negative fi of theta. So why is it continuous? I won't do a full proof, but if you adjust the coefficients of this polynomial a little bit, then the set where it's positive changes but only a little bit. And so this volume should only change a little bit. That's continuity. And why is it antipodal?

Well, if I switch p to negative p-- actually, yeah. Let me put p's here. If I switch p to negative p, then it interchanges these two sets. And so it switches the sign of fi of p. OK. Cool. All right. And now f is going to be f1 up to fn. Now f is a map from where to where? f is a map from polynomials of degree D on RN, take away zero to RN. Why is that?

I was waving my hands a little bit when I said that fi is continuous and fi is not continuous at zero. If the polynomial is zero, then both of these sets are empty. So I guess by the definition, fi of 0 would be zero. But then if I move p a tiny bit so it's not zero, then these sets will appear and they take up most of space. So there'll be a jump discontinuity in those sets.

So the reason we had to do this is that fi is only continuous on the space of polynomials without zero. So if I take a non-zero polynomial and I change the coefficients a little bit, then this set will only change a little bit. OK. Right. So actually, I should say that that's not completely trivial, that fact. It takes a little proving.

It's related to the fact that if you have a non-zero polynomial, then any of its level sets cannot have positive volume. If you have polynomial equals 3, that defines a hypersurface and it cannot happen that that hypersurface has positive volume. That's basically what you need to check that this is continuous.

OK. So this is a space from here to here. And to apply Borsuk-Ulam, I need to have a sphere for my domain. So this sphere is sitting in here as long as n is strictly less than the dimension of this space. This is where we use this hypothesis. OK. And then Borsuk-Ulam says that zero is in the image of f, which is what we're trying to prove.

OK. Cool. All right. So we're getting close to having a tool that we can use to do this decomposition. Now we have a little technical issue or a little mismatch, because what we want is to start with a finite set and have a cell decomposition that kind of cuts it into even pieces. This proposition is about open sets. If we have a bunch of open sets, we can simultaneously cut them each in half. So we need a version where the open sets are replaced by finite sets. That's not difficult to do. It's just a technical thing. That's how it works.

OK. So proposition. If Si contained in R2 or RN, I guess is OK, our finite sets, i equals 1 up to N, and N is less than the dimension of poly D of RN, then there exists a non-zero polynomial poly D of RN. What does it mean? So in some sense, we'd like this polynomial to bisect these sets. But we have to say something slightly weaker, and I'll explain why.

So that Si intersect p positive is less than or equal to Si over 2 and Si intersect p negative is less than Si over 2. OK. So this set could be a set of three points. So it's not possible that half of them will be positive and half of them will be negative. So the best we could hope to is that one would be positive, one would be negative, and one would be zero. OK. OK.

OK. So proof. For any epsilon bigger than zero, let's say O epsilon i is the epsilon neighborhood of Si. So that's an open set. And so we can apply proposition to that set. Apply polynomial ham sandwich to O epsilon i. And we output a polynomial p epsilon, which is in the unit sphere in the space of polynomials. And it bisects.

So we know that the volume of O epsilon i intersect p bigger than zero is half of the volume of O epsilon i. OK. So we have this for every epsilon. And now we want to take some kind of limit as epsilon goes to zero. And the sphere is compact. So the sphere SN is compact. So we have a subsequence p epsilon. Some subsequence of them converges to a limit p, which is a polynomial in SN.

OK. And the claim is that the p does the job. So let's say this is the zero set of p. Over here p is positive, and down here, p is negative. And suppose a particular Si was not split up evenly like this. So more than half of that on that side. Here, let me do that. Yeah. There it is. So why can't that happen? Well, think about what p epsilon would be doing.

So O epsilon i would look like this. And what would p epsilon look like? Well, it's a little perturbation of p. We have some convergence. So that would be the zero set of p epsilon. Here is p epsilon is positive. Here is p epsilon is negative. And that violates this.

All right. So if Si intersect p is positive is strictly greater than 1/2 of Si, then we have O epsilon i intersect p epsilon positive would be strictly greater than 1/2 of O epsilon i contradiction. OK. OK. So now let's prove the cell decomposition lemma. So we prove the cell decomposition lemma by using the polynomial by using this proposition many times.

So we have some set of points X. And first we use our proposition with N equals 1. And so D equals 1. So we're actually an easy case of our proposition. We're only bisecting this one set of points. So here's p1 equals 0. So the output is a polynomial p1 of degree D1 equals 1. And it bisects the set. So we have X plus is the set of X and X where p1 of X is positive, et cetera.

OK. Then we use our proposition with N equals 2, and we get a polynomial P2 of degree D2. D2 would probably also be 1, but let me just put the general thing. So it would be less than our N2 to the 1/2. OK. P2 bisects X plus and X minus. So here's the zero set of P2. And it bisects each half, so now I've divided X into quarters.

OK. And we keep going. So at step K I use my proposition, and I have NK, which is 2 to the K, and DK, which is less than NK to the 1/2. So that's like 2 to the K over 2. I'd output a polynomial PK, and it bisects the 2 to the K minus 1 previous guys. The 2 to the K previous guys. I'll draw one more. I have four pieces, and I find a new polynomial and I bisect all four.

OK. So now, when do I stop? I choose the final K. So 2 to the K is S squared, and that means the final DK around 2 to the K over 2 is around S. OK. So since I bisect at every stage, I get X intersect Oi smaller than X over 2 to the K, which is like X over S squared. So that's my third bullet.

So I'm trying to prove each cell has only a 1 over S squared fraction of the original points. We have that. And then I'm trying to think about how the lines intersect the cells and how the line intersects the walls. So these are the different cells. And the red stuff is the walls. So if I draw a line here. The line intersected with the wall is smaller than D1 plus dot, dot, dot plus DK, because how many times does a line intersect the zero set of a polynomial. It's at most the degree.

OK. So these degrees are bounded by 2 to the k. Over 2. So this is a geometric series. It's bounded by its last term 2 to the K over 2, which is around S. So that's the second bullet. And every time this line moves from one cell to another, it has to cross the wall. So that also bounds the number of cells that the line intersects. So that ends the proof or proof sketch of the cell decomposition lemma. Any questions or comments?