[SQUEAKING]

[RUSTLING]

[CLICKING]

LAWRENCE GUTH: So the first thing I promised to do today was to do the last problem on the first problem set with you all. And so for context, this was a pretty difficult problem and a lot of people didn't get it fully right or made a mistake. And actually, even the people who did get it right, they tended to write something that was pretty long and complicated. So I spent a while thinking about it. I think I could have set up the problem with a little more direction.

But anyway. But I will explain it to everybody and try to say what were the issues and what do you need to know to do it cleanly. OK. So the background of this problem is that we had done a bunch of stuff in class about lines in the plane, say, an Fq2. And the problem was to try to do something similar for lines in Fq3, but it is a little more complicated. So let me first remind everybody what we did for lines in Fq2, and then we'll go up to three dimensions.

OK. So general setup. So first of all, if we have a function on Fqd, then f0 will be the constant part. So f0 of x is 1 over q to the d sum y in Fq to the d of f of y. And then f sub h f high will be the rest of it. So f is f0 plus 1. Oh. So first we studied lines in Fq2. And this is just going to be a recap of what we did in class. So we noticed.

So let's say that L1 and L2 are our lines or characteristic function of lines. OK, so there are two cases. If L1 and L2 are parallel, then you can easily see that their inner product is zero. Parallel means they don't intersect. So you get this. OK. And you could also look at the high part. And if you look at the high part, it will be negative. The inner product will be negative L1 h2h will be negative.

It's easy to compute. There's something I could say out loud. One always has that fg is f0 g0 plus f high g. You can see that from the Fourier representation of these things. So in this case, the inner product of L1 and L2 is 0. L1 0 and L2 0 are both positive so their inner product is positive. So for this all to fit together, this inner product should be negative.

So that's what happens if they're parallel. Now at the time, parallel felt a special case. Two lines are usually not parallel. Once in a while they're parallel. So maybe the more important case is if they're not parallel. L1 and L2 are not parallel. So if they're not parallel, then they intersect. And therefore, the sum L1 of x, L2 of x is 1. But if you take the high part, L1h, L2h, you get 0. So the high parts are orthogonal.

OK. And we discussed a couple of ways of seeing this. So one way to see it is based on the Fourier transform. So if you take L hat of ksi, its norm is q times characteristic function of L perp. And so in particular, the support of L1 hat intersected with the support of L2 hat is L1 perp intersected with L2 perp, which is just the origin if the lines are in different directions.

This is all under L1 and L2 are not parallel. L1 and L2 are not parallel, L1 and L2 hat also are not parallel, and they intersect only at the origin. And then by looking at the high part, we leave out the origin. So there's nothing left where they intersect. And that gives us the orthogonal. OK. So once we had this, we used it to prove a proposition.

That said, if f is the sum over some set of lines of L, then f high L2 squared is bounded by the sum over the lines of L high or not high L2 squared, which then you could work out easily as bounded by the number of lines times q. This is almost an equality. OK. So this was really the key thing. And this came from orthogonality. You can think of it as taking this sum and expanding it out. So this is the sum L1 and L2 in our set of lines of the sum L1h L2h. And for each of these, we plugged in that.

OK. So now let's go to three dimensions, see what's the same and what's different, and most importantly, what should be the analog of this proposition. OK. So lines in Fq3. So they could still be parallel or not parallel. So if L1 and L2 are parallel, then everything is still true. They're parallel, then this is 0. And if we wanted to take the high part, it would be negative. Just the same as before.

OK. What if they're not parallel? If they're not parallel, there are still two possibilities. They could intersect each other or they could not intersect each other. So I guess there's subcases. If they intersect, then the sum on X L1, L2 would be 1, intersecting one point. That looks like before. And then we could take the high part.

But the high part isn't 0. So this is an important thing that's different. So one common mistake was not to do this computation carefully and just assume it should be 0 like it was in two dimensions. But it's not. There's not that much difference between 1 and 1 minus 1 over q, so it's not clear that we're accomplishing a whole lot by focusing on the high part. Yeah?

OK. So to see this, we could do this computation. It's elementary but it's a little bit tedious. But let me see if I can say something conceptual. So the reason that this happened here is that the two high parts are independent of each other as functions. That's a high brow way of saying what's going on.

And that's not true anymore. If you have two lines that intersect in three space, it's not independent of each other. So if you know that you're in the first line, then you're much more likely to be in the second line than if you were just a random point. OK. I don't know if that's helpful, but you could also just check that that's how it goes. They could also not intersect.

If they don't intersect then everything is the same as the parallel case actually. So that's going to be 0. And the sum of the high parts, the inner product of the high parts will be negative. If you do it carefully, it's negative 1 over q. Same thing would be true here. OK. So another direction that this suggested to some people is that we ought to try to think how many pairs of lines intersect, as opposed to the pairs of lines that don't intersect.

That is a totally reasonable way of thinking about this from the beginning. But it's not obvious how to figure that out. It is possible for every pair of lines to intersect. There are two significant ways that could happen. One way is you could take all the lines through the origin. Just pick a point and take all the lines through that point. Then every pair of lines will intersect at that point, have a porcupine shape.

The second way that almost all the lines could intersect is to choose them all in a plane. So if you fix a plane and then you choose lines in the plane, a few of them might be parallel, but almost every pair of lines in the plane will intersect. Anyway, so people tried to explore this a little bit, but it's not so clear how to put a bound about how many pairs will intersect

OK. So then another thought is that instead of focusing on this, we should focus-- we should look at the Fourier transform and just see what the Fourier transform tells us. So the Fourier transform does have a nice description in any dimension, like you did on the first part of the problem set. And it tells us something. So let's think about the Fourier transform. The Fourier point of view.

OK. So if we take the Fourier transform of our line and look at its absolute value, it will be q times the characteristic function of L perp. But notice, so the L is still a line. But since we're in three dimensions, L perp is two dimensional. OK. So then if we were to look at the-- so if L1 and L2 are not parallel. Not parallel. If I look at the support of L1 hat intersected with the support of L2 hat, well, that's L1 perp intersected with L2 perp.

So I have a two plane and I'm intersecting it with a different two plane. And that will make a line. That's a line. OK. So a third common mistake was to not do this so carefully and think that this intersection was just zero. And if the intersection was just 0, when you take the high part, you would get rid of the 0 and that would show that L1 and L2 were orthogonal, which is not actually-- L1 and L2 hat orthogonal, which is not actually the case. OK. OK. Cool.

All right. So to prove some kind of an estimate about how these lines intersect, we're going to try to combine this information and this information. So if they're parallel, they actually are orthogonal, which is very good. If they're not parallel, they're not fully orthogonal. But the Fourier transform of-- the intersection of Fourier transforms is kind of small and special. We'll try to use that. OK. So here's the analogous proposition. And I think it would have been helpful if I had stated this in the problem set.

So here's how it goes. So if L is a set of lines in three dimensions in Fq3 and f is the sum of these lines, then f L2 squared is bounded by-- so what we had before was something like orthogonality. Some L and L of fL or-- and this is with the high part. The high part L2 squared. And that's not true anymore. But it is true if we put an extra factor of q.

OK. And so then that is around the number of lines times each one of these is around q, and there's another factor of q which is around there. OK. So if you compare these, this is like an orthogonality proposition. This is clearly much weaker than orthogonality because we have a big factor here. Nevertheless, it's stronger than nothing.

Even though there's a big factor here, this proposition has some meaningful information in it. OK. So let me explain how to prove this proposition, and then we'll digest it and argue that it does tell us something interesting, even though it's a lot weaker, in some sense, than the two dimensional proposition.

OK. So how do we prove this? All right. So for parallel lines, we have really good information. So let's first think about all the lines in a given direction and add those up. And then after that, we'll add up the different directions. So we'll label the directions by w. So w will be a one dimensional subspace direction. And then Lw will be the lines in our set of lines that are parallel to w.

And we can also make an fw. That's the sum over the lines in Lw of L. So f itself is the sum of the different directions of fw. So the first thing that we'll do is prove some estimates about fw. And this part really is not that difficult. We're adding up functions with disjoint supports. It's not that complicated, what happens. OK. So fw L2 squared is the size of w. Sorry. The size of Lw times q.

Why is that? Because the-- so proof. So sum fw squared is the sum L and Lw sum on x L of x squared. But these lines are parallel. They're disjoint from each other. So there's only-- at any given x, only one of them is making an appearance. So this is sum L and L sum on x out of x squared. And then that's the right-hand side.

It's not clear yet why we would want to look at the high frequency part, but if we did, it wouldn't change much. So fw high, L2 squared is less than fw L2 squared. And so the proof is that this is true for any function. We have it over there. For any function gL2 squared is always g0 L2 squared plus g high L2 squared. So it's at least as big as g high L2 squared.

OK. You can do it with Fourier transform. You could also do it by hand. Let's call it an exercise. OK. Now, the more tricky and interesting point is what happens when we add together the different w. And when we do that, the way their Fourier transforms behave will come into play.

All right. So I guess that this is lemma three. But I'm going to put a star because this is the important lemma. This is where the interesting stuff is happening. OK. Important lemma says that fL2 squared is bounded by q plus 1 times the sum on w of fw L2 squared. OK. Here's the proof. Oh, I'm sorry. And now I need to put high everywhere.

All right. So f high L2 squared is, by Plancherel, it's the sum on the nonzero ksi f hat of ksi squared. So remember that f0 is the zero frequency part of f. f high is all the other frequencies in f. And so this is Plancherel. Oh, OK. And I guess because of the finite field normalization that I always have trouble remembering, there should be this.

OK. Now, f is the sum of fw. This is 1 over q cubed times the sum and the nonzero ksi, sum on w of fw hat of ksi squared. OK. Now, in two dimensions, we could have done this and gotten to this point. And then we would have been very happy because the different fw hats have different supports. They're supported on different lines through the origin. So we would have orthogonality.

The way it would have worked in two dimensions is that for any particular ksi, there is only one w which is nonzero in the sum. And so we could take the square inside. OK. But that's not true anymore. So fw hat is supported on w perp, and in any given direction, any given ksi, there are many different two planes that contain that ksi.

So let's write that down. So we're going to note that the support of fw hat is w perp. And so for any ksi, the fw-- the set of w so that f hat w of ksi is nonzero is contained in the set of w so that w is in ksi perp. Yeah?

AUDIENCE: We still need the high part, right? On the f hat.

LAWRENCE GUTH: Yeah. So the question is-- I think the question is-- so on the left-hand side, I put the high part of f. And the question is, do we need the high part here? Is that the question?

AUDIENCE: Just the zero. Not that one.

LAWRENCE GUTH: So we don't need that, but I should have said it clearer. So let's all remember that if you take the high part of f hat ksi, what you get is f hat of ksi if ksi is nonzero and zero if ksi is zero. This is f high hat. OK. So for any given frequency ksi, I only need to include the w's in ksi perp. So actually, I can write this.

And there aren't as many. There's more than one of them. But there is a non-trivial bound how many are terms are in this sum. So the number of w in ksi perp. So this is a two plane. We're talking about how many lines through the origin there are in this two plane. That's q plus 1. So there are something like q squared different w in total. But for any given ksi, we're only including q of them in the sum here.

That's not as good as orthogonality but it actually is helpful information. And the way we can exploit that information is by Cauchy-Schwarzing this. So I'm going to think of this as the sum of this thing times 1. Of course I didn't change anything. And now I can apply Cauchy-Schwarz. 1 over q cubed sum over the nonzero ksi. The first term will be sum w in ksi perp f hat of w of ksi norm squared. And the second half will be sum w and ksi perp of 1 squared.

OK. And this is q plus 1. So we bring that out to the front, and after that, what we have left is a sum that just involves each fw one at a time. So it's q plus 1, 1 over q cubed sum on the nonzero ksi. Actually, I could put-- well, I'll put the sum on w here. fw hat of ksi squared. But I can switch the order of these sums.

And now if I Plancherel back, I have q plus 1 times the sum on w fw high L2 squared. And that's the statement. OK. Any questions or comments? In that case, maybe I have a question for you. In this proof what was the purpose of taking out the 0 frequency? What did we gain by doing that? Yeah?

AUDIENCE: The zero frequency is equal to 0 every plane contains ksi equal to 0. So the output would be [INAUDIBLE].

LAWRENCE GUTH: Good. Good. So the comment is that ksi equals 0 is special because-- and when ksi is 0, every w perp goes through 0. So there are q squared different w's, and every one of them has a contribution at 0. So if we wanted to do Cauchy-Schwarz, we would have q squared terms here. But at every other ksi, we have only q w perps going through it. And so we have to pay only q over here. That's what's different about 0 and not 0. Good.

OK. So now the proof of the proposition over there is just assembling these lemmas. So the proposition says that I want to bound the L2 norm of f high. How do I do it? Lemma three says the L2 norm of f high is bounded by q plus 1 times the sum on w of those, and lemmas one and lemmas two tells me how big each one of these is. If you put that all together, you get proposition two. OK. OK.

Great. OK. So now let me erase this a little bit and let's think about what proposition two tells us about lines in three space. OK. So actually, let's first review what proposition one tells us about lines in the plane. So f is a sum of lines. f0 is always the number of lines divided by q, and so f0 L2 squared works out to be the number of lines squared. It's not hard to check.

And proposition one, it tells us that f high L2 squared is bounded by the number of lines times q. So if you put these together, you see that if the number of lines is much bigger than q, then f is mostly constant. So then L2 norm of f high is much smaller than the L2 norm of the constant part. The sum of lines looks almost constant with some small perturbations. OK. What's the analog of that in three dimensions? Lines in Fq3.

OK. So in three dimensions, f0 is the number of lines over q squared because each line occupies a 1 over q squared fraction of space. And so then you can compute the L2 norm of f0 squared. It's not difficult. It's number of lines squared over q. And on the other hand, we know that f high L2 squared is at most the number of lines times q squared. So you can put these together.

If the number of lines is large enough, then this is way bigger than this. If you do a little bit of algebra, you see that if the number of lines is much larger than q cubed, then the high frequency part is much smaller than the constant part. So if you add up more than q cubed lines in three dimensional space, you'll get something that's mostly constant with some small perturbations.

OK. Now these are both the correct thresholds for this behavior. Why? So in two dimensions suppose I add up less than q lines, say q over two lines. Well, they only fill half of space. So there's no way that thing is going to look constant. So this will certainly not be true with q over two lines. All right. Now in three dimensions suppose I add up less than q cubed lines, maybe half of q cubed lines.

All right. Now this is a trickier wicket. It depends which lines I pick. If you were to pick half of q cubed lines randomly and add them up, it would be quite close to constant. But if you pick them cleverly, you can make it not look like a constant. And here's how to do it. So example of 1/2 q cubed lines in Fq cubed.

OK. Here's what I'm going to do. I'm going to pick 1/2 q affine two planes Capital Pj. And then in each capital Pj, I pick all the lines, which is basically q squared lines. So a two dimensional plane has about q squared, q squared plus q. So you put this all together, that's about half of q cubed lines.

Now f, which is the sum of L and L, is supported on the union of the Pj. And the size of the Union of the Pj is clearly at most half of the space. So the size of the union of Pj is at most 1/2 q, the number of j times q squared number of points per plane 1/2 q cubed. So this thing is only supported on about half of space. There's no way that it's approximately constant. Cool.

OK. So in a certain sense, this is the right threshold, although you could also imagine asking some more refined questions. Cool. OK. I guess looking back at the homework, I mentioned a few things that people tried that didn't work. And I think the most common thing that people had a tricky time finding was that they should first group the lines by their direction and have these fw's, and then group the fw's together to make f.

Yeah. So I think I could have done a better job giving direction to look that way, look in that direction. But in hindsight, why was that a good idea? There's a saying that I like about trying to learn stuff, or trying to figure stuff out that I learned from a Tai Chi teacher. They said, do what's easy first.

So if you are you trying to, in math, for example, add up a whole lot of things and understand what the sum is like, if there are some pieces that you can add up and you know what those are like, it's usually a good idea to do that first. And adding up parallel lines is fairly easy to understand, and so it won't always help. But sometimes it's helpful to do what's easy first.

Actually, same thing happens in Fubini when I teach analysis at various levels. When is it a good idea to do Fubini to switch the order of integration or switch the order of summation? You want to put the sum that you do first to be the easy one. That's almost always the right way to organize it. OK. So that was the recap of the homework.

OK. Cool. Now we come to the main topic of the day, which is doing some projection theory in Euclidean space and thinking about how projections makes functions smoother. So if you take a function in R to the D and you project it onto something lower dimensional, well, if you just pick one lower dimensional projection, it might not be very nice.

But if you look at all the lower dimensional projections, most of them will be smoother. And we're going to investigate that phenomenon. And I hope you'll see that it's very parallel to the large sieve theory from number theory.

All right. Projections and smoothing. All right. So here's our setup. So f is now a function on R to the D. And v is going to be a subspace. And recall, by the projection of f, what I mean is the integral over V perp f of y plus z d volume v perp of ksi. So this is a map from functions on R to the D to functions on v.

OK. And I'm going to mostly focus-- everything works for all the dimensions, but for ease of notation, whatever, I'm going to mostly focus on projecting to one dimensional subspaces. So if theta is a direction in the circle pi theta of f is going to mean pi-- abbreviation for pi span of theta of f. Direction. Look at the line in that direction.

All right. So here's a theorem that says that if you start with a function, there's very little regularity. Most of the projections are even smooth. So theorem says if f is an L2 function on R to the D, and the support of f is in the unit ball, then if I integrate over this sphere and I take the projection onto the line in the theta direction of f and I take it CK norm, then that is bounded by the L2 norm of f, provided that the dimension is big enough.

So I add the hypothesis at the dimension, which would be dimension D minus 1 over 2 should be bigger than 1/2 plus k. You can unwind that a little bit. It says D is bigger than something. So for example, D is 7. Then you can take k equals 2. So in seven dimensions you take just an L2 function. If you project it down to one dimension in a typical direction, you'll get something that's C2.

OK. I'm actually not sure to whom to attribute this theorem. I'm in the process of doing some historical research and maybe I'll know next class. All right. So this theorem is also based on the Fourier transform and the way the Fourier transform interacts with projections. So now let's think about how that works.

Fourier transform and projections. All right. So f is a function on R to the D. It has a Fourier transform. We've talked about it pi v of f is a function on v. It also has a Fourier transform. I just remind you how it works. So v is a vector space. It basically works the same way as on Euclidean space. OK. So suppose I have a function g from V to C. Then g hat will also be a function from V to C. And g hat of ksi will be the integral over v of g of y, e to the minus 2 pi iy dot ksi dy. But dy means with the volume form for my vector space V. So that's the Fourier transform. Yeah?

AUDIENCE: What do you add the 2 pi key factor here? That's normally not include a general Fourier transform. Is that just a convention?

LAWRENCE GUTH: Yeah. So the question is, why is there a 2 pi here? There are a couple of different conventions for the Fourier transform. One of them has a 2 pi here. The other one has a 1 over 2 pi to the D in the front. Those are, as far as I can tell, equally good. And they both involve some slightly annoying factors of 2 pi. There's another convention where you write neither of those things. That, strictly speaking, is not correct, but it's convenient and it usually is OK.

OK. Now we can do Fourier inversion where we recover g of y. Do Fourier inversion. That's the integral over V g hat of ksi to the 2 pi iy dot ksi d volume V of ksi. OK. So that's how the Fourier transform works on a subspace. And now, what does projection do to the Fourier transform? There is an elegant formula lemma.

Call it a dictionary lemma. Translates between the projection on the physical side and projection on the Fourier side. It's analogous to something that I called the dictionary lemma when we were doing large sieve. All right. It says the following thing. If I have a function on R to the D, then if I project it to V and take its Fourier transform, it's just f hat of ksi for all ksi in V.

All right. So let's process a second. So this is a function on v. Its Fourier transform is only defined on V. So this left-hand side only makes sense if ksi is in V. But it's equal to f hat of ksi, and f hat makes sense for ksi anywhere in R to the D. So here's the proof. Let's just expand out what everything means.

So the left-hand side is the integral over V pi vf of y, e to the minus 2 pi i ksi dot y d volume v of y. But what is the projection? Pi vf of y. That's also an integral, and it's an integral over V perp. So that's the integral over V, integral over V perp, f of y plus z, d volume v perp of z. So that was this thing, e to the minus 2 pi iy dot ksi, d volume v of y.

All right. So now notice that altogether, if you put this stuff together and that and that together, we are integrating over R to the D. And so more precisely, we can do of a change of variables here. We can say x is y plus z and dx is d volume v of y, d volume v perp of z. So this is f of x. This stuff is just dx.

Now who is this? Well, because z is in V perp and ksi is in V, z dot ksi is 0. So e to the minus 2 pi iy dot ksi is the same thing as e to the minus 2 pi i y plus z dot ksi. Changed it by z dot ksi, which was 0, which is e to the minus 2 pi ix. So I can rewrite everything in terms of x and it's just my integral f of x, e to the minus 2 pi x dot ksi dx. All right. So that's f hat x ksi, and that's what we wanted to prove. All right. Good. OK.

So let's see. How might we use this? What might this tell us? So we're interested in this projection thing. And the projection plays very nicely with the Fourier transform. And so anything else that plays nicely with the Fourier transform, we could try to figure out how it interacts with projection. And one of the fundamental things that interacts nicely with the Fourier transform is the L2 norm, because we have Plancherel.

So we can write f L2 in terms of f hat. And f hat is closely related to pi v of f hat. Let's just see what happens. What does that tell us? So I'll call this relating f L2 and pi v of f via the Fourier transform. And I'm going to focus on pi theta of f, because that will make this the easiest.

All right. So f L2 squared is f hat L2 squared. Write that out. It's the integral over R to the D f hat of ksi squared d ksi. OK. And the different-- if I take pi theta of f hat, that will correspond to f hat on different lines through the origin. So that suggests-- let's separate this out in different directions by writing it in polar coordinates.

So this is equal to the integral over the d minus 1 sphere integral from 0 to infinity, f hat of r theta squared r to the d minus 1, dr d theta. OK. And let me do this little trick to also include negative r, so that once I fix theta I have a nice integral from minus infinity to infinity.

All right. So this guy is related to pi theta of f by this lemma. So this is 1/2 integral over Sd minus 1 d theta. So I'm averaging over all the different directions. And what am I averaging? Integrating from minus infinity to infinity, pi theta f hat of r theta. r to the d minus 1 dr.

What is this thing inside here? At first you might have thought you might get the L2 norm of pi theta of f squared. So that would be the L2 norm of pi theta of f squared. This has this extra factor here which heavily weights large frequencies and gives a very low weight to small frequencies. Yeah?

AUDIENCE: That's Sobolev norm.

LAWRENCE GUTH: That's right. This thing is a Sobolev norm. So What naturally appears here is Sobolev norm. And now let's pause and we'll review Sobolev norms and some important facts about them. All right. OK. So definition. If I have a function g on a vector space, like a subspace of-- a space with the Euclidean norm. Yeah?

AUDIENCE: This is about the dictionary lemma. I'm wondering what class of functions that applies to. Because if f is an L2 function so f hat is L2 function, how is its value on a measure 0 subset being meaningful?

LAWRENCE GUTH: Yeah. OK, great. So the going back to the dictionary lemma. The question was, for which class of functions does this actually make sense? And if you look at what I wrote on the board, it's a little suspicious. It just says that f is a function. And if f is just a function, actually, none of the sides of this equation in general makes sense. One thing that would be clearly sufficient is if f was a Schwarz function.

If f is a Schwarz function, f hat is a Schwarz function, the projection is a Schwarz function, this is a Schwarz function. So that's all OK. If f is just L2, it's not immediately obvious how all of these things would be defined. So let's think about this way for now. It's fairly often the case that we can prove things about general L2 functions by first proving them for Schwarz functions, and then proving-- then using the fact that Schwarz functions are dense in L2. OK. OK. Good.

OK, so we were remembering what is a Sobolev norm. So if I have a function on v, then g HS dot squared is the integral over v. I'll take the norm of g hat of ksi squared. And then I'll put here the norm of ksi to the 2S. This makes sense for any real number S, and it has something to do with the smoothness of g. And in particular, if S is a natural number, then it has an interpretation in terms of L2 norms of derivatives.

So g H1 dot squared is the sum over different directions j of the j-th derivative of g L2 squared, at least approximately, maybe exactly, and then higher values of S-- higher values of S correspond to taking more derivatives. So I'll just put one more and you'll see the pattern H2 dot squared is around the sum over j1 and j2 of the second derivative in the j1 j2 directions of g, L2 squared, and so on.

OK. And if S is not an integer, then there's not a direct interpretation of this form, but it is still kind of measuring how smooth the function g is. There is also a regular. So this is the Sobolev norm with a dot. It's called the homogeneous Sobolev norm.

There's a version without the dot, which is defined like this. g HS squared is the integral over v g hat of ksi squared 1 plus the norm of ksi squared to the S d ksi, which is roughly the same as if you took out the biggest power. So it'd be 1 plus ksi to the 2s, and the one contributes g L2 squared, and the ksi to the 2S is this thing. So it contributes HS dot squared. OK.

Cool. So there's a well-known set of theorems called the Sobolev embedding theorems, which imply that if g is in a Sobolev space for a big enough S, then it is continuous, and if even bigger S, then it's C1, and so on. So I will write a summary. Sobolev embedding. If S is bigger than 1/2 the dimension of V, then g L infinity is bounded by g in HS.

And if g is in HS, then g is continuous. If S is bigger than some k plus 1/2 the dimension of V, then the k-th derivatives of g are bounded by the HS norm. And if g is in HS, then its CK. OK. Any questions or comments about Sobolev spaces?

OK. So now that we have recalled that, if we look back here, we will recognize what appears here as a Sobolev norm of the projection. So the L2 norm square root of f is this integral. And this thing here is the HS norm of pi theta of f. So this is 1/2 integral over the d minus 1 sphere pi theta of f H d minus 1 over 2 dot squared d theta.

OK. Where did this come from, and why does it have a dot? This r is the norm of my frequency ksi. So this is the norm of ksi to the 2S. So S is d minus 1 over 2. And it's homogeneous. We don't have this extra one. So there's a dot. OK. Cool. Cool.

OK. So now you can probably have a good guess of how we would like to prove this theorem. The L2 norm controls an integral like this with the Sobolev norm. The Sobolev embedding tells us that the Sobolev norm controls the CK norm. So that's the plan. OK. But technically, there is a small technical issue, which has to do with the dot or the no dot. Dot and no dot, homogeneous, not homogeneous are both natural and important.

Homogeneous one will occur naturally in some places. It naturally occurred here. On the other hand, Sobolev embedding is a theorem that requires the non-homogeneous Sobolev space. So we need to do a little bit of work to get between.

OK. So now we can do the proof of theorem. All right. So integral SD minus 1 pi theta f CK d theta is bounded by integral SD minus 1 pi theta f H d minus 1 over 2 without the dot squared d theta. So this is Sobolev embedding. And the hypothesis about d up there is just exactly the hypothesis that you need for this to be true when you plug in Sobolev embedding.

All right. Now let's open up the definition of this thing. And we'll almost see the f L2 norm squared. But not quite. So this is around integral of SD minus 1, integral from minus infinity to infinity, f hat of r theta squared. And we'll have 1 plus r to the d minus 1 dr d theta. So we don't have a dot that requires us to put this extra one that we didn't have before.

All right. So you have a sum of two pieces. In real analysis, we usually think about which of the two pieces is bigger. So if r is bigger than 1, this piece dominates. And if r is smaller than 1, that piece dominates. So the integral SD minus 1 integral r bigger than 1 f hat of r theta squared r to the d minus 1 dr d theta. That's called that term number one. Plus the integral SD minus 1 integral r smaller than 1 f hat of r theta squared 1 times 1 dr d theta. Let's call this piece two.

OK. So piece one is clearly smaller than what we had before, which was f L2 squared. So piece is 1 smaller than f L2 squared. Piece two. OK. In piece two, I think it's no longer really helpful to think of this as being related to polar coordinates. We're integrating over a sphere across an interval, and that's a finite volume. And so what I'd like to put here is just that it's bounded by a constant, which is the volume times the biggest possible size of this.

OK. So how big is that? So f hat of ksi is bounded by the integral f of x, e to the minus 2 pi i x dot ksi dx. And the most common approach, the easiest approach is to use the triangle inequality. So we have this. So now we have the L1 norm of f. So we have two pieces. One piece is controlled by the L2 norm of f. The other piece is controlled by the L1 norm of f.

I could have written the theorem that way, but I wrote it a little bit differently. I thought it was cuter. I had the support of f is in the unit ball. If the support of f is in the unit ball, then we can control the L1 norm by the L2 norm. So since the support of f is in the unit ball, integral of f of x dx looks like that. It's bounded by the integral of f of x squared to the 1/2 times the integral over the unit ball of 1 to the 1/2. So that's bounded by f L2.

OK. Using the support, I can bound both 1 and 2 by f L2 squared. OK. That was what I had thought of as the proof of the theorem. But now is a good moment to come back to the question that was asked earlier. What do you actually have to assume about f for all of this to make sense? OK. So let me add some fine print, write what we actually proved.

All right. So the fine print is for the time being, suppose that f is a Schwarz function. We needed f as a Schwarz function to justify our formula about the Fourier transform of the projection, and in a few other places to make sure all these Fourier transforms make sense. If you assume that f is Schwarz, everything we wrote makes sense and we get this estimate. And this is an inequality.

And then it is now possible to do some extra work. If you have a function that's not Schwarz, just a function in L2, you could have a sequence of Schwarz functions that approximate it. And for each one of those approximating functions, this bound would be true. And then with some work, you could show that the bound is also true for the limiting function f. OK. So I'll put a remark. Can remove the hypothesis of f Schwarz. OK. So any questions or comments about the proof of this theorem?

AUDIENCE: So it's a condition that the support of f is essential to this theorem?

LAWRENCE GUTH: Yeah. OK. So the question is, is the assumption that the support of f is in the unit ball essential? What would happen if we removed this assumption? Yeah. So looking at the proof, we could-- let me get another color. Looking at the proof, we could, if we want to, remove this assumption. And then we would need to add the L1 norm, as well as the L2 norm to deal with.

Due to my brilliant neighbor labeling scheme, term one is bounded by the L2 norm and term two is bounded by the L1 norm. OK. I haven't thought about it carefully. You could ask if we did this, would it still be true? I suspect that's not true, and the natural example to look at is a function on a really big ball. I haven't thought it through carefully. It's a good little project to figure out what happens.

OK. OK. So the big picture takeaway of this theorem is we take a function in high dimensions, which is not very regular, and when we project it to low dimensions, most of the projections are much more regular. And I was trying to think of places that this relates to or we might have seen something like this before. And one place where we see something like this is the central limit theorem in probability. So let me remind us how that works.

OK. So suppose that Xj are random variables which are uniformly distributed in 0, 1 and there are a lot of them going up to n. OK. So in the central limit theorem, we might think what happens if we add up all these random variables. So then we could look at the sum of the Xj. This will be a random variable. Actually, let me put it-- not that important. But I'm going to put minus 1/2, 1/2. The central limit theorem is a little bit nicer.

OK. So what happens when I add them up? When I add them up, the central limit theorem says that the sum of these random variables is approximately a Gaussian with mean 0 and a standard deviation, something like square root of n. And notice in particular that it's quite smooth.

OK. So the probability, the joint probability density function of these Xj's. Xj. Is f of X1 up to XN. So the characteristic function of the cube. So this is a typical-- this is an example of a function that's in L2 but is not continuous. OK. Now what is the probability density function of this thing? Well, we have a map. I guess the summation map goes from RN to R and it just takes X1 up to XN and maps it to X1 plus dot, dot, dot plus XN. Add them up. It's a linear map.

So the probability density function of the sum of the Xj's is the pushforward of f dx. And this is a slightly different language, but up to a small change of variables, it's the projection of f on the main diagonal. So actually, we can make this look nicer if we normalize in the central limit theorem, which is a fairly standard way of writing it. So let me write the central limit theorem this way.

So now my function is that. And let's say that theta is 1 over square root of N 1, 1, which is a unit vector. So this is just the projection. I'll call it theta 1 in the theta 1 direction of f. So I take the characteristic function of the cube. If I project it in that direction, I get something that is almost a Gaussian. And that's the central limit theorem. And notice in particular that that projection is quite smooth.

So if I were to take the cube and I were just going to project it onto the x1 axis, well, the other variables would be irrelevant. And I would just get the characteristic function of an interval. It would not be any smoother than it was before. Still be discontinuous. And that's the same. I could project it onto any of the axes. They're all the same. That would be a characteristic function of an interval.

But if I project it onto the main diagonal, then I get the thing in the central limit theorem, which is extremely smooth. And the more I increase the dimension, the more it looks like a Gaussian. And in particular, the smoother it gets. OK. Now in the central limit theorem, it actually is not terribly important that all of these are exactly 1. It's just important that they're roughly equal to each other.

So as long as you take a direction, which doesn't emphasize any of the coordinate axes too much, you'll see something like the central limit theorem. So in all of those directions, the projection will look quite smooth. And in fact, it will look a lot like a Gaussian. But if you were to project it in the bad directions, in a coordinate direction or a combination of a couple coordinate directions, then you'll get something that's not so smooth.

OK. So that's an example of this theorem. And in this special example, we can say much more refined things about which directions are good and bad and exactly what happens. But this theorem is a lot more general, because you don't have to put in the characteristic function of a cube. You can put in quite a broad range of functions, and qualitatively similar things happen. Yeah?

AUDIENCE: [INAUDIBLE] but this theorem and the theorem you did earlier, it's really the same. Like when you integrate, you make it smoother because the production is an integration, kind of.

LAWRENCE GUTH: Yes. Yes. The comment is that when we project, we project by integrating some of the variables. And so the comment is when we integrate something we make it smoother. That's a really nice comment.

And it reminds us of something more familiar and fundamental, which is like if you had a function of one variable that was just an L2 site or L1, if you integrated it, it would get smoother. If you integrated it a whole bunch of times, then you would get something smooth. Yeah. So this is a little bit like-- this is a little bit like that. And I guess this smoothness does come from doing the integration. In this situation-- yeah? What's that?

AUDIENCE: The bounds are interesting still. Like how smooth it is is interesting.

LAWRENCE GUTH: Yes. The bounds are interesting. Yes. Yeah. And I guess one of the differences here is this function depends on many variables. We're going to integrate some of them but not others. And so-- yeah. Anyway. And then it depends which ones we integrate whether or not it gets smoother. But it turns out that usually it does. Yeah. But yeah, I think that is a good way of looking at it. Yeah. Cool. Cool.

OK. So let me end by mentioning. So as part of the class, I will try to mention different ways that projection theory is related to various parts of math. And so this reminds me of something I heard about, which is about trying to generalize the central limit theorem. So what if the input was not independent random variables, but something somewhat more general? Would it still be true that when we take a projection like this, you would see a Gaussian?

And so there's a big body of work on that. But one interesting direction I heard about is initiated by Keith Ball and its connection between probability theory and convex geometry. And the point of view is that maybe to some extent, what's important about the cube is just that it's a convex set. So instead of the cube, take any convex set in R to the N and it's large.

OK. Now there's no distinguished-- there's nothing special about the coordinate axes anymore. The coordinate axis projections aren't necessarily the ones that are bad, but this one isn't necessarily good. But the question he raised is maybe you take any convex set and look at its projections onto one dimensional spaces. Maybe most of them are almost Gaussian.

OK. And that's a recent topic of research, and a significant amount like that is true. Keith Ball gave an ICM talk about it and I put some references in the homework if anybody is interested in reading about it. Cool. OK. Well, let's stop there for today. The plan for Thursday is to come back to the large sieve, which is done, which is from number theory. And I'll try to tell you some things, what kind of applications in number theory does the large sieve have.