[SQUEAKING]

[RUSTLING]

[CLICKING]

LAWRENCE GUTH: All right. Well, last class, we started to talk about the Fourier method in Euclidean space in R2. And the big goal for the class is to digest the Fourier method, finish that story, but also just to digest it in general. And then after that, if there's time, we'll start to introduce sieve theory, which has a extremely parallel story. But that was developed by a different group of people, and with somewhat different goals.

OK, so to digest the Fourier method, I wanted to go back for a second to the finite field Fourier method, where everything is a little bit cleaner, and to make a point about it. OK, so our main lemma in the finite field case was this. If you have a set of lines in FQ2 and then you add up their characteristic functions, then it has a big constant part, and then a small, high-frequency part. And then the conclusion is that f is f0 plus f high, where f0 is a constant function, which is the number of lines over q.

And from that, you can compute its L2 norm. L2 squared is number of lines squared. And f high means 0. And we have a bound for its L2 norm.

OK, so one thing that you can get out of this is a bound for the L2 norm of f. If you add these together, you get a bound for the L2 norm of f. And it's worth saying that that bound for the L2 norm of f has a elementary proof that's really simpler than this. Let me show it for comparison to help put this in perspective.

OK, so let's call this, elementary L2 bound. OK, so there's a first lemma that if you have two different lines, then you take the sum over x, L1 of x, L2 of x, this is at most 1. Could do that. It's just saying that two lines intersect in at most one point. OK, and we can use this to make an L2 estimate. So here's the L2 estimate. If f is the sum L in L, of L of x, then f L2 squared is bounded by the number of lines squared plus number of lines times q.

Here's the proof. So f L2 squared is the sum over L1 and L2 of sum on x, L1 of x, L2 of x. And now each one of these sums, we can control with this lemma. All right. But you have to be a little bit careful. If L1 is equal to L2, that's a different situation. A line intersects itself in a lot of points.

So we're going to split up this sum. So it's the sum L1 equals L2 plus the sum L1 does not equal L2. OK, so here, we have only L choices for L1, or only L choices appear in this sum. And each one of them contributes q. And then over here, we have L squared choices for the lines. But each one of them contributes 1. So that's the proof.

All right. So if you compare this with this, something looks kind of similar, right? We get an L2 bound up there of L squared plus Lq that matches L squared plus Lq that we get here with this quite simple argument. OK, but the main lemma has more information than that. Well, there's actually really two cases. If q is bigger than the number of lines, then this term dominates here, and this term dominates here. And in that situation, the main lemma doesn't really have much more information than this, because the function fh is not that special.

But the other regime is that the number of lines is much bigger than q. So this dominates that. And in that situation, most of the L2 norm is coming from f0, which is very special because it's constant. So the main lemma has an important piece of information beyond this that when we have many lines, this sum is almost constant. And then it's a constant plus something that's much smaller.

OK, so that is also true in spirit for the main lemma in the real case. But the main lemma in the real case is a little bit more complicated. So I wanted to do this first, and then we could look for the analogous thing in the more complicated real case. OK, so what did the main lemma say in the real case?

Main lemma in the real case. So now, we have T is a set of 1 by R rectangles. And psi T is a smooth approximation of the characteristic function of t. And the exact meaning of that is explained in agonizing detail on the homework. OK, and then we have a function, which is the sum over all the T and T of this smooth characteristic function of T.

OK, now our bounds are going to depend on how these tubes are clustered. So N sub T of R is the maximum over T tilde is a 2 little r by 2 big R rectangle, or tube, of the number of tubes in there. So picture this. We have this T tilde. Let me see how many little tubes are in T tilde.

OK, so those are the hypotheses of the main lemma. And then the conclusion is that you can write f, you can break it up into pieces of different frequencies. So the pieces are labeled by a sort of a width parameter, little r, which is a dyadic number between scale 1 and scale capital R of f sub little r, where the f sub little r has two good properties. It has a frequency support. The support of f little r hat is contained in the ball of radius 1 over r.

And it has an L2 norm bound. f sub little r L2 squared is bounded by the number of tubes times this packing, clustering, parameter at the scale little r times big R over little r. OK.

OK, now one thing you could take away from this is you could take away from it a bound for the L2 norm of f by just adding up all of these. And like before, that bound for the L2 norm of f has a simpler proof. And for context for understanding this, I'm going to show you this simpler proof. Yes.

AUDIENCE: Are the fr's orthogonal?

LAWRENCE GUTH: The fr's are orthogonal. Yeah, the fr's are orthogonal. If we didn't know they were orthogonal, we would lose a little bit. We'd have to do Cauchy-Schwarz to compute the L2 norm of f. But the amount that we would lose would be like a log r, which is relatively minor. Yeah. OK.

OK, so let's do an elementary L2 bound. All right. So like before, when we expand out the L2 norm, we'll have to estimate the integral of 1 tube times another tube. But one tube times another tube is a little bit more complicated than one line times another line. Two lines always intersect in a point, but two tubes intersect in a different amount of area depending upon the angle between them. So we're going to keep track of that.

All right. So let's make a definition. The width, r, in between T1 and T2 is the minimum r, so that there exists a 2 little r by 2 big R rectangle T tilde that contains them both. OK, so if this is T1, and this is T2, then that would be T tilde. And that would be r.

OK, so now, if I want to integrate T1 of x, T2 of x, dx, that's going to be about big R over the little r associated to these two tubes. Why? Well, the integral is the area of their overlap. And if I'll sort of sketch their overlap over here-- it's not a great picture. This is around 1 because it's contained in one of the tubes. And this axis here is like big R over little r. As little r gets smaller, they overlap more and more.

OK, good. So now, we can make an L2 bound. Let's say that the integral of f squared is bounded by the sum over little r dyadic of T times N sub T of little r times big R over little r. OK, and the proof is that we'll just break up this sum as a bunch of integrals like this, group them according to the value of little r, and we'll see what happens. This is what's going to fall out. So here's the proof. So the integral of f squared is the sum T1 and T2 in our set of tubes integral of T1 of x, T2 of x.

Now this integral here, it depends on the value of little r. So let's group these by the value of little r. So that's the sum over little r dyadic sum over T1 and T2 with that value. So the R associated to T1 and T2 would be about this r. And then I have this integral, which I can bound by the lemma. So I get big R over little r.

OK, now I need to know how many pairs of tubes are in this list here. Well, I can choose T1. And T1, as far as I know, could be anything. So I have number of tubes choices for T1. But once I choose T1, I need to choose another tube, T2, that fits into the same rectangle with T1. So the number of choices for that is about N sub T of little r. That's what it's measuring. So this is bounded by sum over r dyadic of number of tubes and N sub T of r big R over little r.

OK, great. So let's compare this with this. This is exactly what we get by just adding up all those L2 norms for the f sub r. OK, so this proof is a lot easier than the proof of the main lemma that we did last time. Let's compare the L2 bound and the main lemma. If I were to erase this information in the main lemma, then all that would be left is the L2 bound. But there is this information. And this is helpful. This is telling us something more.

OK, now depending upon what we know about this thing here, different terms might dominate this sum. If it so happens that the term little r equals 1, has the biggest contribution here, then we don't learn very much because our information about the support of f1 hat is the same as the information we already had about the support of f hat. So we'd have nothing new.

But it may well happen that the term where this is biggest has little r much larger. And then f little r has this frequency support condition. It gives it some structure, some extra information there that's going to help us.

So let's talk about, what does it mean about the function f little r that its frequency is supported there? So intuition, we're going to call this the locally-constant intuition, says that if I have some function g and the support of g hat is contained in the ball of radius 1 over r, then the function g is approximately constant on balls of radius r.

So if the support of g hat is in the ball of radius 1 over r, then g hat is a combination of waves, cosine waves, where all of the cosine waves have frequency at most 1 over r. So each one of those cosine waves is not changing very much on balls of radius r. So it's sort of intuitively plausible that after we add them up, also, we'll get something that is roughly constant on balls of radius r. We'll come back in a little while and make a precise statement about this.

OK, so based on this, let me make a picture illustrating main lemma r. Main lemma and the real case. OK, so OK. So this will be, well, this will be like each one of those lines is going to be a 1 by r tube. So I may have a whole bunch of them over here sort of like this, and a whole bunch of them over here sort of like this. And over here in this part of the board, I might have some that are not so clumped into thick tubes. They might look sort of like this.

All right. So over here, a picture I wanted to give you is that I put so many thin strokes of chalk, one by our rectangles, that they sort of smooshed on top of each other, and they produced, like, a thick stroke. And so this part over here might be dominated by f of r. So maybe this scale might be r and f sub r be sort of a sum of this blob and that blob. And on the other, so this over here in blue would be f sub r.

And on the other hand, this over here would be F sub 1. All of this stuff is pretty high frequency. OK, and notice that this f sub r is kind of roughly constant on any ball of radius r. And so if I were to visualize the set where F is big, well, there might be some bits in here where F is big. But then there also might be some piece in here where F is big. OK, so over here, this is where F sub 1 is big. And over here, this is where F sub r is big.

All right. And now, the kind of trade off that happens is we might have a very strong bound for the L2 norm of f1. And so that bound tells me that this set where f1 is big is small. And then we have a bound for the L2 norm of fr. It might not be such a good bound. It might be bigger. But on the other hand, fr has this nice structure that it is constant on balls. So the set where fr is big is group is clumped into these big balls. And so that gives us--

So the area might be larger than this area. But the fact that it's clumped gives us some other geometric information about it. It's kind of simpler. OK, so let's pause there. That was recapping the main lemma in the real case. And I sort of tried to have a picture of what it's telling us.

OK, OK. So the next thing that we should do is this is sort of nice, but it is not a actual precise math statement. So in order to really make use of this idea, we need to now replace this intuition by an actual precise thing. So let's do that.

OK, so here's a precise version of the intuition. OK, OK, so how do we use the fact that the support of g hat is contained in something like B of 1 over r? There's a nice trick for using this, which goes like this. So suppose I have a function, eta so that eta of c is 1 for all the c in the ball of radius 1 over r. There are many such functions eta, and probably the most useful kind is to suppose that eta is also smooth and compactly supported.

OK, so g hat is equal to g hat times eta, because g hat is supported on this ball, and eta is 1 on this ball. And now, I can take the inverse Fourier transform of this, and I get that g is g convolved with eta check. OK, so this follows from, and is almost equivalent to, this. But it's often a little bit easier to work with.

So next to work with this, we should figure out a little bit what this function eta check is like. So properties of eta check. OK, so for one thing, eta check of x is less than r to the minus 2 in the plane. That just follows from the triangle inequality. To compute eta check, we integrate eta times the complex exponential. And eta has size 1, and it's supported on this ball. So we get this. But in addition to that, if x is large, then we get cancelation.

So the next thing is that eta check of x decays rapidly. So I'll write something precise in a second. Decays rapidly if x is bigger than that. And so the reason for that is eta check of x is the integral eta of C e to the 2 pi i xc dc. And we can bound this integral by integrating by parts. And this is a nice and smooth thing. And if x is bigger than this, we integrate by parts many times, we get a strong estimate for this.

OK, the estimate has the following shape. I start with an r to the minus 2 just to compare to what I had there. And then I have x over r. And I can put here any exponent I like. But negative 1,000 will be fine for us. So if x is more than a teeny bit bigger than r, this thing is very small. OK.

All right. So now let me set C sub r to be the norm of eta check. And so C sub r has bounds that we just computed. So C sub r of x is bounded by r to the minus 2. And it's rapidly decaying if x is much bigger than r. And now, we can state a lemma. So if the support of g hat is contained in the ball of radius 1 over r, then the norm of g of x is bounded by the norm of g convolved with Cr x.

And the proof is basically just what we have. So we started with g hat is g hat times eta. And so g is g convolved with eta check. Now take the triangle inequality. The norm of g is less than or equal to the norm of g convolved with the norm of eta check. And that's the proof. The norm of eta check is called Cr.

OK, and since we're going to eventually be thinking about L2 norms, it is helpful to have a tiny variant of this lemma, where we have the norm of g squared instead of the norm of g. It's sort of convenient to prove it ahead of time instead of in the middle of what's coming. So lemma 2. If the support of g hat is contained in the ball of radius 1 over r, then the norm of g squared is bounded by the norm of g squared convolved with Cr.

OK, so before we do the rigorous proof, let me say something about how I think about this. So this here says that the norm of g is bounded by-- this is sort of like taking an average. It's like the norm of g is bounded by the average value of the norm of G on a ball of radius r. And then if you had that by Cauchy-Schwarz, you would also have that the norm of g squared was bounded by the average of the norm of g squared on the ball.

OK, so that's what we'll write down now. So the norm of g of x squared is bounded by the first lemma by the norm of g convolved with Cr at x squared. If I write out what this convolution is, it's the integral norm of g of x minus y, Cr of y dy all squared. And now, I can do a Cauchy-Schwarz. I want to do a Cauchy-Schwarz to get this guy squared with one of those. So when I do my Cauchy-Schwarz, my first factor will be integral norm of g squared x minus y Cr of y.

And my second factor will just be the integral Cr of y dy. So you get to see that this is a valid Cauchy-Schwarz. The total number of exponents of g is 2. The total number of exponents of C is 2. And that matches what's on the other side. And now, this thing here is bounded by 1, which follows from our bounds for how it behaves, and which goes along with the intuition that this integral is 1. And so this convolution is like doing an average.

OK, and this thing here is just the right-hand side. This is g squared convolved with C. OK, so this is the way we will rigorously make use of the fact that fr has frequency in there. And it corresponds visually to this picture that I think I'll erase now that fr looks of like this. And the places where it's big consists of some blobs of balls of radius r. OK. OK.

So now, we can use our main lemma to prove some projection theory estimates. All right. So our setup is that we have x is a set of unit balls in the two-dimensional ball of radius r. D is the set of directions. So it's a subset of S1, which is 1 over r separated. And S measures the size of the biggest projection of x in here. So it's the maximum over all the thetas in my direction set of the theta projection of x.

So that's our setup. And we're going to try to bound these different quantities in terms of how the directions and how the spheres are clustered. So remember, Nx of r is the maximum over all possible center C of the amount of x in the ball of center C and radius r. And similarly, N sub d of rho is the maximum over arcs and S1 of length rho of the number of directions in the arc.

So that's our setup. And the projection theory estimate that comes from this method, I'll call it theorem 2R, theorem 2 in the real case, says that if we have this setup, then the number of directions is bounded by S big R over x times the maximum over little r of Nx of little r ND of little r over big R over little r squared.

OK, so this is a little bit messy. And we'll eventually make some hypotheses that allow it to simplify it. But just notice for now that this estimate takes account of how the set of directions is clustered. And it takes account of how-- sorry. It takes account of how the set x of unit balls is clustered. And it takes account of how the set of directions is clustered.

OK, so the proof follows the idea of the proof we did in finite fields, and makes use of the main lemma. And here is our setup. So for every direction in our direction set, let's say T theta, is a set of at most, S. We could say exactly S. 1 by r tubes T that cover x. So this projection has size S. Then looking at the fibers of the projections, I get S tubes that cover my set x. OK, then T, my master set of tubes, will be the union over all the directions. T theta. And f is going to be the sum over all these tubes of the smoothed characteristic function of the tube.

OK, so we observe that for every x in x, f of x is at least the number of directions. All right. So now, we use our main lemma and we break up f into these different fr's. We use main lemma. And there are only logarithmically many different choices of r. And so one of them has to make a pretty big contribution at each of these points. And so I can choose r so that-- all right. So I have x times D squared. That's smaller than the integral over x of f squared.

And now, there are only a few r's. So that should be somewhat smaller than a log times the integral over x of r squared. OK, now the first thing I might try here is if you look at the statement of the main lemma, it has an upper bound for the L2 norm of fr. And I could plug in that upper bound, and I would bound this. And I would get something out of that. But that's not the optimal thing to do here.

Well, depending on what we know about x, depending on what we know about this. So to get some intuition, let's look at this picture that I haven't managed to erase yet. All right. So we're trying to understand this thing, the integral over x of fr squared. So let's add an x to our picture. Do I have any more colors? So x in this picture is going to be yellow. And x may look something like this.

It depends a little bit on Nx of r. If Nx of r, the biggest that Nx of r could be is r squared. If it's that big, then x could fill in this whole thing. But if Nx of r is smaller than r squared, then it's sort of like the picture that I drew, and x is only a small fraction of this ball. And in this picture, it would be lossy to estimate the integral over just x of fr squared by the global integral of fr squared, which would include the whole orange ball.

OK, so we're going to try to do a bit better than that by using our lemmas about the locally-constant property of the function fr. All right, I'm going to erase this. All right. I'm going to leave this picture here for our intuition. All right. So we have integral over x of fr squared. I'll just rewrite it as the integral of 1x fr squared.

OK, now if I want to use the locally-constant property of fr squared, it means I could use these lemmas. And in particular, I can use lemma 2, which is sort of well set up to deal with the square. So this is bounded by the integral of 1x times fr squared convolved with cr. And if you write out what this is and do Fubini, this is the same as the integral of fr squared times the characteristic function of x convolved with cr. Do you see that?

Yeah, OK. Let's do it. So the algebra is like this. I have the integral 1x of x. And then I have this convolution, which is the integral fr squared of x minus y cr of y dy. No, let me do it the other way.

OK, so now what happens if I Fubini that? That's the integral fr squared of y integral of 1x of x cr of x minus y dy. OK, it doesn't look quite right yet. And it actually reminds me that there was a little something else that's helpful to say about cr. So you can also add here that cr is radial. Implies that it's symmetric. Yeah, I should probably have mentioned that earlier in my plan. So the reason you can do that, remember, cr was the Fourier transform of some bump function eta.

If your bump function eta is radial, its Fourier transform will be radial. So this guy is radial and symmetric. OK, and then because that's symmetric, I can switch those. So that's the integral fr squared of y integral 1x of x, cr of y minus x. Sorry, this one is dx. All right. So that's the integral of fr squared of y 1x convolved with cr of y.

I feel like there also should be a nice image of what's going on in this integral. We have this-- imagine for a second that this was the characteristic function a set. It's not quite true, but it's not far off either. So I have one set, and I have another set. But before I multiply them together, which would mean intersecting the sets, I fatten out this set here by a distance r. So I take this, I call this set y, take y, I fatten it out by r, and I intersect it with x. And that's kind of the same thing as taking x and fattening it out by r, and intersecting it with y, because I could describe both of them like this.

I'm looking for pairs of points, 1 in x and 1 and y. The distance between them is at most r. If I say it that way, it's symmetric. And that's what's going on here.

Yeah, I guess that's a fancy way of saying, I just took this integral and I wrote it like this. And now, remember that the function, cr, is symmetric. Now the whole thing is symmetric, symmetric in exchanging these two functions. OK, cool. OK, so now, we've written it this way.

And the intuitive meaning of this is I've averaged the characteristic function of x over balls of radius r. And so this is bounded by Nx of r over r squared. So this is the integral of that guy over a ball of radius r. And then I'm taking the average. So I divide by the area of the ball of radius r. OK. So to check this a little bit more carefully, we have to use the bounds that we have for this cr thing, which are over here.

So it's bounded by 1 over r squared on the ball of radius r. That would give the average. Then it also has this rapidly-decaying tail, which to be careful, we should take account of. But what it contributes is smaller than the first term, the main term.

OK, so that's how we bound this. This is now a number. So we can take it out of the integral. And we just have left the L2 norm of fr squared. So we plug in our bound for that. So we have, it's bounded by Nx of r over r squared, you can think of this thing as the density of x at scale r, times the integral of fr squared OK, and then I just plug in the bound for the integral of fr squared. And we'll write that out. So OK, we have number of tubes. We have big r over little r. We have Nt of r. That's this thing. Then we have Nx of r. And then we have r squared.

OK, we are almost done. But what we have here is number of tubes in a little-r-by-big-R guy. And the information we want to be using is the number of directions. So the last step is that the number of tubes in the little-r-by-big-R tube is bounded by the number of directions in a little-r-over-big-R angle times little r. So here's a picture. Here's our little-r-by-big-R tube. Inside of it, I can fit in some tubes in different directions.

So if I pick one direction, I can pick little r of them. That's this little r. But they don't have to go that way. I could also have them go maybe the most extreme directions. I could have them go that way. And this angle here is little r. The angle there is little r over big. So all of the different tubes that appear here, they have to have an angle within a range of little r over big R. So the number of directions is like that. And then for each direction, then the number of possible tubes is at most, little r.

OK, so if you substitute this for that, then this expression becomes the right-hand side of theorem two. Yeah.

AUDIENCE: I remember kind of use the fact that the tubes that are at an angle, that you can only fit less of them in. Does that change the boundary of anything different?

LAWRENCE GUTH: Yeah, OK. So the question was if we knew that in a given direction, these tubes had to have some spacing, so we could also measure the clustering of the tubes in a given direction. So we add some information about that, then maybe this would have to be smaller than little r.

So our setup involves three things. But we only thought about the clustering of two of them. Our setup involves a set of unit balls, x, and we looked at how it's clustered, Nx of r. Set of directions, D. And we look at how it's clustered, Nd of rho. And also, a set of projections, which each have size at most S. But we didn't say anything about how they were clustered. So you could add to this setup something about how each projection is clustered. And you could then use that to improve this thing here. Yeah. So that could be a little homework project if anyone is interested in doing it.

Yeah, OK. Any other questions or comments? OK, OK. So this theorem sort of naturally falls out from using the Fourier method. But it's a little messy looking. And the thing that's particularly messy looking is this maximum over little r, and x, and d, da, da, da. And there's a fairly general situation that we mentioned last time, where that messy bit gets a lot cleaner. I don't know that that's so philosophically important, but I think it's worth mentioning. And this is the way the results usually appear in the literature.

All right. And this will allow us to finally prove the result that we stated last time. All right. So definition, we say x is Hausdorff, or Hausdorff spacing. f and x of r to the beta is bounded by around x to the beta for all beta between 0 and 1.

So when beta is 1, this is trivial. This is always true. But this is saying that at smaller scales beta, this x is not too clumped compared to how it is at the top scale. Yeah.

AUDIENCE: How is the [INAUDIBLE]?

LAWRENCE GUTH: Ah, sorry. Thanks. Yeah, there was one other thing I forgot to do. OK, right. Sorry. Thanks. So OK, so the question is, how did we actually do all this algebra and get the equation that the inequality is actually stated in theorem two? And thank you for the question, because I left out some of the steps.

OK, so now that we have done the main ideas, we're just talking about the algebra, I will erase this. And so what do we have so far? If you put together the inequalities we have so far, we have x times the number of directions squared. That's where we started over here. Then we did some stuff, and we ended up here. So that's bounded by number of tubes times big R over little r times Nd little r over big R times r. That was this thing for there. And x of r over little r squared. And we can cross those guys.

Now we still have this number of tubes. What's that? Well, we have S tubes in each direction, we have D directions. So this is S times D. OK, so now, I've gotten rid of all the extra letters. And we only have left the letters that appear in the inequality we want. And then the last thing you do is you get D by itself. So I take one factor of D, and bring it over there, and cancel the two. Then I take the x and bring it over here. And then what's left is what's there. Yeah.

AUDIENCE: And that's inequality be made sharp with any [INAUDIBLE]? Like in any choice of S, x, and D [INAUDIBLE]?

LAWRENCE GUTH: Yeah. Yes. OK, so the question is, is this inequality ever sharp? And it is sometimes sharp, although not super often. And the cases that I have worked out so far when it's sharp are within this Hausdorff setting. So let's maybe do that. It'll all get a lot simpler. And then I can say what the cases I know are.

OK, so the Hausdorff spacing is a bound on Nx of little r in terms of the size of x that looks like this. And I coined the term, Hausdorff spacing, because it comes up naturally when you talk about Hausdorff measures and things like that. OK, and it's useful in our setting because if x and D have Hausdorff spacing, then when you take the maximum of, actually many expressions, but the one that comes up for us here is this, OK, what happens?

So I assume that both x and D have Hausdorff spacing. It's important to have both of them. Then each one of these things get replaced by some power of x or D. And once you have a power law, so the whole expression will be, or another way to look at it is the whole expression will be some power of little r. A power could be positive or negative, but the maximum will occur either at 1 or at big R.

So this guy basically is equal to what happens at 1. Nx of 1 is always 1. ND of 1 over r is always 1. So we just get 1. And when little r is big R, and x of big r is all the points, MD of 1 is all the directions, And we have that. So it's less nasty. And so we get the following corollary. So corollary says if we have our setup, and x and D have Hausdorff spacing, then the number of directions is bounded by SR over x times 1 plus SR over x times this thing. So SD over r.

OK, now this one has an interpretation. If this term dominates, you can cancel the D's, and it tells you that S is almost r. So we get either S is almost as big as R. R is the biggest it could possibly be. So S is almost as big as it could possibly be. Or D is bounded by SR over x. So that's what we stated last time. The formula is not as nasty as the general formula. Yeah.

AUDIENCE: There's also a factor of log [INAUDIBLE] theorem two, because you're choosing R.

LAWRENCE GUTH: Yep. Thanks. Yeah, so the point from the audience was that there was an extra factor of log R that I forgot to write down. So I'll put this to represent that. Yeah, thanks. OK.

OK, so this relatively clean estimate, it exactly matches our theorem two in finite fields. I won't write it down, but the numerology is the same. And it's a little nicer to look at. It still takes some digesting. And this is sharp in some cases. It's sharp in the case that S is pretty close to R. So because I need this, the smallest I could make S would be something like R over log R. And if you take S around that, R over log R, or R to the 1 minus epsilon, then this is sharp in the example of a grid of points, grid of unit balls.

Yeah, so that's the main interesting example when-- that's the main example that I know when it's sharp. And there are plenty of examples where it's not sharp. But that grid example is important. Yeah.

OK, I guess it's also a natural question, if we go to the regime that's not Hausdorff, this theorem added some junk here to account for things being more clustered. And it's a natural question, is there any scenario where this junk is exactly sharp? And I wondered about that when I was working on this lecture. But I don't know the answer yet. So if anybody wants to try, that's also a nice little project. Cool.

OK, so that was the meat of the class. And that was my discussion that I planned about the Fourier method in the real case. So it's another good moment to see if you have questions to digest. Then we'll start something a little different. Yeah.

AUDIENCE: For the maximum term, could you explain why it has to be a maximum and it's-- the part of the proof that requires taking the maximum over, or instead of, any given r?

LAWRENCE GUTH: So I think the question is, why do we have a maximum over r instead of maybe like a sum over r?

AUDIENCE: Yeah.

LAWRENCE GUTH: Yeah.

AUDIENCE: Just where does the maximum come from?

LAWRENCE GUTH: Yeah. So you could put sum over little r. And the difference isn't very important, because the number of terms is only log of r. And the log of r has got absorbed here. But where did this come from at all? Well, we had this function that's the sum of all these tubes, which describes our projection situation. And then we are-- the key thing is that we applied the main lemma to this function. And the main lemma said you can break up this function into a few--

AUDIENCE: [INAUDIBLE]

LAWRENCE GUTH: Yeah. And so one of those pieces is sort of the important one. And that's the one that appears in the maximum. OK. Yeah.

AUDIENCE: Yeah, I'm wondering. So in the proof, we had to keep track of all the r's. In the Hausdorff spacing case, it seems like just there's only two cases. So I'm wondering if you do have cluster spacing, is it really necessary? Is it still important to do the static division?

LAWRENCE GUTH: OK, so the question is, in the Hausdorff spacing case, it turned out at the end of the day, that only two values of little r mattered in the final inequality, namely 1 and capital R. And the question was, would it be possible to write the proof so that only those two appeared and we didn't have all this rather complicated business of keeping track of all the intermediate little r's? I don't know any way to do that.

I think the only way I understand it is to keep track of all the little R's, and then notice that under these hypotheses, the intermediate R's are bounded by the extreme ones. OK.

OK, so I thought I would say a little bit about the history of these ideas. I haven't been putting names next to the theorems. And the reason is that these theorems are small variations of theorems from the literature. And there are many kind of related theorems from the literature. But let me tell you a little bit about the history. I think it's interesting. I think these are kind of fundamental ideas. And they were discovered independently, or somewhat independently, by different people working on different problems.

So the thing that we have been calling the Fourier method, so let's call this history, the thing that we are calling the Fourier method appears in a bunch places, including the following. The earliest that I am aware of is Linnik's work in sieve theory, which I will introduce sieve theory and tell you what this work was today. So that was in the 1940s.

The second place I'm aware of is Roth's work on the Heilbronn triangle problem, which at some point in the class probably, we'll talk about a little bit, in the '70s. He knew sieve theory. He knew Linnik's work well, and he'd worked on it. But in this problem, he did something similar that's in Euclidean space. And the setup is kind of different. And so I put it on the list.

Then it was something like this was done by Falconer in geometric measure theory. All right. So Falconer essentially proved the corollary that I just erased. And that was around 1980. And then it was done by Vinh in the setting of finite fields. So lines in Fq2. And I'm not positive about this, but I think that many of these people were not fully aware of others of these people.

OK, we also mentioned the double counting method. And the double counting method that we did, so it was done by Kaufman in the context of geometric measure theory in the 1960s. And it was done by Gallagher in the context of sieve theory I think also in the 1960s. I'm pretty sure that these two people did not know what the other one was doing. And I guess it was probably also done in the setting of lines in finite fields in the '60s or '70s. But I don't know. I don't have a reference.

And those people probably didn't know about either of these people. OK, so now, let me tell you what sieve theory is, and how it is analogous to what we've been doing.

OK, so sieve theory takes place over the integers. And the main character in sieve theory is taking the integers and reducing it modulo q for different bases q. OK, so z, of course, is the integers. I'm going to write zq for-- some people write z mod qz. This is the integers modulo q. And there's a basic map pi sub q z to zq, which takes an integer and reduces it modulo q.

OK, this map is going to play the role of our projections. So let me mention what is analogous. So we've been talking about maps pi theta that go from R2 to R, or one-dimensional subspace, or from Fq2 to Fq. And now, we're talking about the map pi cubed that goes from z to z mod q. What do all these things have in common? They are all group homomorphisms between abelian groups.

OK, so what we're going to do in sieve theory is take a set, which is going to be now a set of integers, and think about all the different projections of its integer of the set, and think about how they're all related to each other, and how they're all related to the original set. So I have another little piece of notation. So this thing is the numbers from 1 up to n, which is a subset of the integers. And let me illustrate projection theory. Let me illustrate sieve theory with an interesting example of a set whose projections behave in a funny way.

OK, so let's do an example. My example is going to be basically the set of square numbers. So it's going to be the set of little n squared, where little n goes from 1 up to big N to the 1/2. So it's a subset of the integers from 1 to N. And the cardinality is about the square root of N.

What happens when I reduce this set modulo p? So if p is prime, when I take pi p of x, I get the quadratic residues. And the number of quadratic residues is p plus 1 over 2. So about half.

So that's kind of an interesting and surprising fact. Of course, if p is bigger than square root of N, obviously, the size of this projection is smaller than x. So I might not even get all the quadratic residues if I take a big P. That's not surprising. But I could take p significantly less than the square root of N.

And then when I do this projection, I only hit half of the elements in z mod p, and I hit each of those many times. That's not something that's likely to happen randomly. It's kind of striking. So one sample question from sieve theory is, how big could a set x be and have all of its projections be small like this? And what examples are there? How many sets like this can you find?

All right. So using his ideas about sieve theory, one of the things that Linnik did was to answer some of this question. So theorem of Linnik in the 1940s says, if I have a subset of the integers from 1 to N, and if pi p of x is at most, p plus 1 over 2 for every P. So there's nothing so special about this. We could put other things here, but I'm just using this to compare with the squares. Then the conclusion is that the size of x is at most around N to the 1/2.

So it says that the square numbers are just about the biggest set with this property. Cool. So we will give two proofs of this. We'll give a proof using double counting, and we'll give a proof using the Fourier method. There are lots of hard open problems in sieve theory. And one of them is to understand what the examples are. So the only known examples, only known examples where this theorem is roughly tight are the square numbers and close cousins.

So sorry. It's messy. The square numbers, and close cousins of the square numbers. You could replace this with another degree 2 polynomial in little n. And it's not hard to see it's kind of equivalent. And those are all the examples. OK, but people know basically nothing about classifying the examples. Cool. Cool.

OK, so we'll give two proofs of this, one based on double counting, and one based on the Fourier method. Double counting, I think, we can do today. And the Fourier method, we'll do next time, next week. And first of all, I think that these results are pretty cool. But I also hope that you see that they're really extremely parallel to the results in geometric measure theory that we've just been talking about. OK.

OK, let me make a little remark before we start. So I'm going to focus on projections mod p, where p is a prime. You could also reduce mod q, where q is not a prime. And all of the ideas still work. It's a little bit more complicated because if you take two numbers, q1 and q2 that share a common factor, then reducing mod q1 and reducing mod q2 are kind of related to each other. And so if you write everything down, you have to keep track of that. That's just an extra thing to keep track of. And not so difficult, but it makes everything a little longer, and more complicated.

And sometimes, you might want to do that, but in applications. But just to show the ideas, I'm going to focus on mod p. OK. All right. So here's the double counting method. So I'll call it theorem 1S. So it says if x is a subset of the numbers from 1 to N, and D is a set of primes less than or equal to N, and for every p in our set of primes, pi sub p of x is less than or equal to S, then we get a conclusion.

Either x is bounded by 2S, or the number of directions is bounded roughly by S. So I guess this is most interesting if the number of directions is significantly bigger than S. So if we look at the squares, so suppose I know that pi p of x is bounded by p plus 1 over 2. So that's my S. And I need to do this-- yeah, let's see. Well, this may not quite work for this. Right.

Actually, I'm not sure if this will imply Linnik's theorem. Linnik proved his theorem with the Fourier method. Let's come back to that next time. I think we have just enough time to do the nice double counting argument and then do the Fourier next time. OK, so proof, we're going to count coincidences. That's the set of x1, and x2, and x, and p in our set of directions, or a set of primes so that pi p of x1 equals pi p of x2.

OK, so we'll count this two different ways. One is a lower bound. So star is at least for every direction, when I do this projection, there are only S images. So a typical image has x over S preimages. Square that for my choice of x1 and x2, and then multiply by S for the number of choices of the projection. So that's x squared D S inverse.

OK, now what's our upper bound? So now, imagine fixing x1 and x2. And think about how many P's can satisfy this. OK, so if I have pi p of x1 is equal to pi p of x2, that tells us that p divides x2 minus x1. And there can't be that many different primes that divide x2 minus x1. x2 minus x1 has size n. So it could be, at most, log of N that do this.

So star is upper bounded by, OK, so there is a case where x1 equals x2. So that gives me x times the number of directions. And then if x1 is not equal to x2, I have x squared choices for x1 and x2. But I have only log of N choices for the direction. OK, and now, I'll compare these to each other.

OK, so I have x squared times D times S inverse is less than x times D plus x squared times log N. So there are two cases. If this one dominates, then I would have that x is bounded by 2S. Or if this one dominates, then the x's disappear, and I get that D is bounded by log of N times S. And that's the theorem.

OK, so let me give you an example of this and argue that this is-- well, I'll give you an example, a flavor of what this is telling us. So suppose that pi p of x is bounded by N to the 2/3 for 10 log n times n to the 2/3 different p. So if that's true, then the theorem tells us that x is bounded by N to the 2/3.

So these primes might have size around this big, or a bit bigger. So these primes are much less than N. So each projection, the fact that each projection one at a time is small, it's not that convincing in terms of x being small. But knowing that all of these projections, or this large number of projections is small, it means that x is really that small.