[SQUEAKING]

[RUSTLING]

[CLICKING]

LAWRENCE GUTH: OK. So the topic of today is sieve theory and the large sieve, which was the first place that I know of where our proofs like the proofs that we've been talking about appeared, and they appeared in the context of analytic number theory. So we'll talk about that, and then if there's time, I'll try to show you a really very closely analogous thing in the context of real analysis. And we'll compare them. OK.

All right. So the setting of the large sieve is that we have a function on the numbers from 1 up to n. So remember that that notation is the integers from 1 up to n And I have a function. OK. And then I'm going to think about what happens when I reduce these integers mod p. So I'll have pi p of f. That's a function from Z mod p to C.

And the definition is like this. Pi p f of a is defined to be the sum over all the n equal to a mod p of f of a. OK. And the setting of the large sieve is I have one function f, but then I consider-- this is a projection of f, and I consider pi p of f for many different p's and see what the typical behavior and how the different p's are related to each other. OK.

OK. Now, as before, we've seen it sometimes helpful to separate a function out into its constant part and its mean zero part. So f0 would be-- so take the average value. So 1 over n, sum n equals 1 to n, f of n. So that's the average value. And then maybe still I still have a function on numbers from 1 to n to do. OK. And then the high part of f is f minus the constant part. So the sum on n f high of n is 0. So this is a dangerous notation with the-- I'll make it a capital H.

OK. And you can do the same thing with the projections. So if I take the projection of f0, it's 1 over p sum a and Z by p pi pf of a. It's a constant function. And pi p of f pi is pi p of f minus pi p of epsilon. OK. A little tiny remark, because I was careful about where to put the parentheses, but it doesn't actually matter. So pi p of f pi is the same thing as pi p of f high. And pi p of f0 it's the same thing as pi p of epsilon.

AUDIENCE: Sorry. Can you just explain what the bracket means on the definition f0 again?

LAWRENCE GUTH: This one here?

AUDIENCE: Yeah, yeah.

LAWRENCE GUTH: So f0 is a function of, let's say m. And inside the large bracket is a number, which is like the average value of f.

AUDIENCE: So those are just parentheses.

LAWRENCE GUTH: Those are just parentheses.

AUDIENCE: OK.

LAWRENCE GUTH: OK. So the theme of the large sieve is that if you take almost arbitrary function here and you look at many different projections, then the oscillating part of-- the high frequency part of the projections will on average be pretty small, whereas the constant part-- well, it only depends on the constant part of f, but it could be typically bigger.

OK. So let me write a theorem which makes that useless. I might need just a little more notation to state our theorem. OK. So let's say capital PM is the set of prime numbers p prime in the range M over 2 less than p less than M. So here is the main theorem, which was proven by Linnik.

It says that if you have a function N goes to C and we have an M which goes up to square root of N, then if I take the sum p and PM of pi p of f high L2 squared, then that is bounded by N over M times the sum on N of f high of n squared. OK.

So it takes a little bit of work to digest the statement of this theorem and what kind of a bound this is giving us. So let's digest it together. And by doing-- we'll digest it by doing some applications. And then after that, we'll come back and we'll prove this theorem. OK. The character in our story is the primes of size about M, and it's useful to know about how many there are. So a background fact from analytic number theory is that the number of primes about M is around M over log M, which very roughly I'm just going to say is about M.

OK. So there are about M terms in this sum. It sometimes is helpful to replace the sum by an average. So a corollary is that the average p and PM of pi p of f high L2 squared is less than N over M squared times the sum on M of f high of N squared.

OK. So as a first application of this, let's prove an estimate about-- so last time, I gave the example of the square numbers. The square numbers have the interesting property that if you reduce the mod p for any p, you get the quadratic residues and there are only p plus 1 over 2 of them. So let's think about a situation like that. Suppose you have a set, and when you reduce it modulo p, you get significantly less than all p of the residue classes. What does that tell us about the set?

So there is a corollary if N is a subset of 1 to N, and pi p of A is significantly smaller than p. So I'm going to say it's less than 0.99 times p. Just some constant that's less than 1. And for all p would be interesting, but I'm going to say for all the p and PM to the 1/2. So we only actually are going to use the primes of size about N to the 1/2.

Then the conclusion is that the size of A is bounded by around N to the 1/2. This would be the case if was the set of square numbers. If A was the set of square numbers. This would be basically half of p. This would be all the primes, so in particular all of those primes. And this set of square numbers would have cardinality N to the 1/2. OK. So let's make our function the characteristic function of A.

All right. OK. So if I look at the sum A in Z mod p of pi p f of a squared. So let's say that p is in this, PM to the 1/2. OK, so how big would this be? Well, if each of these entries, if each of these guys were the same as each other, then the size of this thing would be A divided by P.

And it would be squared. And I'd be summing over p terms. So if all these values pi pf of a, they were all equal to each other, they would each have size size of A over p. And then the sum would be like this. They could be unequal to each other, and then by a Cauchy-Schwarz argument, this would be even bigger. OK. And then simplifying this a little bit, it's around A squared N to the minus 1/2. Because p is around N to the 1/2.

OK. Now, what happens if we take the high part? Well, because pi p-- so the support of pi p of f that's contained in pi p of A. So the size of the support pi p of f is less than 0.99 times p. That's enough to say that this function is not super close to a constant function in L2. So that implies that the sum a in Zp pi p of f pi squared is around the sum without the high.

So all of these L2 norms, they're all big. Does that feel intuitive to people? Should we talk about that more? Yeah. OK. So let's call it a lemma. So lemma is that if I have a function G on Zp to C-- it's not really important that this is Zp. It's just a finite set. And the support of g has size significantly less than p.

Then the conclusion is that the high part of g L2 squared is around the same as g L2 squared. Right. OK. So g is g0 plus g high. And g0 and g high are always perpendicular to each other. So the sum of-- so g L2 squared is around g0 L2 squared is equal to g0 L2 squared plus g high L2 squared. OK.

So OK. If g0 L2 squared was much smaller than g L2 squared, then clearly g high L2 squared would be bigger. So if g0 L2 squared is less than 1/2 of g L2 squared, then we're done. Otherwise, we go the other way. So let's say S is the support of g complement. And notice that S is at least some definite fraction of p.

And on S, g high is equal to, I guess negative g0, because they have to add up to 0 because g vanishes on S. So we get g high L2 squared is at least the sum in S of g0 squared. But since g0 is constant, this is at least S over p times g0 L2 squared, which is then at least-- so S over p is on the order of 1 and g0 L2 squared is on the order of g L2 squared. OK.

OK. So the high part of this projection has an L2 norm that's comparable to the L2 norm of the whole projection, which is at least this big in terms of the size of A. But the L2 norm of the high part of the projection shouldn't be too big by our theorem. So the theorem implies that the average over p PN to the 1/2 of this same thing is bounded above by N over M squared.

My M is the square root of N, so this just cancels. And then I get the sum of f of N squared. So this would be bounded above by the size of A. OK. So to summarize, this is bounded by this is bounded by this. A squared N to the minus 1/2 bounded by A. A is bounded by N to the 1/2. Any questions or comments about it?

OK. OK. So I feel like it's interesting that it matches the examples of the square numbers. It's sharp in that context. But also it would be nice to discuss more and look at more examples to try to get a feel for the numerology of the equations. And a helpful reference point for me is to look at a random set. So reference point. A random set.

OK. So here we're going to take a subset of the numbers from 1 to N and we include each number in A with probability 1/2. You could also, as an exercise, play with this probability. But let's just talk about 1/2 independently. OK. So what would we expect? Pi p characteristic function of A. How would we expect this to behave?

So this is the number of N from 1 up to N so that N equals A mod p and N is in capital A. And each one of those candidates was chosen with probability 1/2. So the expected value over our choice of A. So over the random choice of this set A, of pi p1A little a would be half of the number of guys here. So that's essentially 1/2 of N over p.

OK. Well, you wouldn't necessarily expect this number to be exactly N over 2p every time. We choose randomly, you'll see some ups and downs. So the standard deviation of the number that gets picked here would be the square root of the size of this. So it also would see that with high probability, the actual size of the projection minus the approximate size would be bounded by the square root of this.

And it typically would have size around this square root. So for example, if p is in capital PN to the 1/2, then the actual size of the projection minus its expected size would be around N to the 1/4.

OK. So let's compare this with what our theorem says about an arbitrary set. So what we concluded is for every P in here and for every A in Z mod p, we would have this. So if we took a random set and then for every P and every A, the difference between the actual size of the projection and its expected size would be square root of the number of choices, which would work out to N to the 1/4.

We're going to compare this to what our large sieve tells us about an arbitrary set. Any questions or comments about the reference setup? OK. So corollary three. If A is any subset of numbers up to N, then if I take the average over primes of size around N to the 1/2, and then I take the average over A and Z mod p, I take the actual size pi p1A of A, and I subtract off the expected size A over p.

So this is what-- so we have this many elements and we reduce them all mod p. I would expect, on average, that this number of them would be congruent to little a mod p. So then this is bounded by N to the 1/4 and this matches this. OK.

So the proof is copy down what corollary one tells us. It basically tells us this. So we're going to plug in corollary one where p is PN, where M is N to the 1/2. That will make this disappear. And of course, corollary one tells us that the average p and PN to the 1/2 sum A in Z mod p of pi p1 A of A minus A over p squared is bounded by the L2 norm of f, which would be the size of A.

OK. And that's smaller than N. Then if I want to replace this sum by an average, then I'm going to divide by p, which has size around N to the 1/2. So we replace the sum by the average, then this is bounded by N to the 1/2. And there's a square here. So I want to get rid of the square. I would use Cauchy-Schwarz. Get the average on p, average on A. Get rid of the square. So that's the proof of the corollary.

OK. Right. So the large sieve is telling us that if you look at a residue class with a random p and a random a, then it behaves-- so you take an arbitrary set, a worst case, complicated, any set at all. But then you look at a random residue class with a random p and a random a, then it behaves a lot like you had taken a random set.

So a cute application of this is if A is the set of primes. And there's a fun question, an interesting question in number theory. How many primes are there in some arithmetic progression? So that's exactly this thing. A is the set of primes. And this is telling us how many primes there are that are congruent to a modulo little p.

So the question is, how evenly distributed are the primes among these different arithmetic progressions? And the conjecture would be that for every little p and every little a that's not zero, this is always true. And this corollary actually makes some progress towards that. It says we don't know that it's true for all the little p's and all the little a's, but we know it's true for most of them. Now, it's a little bit silly to call that a corollary about the primes, because it uses nothing about the primes. It just uses the fact that the primes are a set of numbers.

OK. But that line of reasoning is actually important in some of the things we know about the questions. How well are primes distributed modulo p? So next class we'll come back to the question, how much do we know about this function when A is the set of primes. And we'll do the Bombieri-Vinogradov theorem, which is not quite the state of the art, but for many, many years was the state of the art about this question. Yeah. It uses these ideas in a crucial way. Cool.

OK. So that was my survey of some applications of this large sieve inequality. And next we'll prove the large sieve inequality itself. Oh. Actually, there was one other thing I wanted to say. OK. So imagine that A was a set of size N over 2. Then this thing here would have size about N over p, which works out to about N to the 1/2, and this thing has size about N to the 1/2.

And the error is about N to the 1/4. So this thing here is the pi p of f sub 0. And so what we're seeing here is that the pi p of f0 is much larger than pi p of f high at most of the points. So say in L2. So when we take a set of size N over 2 and we look at all these different projections. A typical projection looks almost constant. It's a constant function plus something that's much smaller.

So this process of taking a random function takes a set with no structure. And it produces something that's almost constant. And people sometimes use the word that the projections are getting smoother. And when we do the real analysis version a little bit later in class, we'll see that smoother is exactly the right word. Cool. OK.

Yeah. So I think that's a good way to summarize what the large sieve is telling us. It says if we take a set A, which is a pretty decent fraction of the numbers from 1 to n, and then when we look at all the projections of it, most of those projections are almost constant.

And that's relevant for this question up here, because if we hypothesize that pi sub p of A has size smaller than 0.99 of p, that's telling us the projection is really not very constant. That's only possible for small sets. So that's where we get that upper bound on the size of A. If A was bigger than that, then it would start to be the case that the projections of a would look pretty constant, and it would then it would be impossible that the support would be that small. Yeah?

AUDIENCE: What does p tilde capital P N to the 1/2 mean?

LAWRENCE GUTH: What does--

AUDIENCE: Under the average.

LAWRENCE GUTH: This one? Oh, sorry. That's P in. So capital P N to the 1/2 is the set of primes of size about N to the 1/2. This prime is one of them. Yeah. Thanks for-- thanks. OK. OK, so next, let's prove Linnik's theorem.

So the proof of Linnik's theorem is based on the Fourier transform. So we're going to take the Fourier transform of all of these characters, and then we're going to see how the different Fourier transforms are related to each other. So first let me remind you how the Fourier transform works for functions on the integers and functions on Z mod p.

So f is going to be-- think of it as a function on the integers, which is supported in this set 1 up to N. And so if I take its Fourier transform, what does that mean? The Fourier transform is going to be a function on R mod Z, and it's defined by f hat of Z is the sum on n f of n, e to the minus 2 pi i ksi dot ksi. That's the Fourier transform. And it has a Fourier inversion and it has a Plancherel. And they go like this.

Fourier inversion that tells us that f of n is the integral from 0 to 1, f hat of ksi, e to the 2 pi i ksi n d ksi. And there's a Plancherel, and it tells us that the sum on n of f of n squared is the integral from 0 to 1 norm of f hat ksi squared d ksi. OK. So this is probably the most familiar backwards from the way I've written it. If you started with a function on R mod Z, a periodic function, you would take its Fourier series.

And if you sum the Fourier series you get the function. This is the same thing. We've just chosen to label the Fourier series by f and the function on the circle by f hat. OK. And then we will also have a function on Z mod p. So suppose I have a function Z mod p goes to C. Then its Fourier transform will also be a function Z mod p goes to C.

And the definition is g hat of alpha. So I'll say that this variable is called a and this variable is called alpha. So g hat of alpha will be the sum a in Z mod p, g of a, e to the minus 2 pi i a alpha over p. And then we have Fourier inversion, which tells us that g of a can be recovered as 1 over p sum alpha in Z mod p, g hat of alpha, e to the pi i a alpha over p.

And Plancherel. That tells us that the sum on a, norm of g of a squared is 1 over p times the sum on alpha norm of g hat of alpha squared. OK. So that's how the Fourier transforms works for each of these worlds. And now we're going to put it together. So if I start with a function on the integers, I reduce it mod p to get a function on Z mod p. How does this Fourier transform relate to this Fourier transform? OK.

OK. So lemma, I'll call this lemma dictionary. This is a dictionary between the integer world and the mod p world. And it says that if I take pi p of f hat of alpha, it is f hat of alpha over p. OK. The proof is just unwinding the definitions of what all these Fourier transforms mean. So what does it mean, pi p f hat of alpha? Pi p f hat of alpha. That's the sum a in Z mod p, pi pf of a, e to the minus 2 pi i, a alpha over p.

OK. Now who is pi p f of a? Sum a and Z mod p, open parentheses. Sum n equals a mod p of f of n. e to the minus 2 pi i a alpha over p. All right. Now if you look at this double sum, we're just summing over all the n's. And now let's look at the thing that we're summing.

So n equals a mod p. That tells us that e to the minus 2 pi i n alpha over p is the same as e to the minus 2 pi i a alpha over p. Why? Different n's that are congruent to a mod p. So I think of adding a multiple of p to this n. That multiple of p-- the p cancels with that p, so we're changing this by an integer, and it doesn't change the value of it.

OK. So this is just the sum over n, f of n, e to the minus 2 pi i n alpha over p. OK. Another little remark. This is something we basically saw last time, that if you take the high part of pi p of f and its L2 norm squared, you take the high part of it, what it means is that we removed the zero frequency.

So I guess I'll also call this a lemma, but it's very similar to things we've already proved. So pi p f pi L2 squared is the sum on the non-zero alphas. OK. OK. So now we have a way of writing-- so let's look at the theorem we want to prove. We're going to take a sum over different p's of the L2 norms of the high parts of pi p of f. OK?

These lemmas give us a way of writing that and translating it into a sum that involves the Fourier transform of F. Right. OK. Yeah. I guess one last remark is. So pi p of f hat of alpha is f hat of alpha over p. That's for any function f. And so if I were to take pi p of f high hat of alpha, that would be f high hat alpha of p.

So now let's write out the left hand side of Linnik's theorem using the Fourier transform. OK. So the left-hand side of our theorem, I'll just copy it. It's the sum p in primes of size M. Pi p of f, pi hat L2 squared.

OK. So that's now some p in PM, some on the non-zero alpha in Zp. Pi p of f high hat of alpha squared. Which is equal to sum p in PM, sum alpha not equal to 0, and Zp of f high hat of alpha over p. Norm squared. All right.

Cool. OK. So we're taking the Fourier transform of the high part of f and evaluating it at a set of points norm squared and adding them all up. Let's visualize that set of points. So let's say QM is the set of alpha over p, where p is a prime of size around M, and alpha is strictly positive and less than or equal to p minus 1.

OK. So the size of QM is roughly M squared. So we mentioned before how many choices are there for this prime p. It's about M over log M, which is roughly M. And then for each prime, we're basically p choices of alpha, which is also around M, roughly M squared of these quotients. And an important fact about these quotients is that they're distributed quite evenly on the real line. So there are about M squared of them, and the distance between any two of them is at least M squared.

So here's a lemma. If alpha 1 over P1 and alpha 2 over P2 are in QM and they're not equal to each other, then the distance between them is at least 1 over M squared. OK. Proof. We're just going to put it over a common denominator. Alpha 1 over P1 minus alpha 2 over P2 is alpha 1, P2 minus alpha 2, P2 over P1P2.

By assumption, the difference is not zero, and so this numerator is not zero. It's an integer, so it has norm at least one. So it's at least that big. OK. I guess to be careful I should also say that if P1 is different from P2, then these two are never equal to each other. So we never-- so the importance of making these primes is that we never have the same number written as a fraction in two different ways here.

So that's also make the remark that if alpha 1 over P1 equals alpha 2 over P2 in QM, then it must be that alpha 1 equals alpha 2 and P1 equals P2. OK. Cool. All right. So I'm going to erase the dictionary lemma and I'll make a picture of this set and of what we're doing.

OK. Here's a picture. So here's a line from 0 to 1. And on this line we have this set, Q sub M, which is maybe not quite that evenly spaced, but it's quite evenly spaced. So that's Q sub M. OK. And then we're interested in-- our formula involves the norm of squared of the Fourier transform of f. Let me put that on this axis.

OK. So this will look however it looks. OK. We'll talk about this a little bit more. All right. So what we're taking in that sum is we're finding the actual values at these points and we're adding them up. And this sum might remind you a little bit of a Riemann sum that would help to give an approximation of this integral. And indeed we're going to compare this sum to this integral.

OK. Now there is a way that the sum could be a lot bigger than the integral, which is a narrow peak like this. If the function is very big at this one point, but the peak is extremely narrow, then it wouldn't contribute much to the integral, but it's still there as part of our sum. So it's important to know how narrow a peak like this can be.

And the heuristic, which is similar to heuristic we've talked about before, is that the function f high hat of ksi is roughly constant on intervals of length 1 over n. Because remember that f is supported on the numbers from 1 to n. So we'll make this precise in a moment. So I'll draw in what it means, that this peak here should have width at least 1 over n. And I'll also make a remark that M is less than or equal to N to the 1/2.

That's one of the hypotheses in Linnik's theorem. And so that guarantees that the spacing between consecutive points of QM is smaller than this 1 over N. So the picture is right that this 1 over N is smaller than the distance between the different points of Q. So this is just a heuristic, but let's follow it for now. And we'll see that it basically gives the inequality that we want. And then we can come back and prove things rigorously.

OK. So the heuristic tells us that the sum ksi in QM f high hat ksi squared is bounded by N times the integral from 0 to 1 f hat high ksi squared. Why? So take one term of this sum, say from this point. That one term is bounded by N times the integral on this interval around here of width 1 over N, which is a piece of this integral. OK.

OK. Good. So the rest is just algebra. It's a little bit annoying to get all of the constants right in the algebra, all the exponents. And in fact, I have already mastered up a little bit. So if we go back here, the left-hand side of the theorem is indeed this sum. I would just copy that correctly. And then for the sum, I did Plancherel. But when you do Plancherel mod p, there's this factor of 1 over p that I forget frequently.

So there's a 1 over p there. Now the 1 over p, the p has size about M. This is around 1 over M times that. OK. All right. So the left-hand side of the theorem is actually around 1 over M times this sum over here. So that would be bounded by N over M times this integral. And then we use Plancherel again, and this time there's no factor that I have trouble remembering. So that's N over M sum on N f high of N squared. OK. So this is the Linnik theorem.

OK. Cool. OK. Yeah. Let me mention in this picture. So we had this theme that if you take one function and you reduce it mod p for many different p's, then most of them are almost constant. So why the constant frequency is special compared to all the other frequencies in this story. OK.

So let me shade in for one particular prime, the frequency is alpha over p. So that would be frequencies alpha over P1 alpha and z P1. Let me do it for another prime. In the color blue, I'll draw alpha over P2, alpha in Z mod P2. What would that look like? Well, crucially, it would still have zero but all the other ones would be different.

OK. So now imagine doing this with all the different primes in our set of primes P sub M. I would get a whole rainbow of different stuff. Zero would be in every color, but other than that, each point is only one color. There are no other overlaps. OK. Now, f hat can be distributed anyway. We've assumed nothing about the function f, except that it's supported from 1 to n. So we know nothing about its Fourier transform except for this heuristic.

So f hat could look like anything, but-- well, most of the places that you put f hat-- OK, so one thing we do know is let's say we know the sum of f-- the L2 norm of f. So we know the L2 norm of f hat. If you stick some mass of L2 norm of f hat over this point, it will contribute to this one prime, but none of the others. But if you stick some mass over zero, it will contribute to everybody.

So zero frequency is being counted very differently from any other frequency. Every other frequency only contributes to one prime, but the zero frequency contributes to all primes. So therefore, if I start with my function f, if it has a decent amount in the zero frequency, then all the different p's also will see a decent amount in the zero frequency.

And then my function may have a bunch of other frequencies, but each one of them is only contributing to one prime. If I average over all the primes, those get damped down a lot compared to the zero frequency. So that's what's going-- that's a possible intuition about what's going on in the large sieve. OK. So the last thing to round out our proof of the large sieve is we'll actually rigorously prove this heuristic. Any questions or comments before we do that?

AUDIENCE: Well, it seems like we have more freedom to choose f here than in the real sieve. We always just choose [INAUDIBLE].

LAWRENCE GUTH: Yeah. So the question is, do we choose f? So here, f doesn't have to be the characteristic function a set. We've often talked about f being the characteristic function of a set. What's up with that? We'll also do something at the end of class, we'll mention something in the real setting where f doesn't have to be a characteristic function.

Yeah. So there are some theorems that become interesting where f is a characteristic function of a sparse set, and that sparsity comes into play. And there are others where f could just be anything. They're all part of the theory. OK.

This heuristic, this locally constant heuristic is similar to the one that's on the homework that you're thinking through that's due tonight. And by the way, there are office hours after the class for anybody who would like to talk about that or anything else. Yeah. So I'll also prove this in class, and yeah. All right. So how does it work? I'm going to choose a function psi N, a function Z to C so that psi N of little n is 1 if N is an integer from 1 to N.

And psi N is also smooth and rapidly decaying. OK. So then when I take the Fourier transform of psi N, it behaves like this. So it's around N if the frequency ksi has size less than 1 over N, and then it's rapidly decaying. So in pictures-- so ksi is only well-defined modulo 1.

And in this context, it's nice to draw the modulo 1 is going from negative 1/2 to positive 1/2, 0. So this axis has the norm of psi N hat of ksi. This is N, and this distance here is 1 over N. That's what it looks like. OK. And rapidly decaying could mean rigorously N times ksi times N. So once ksi is bigger than 1 over N, this thing starts to be bigger than 1 and kick in. You could throw in that. Yeah?

AUDIENCE: So what does smoothness mean for a map from the integers?

LAWRENCE GUTH: Yeah. What does smoothness mean? Right. You could think of psi as a map that's defined on the whole real line, and it's smooth on the whole real line, and then we happen to be evaluating it at the integers. Yeah. OK. Cool.

OK. Now this is helpful because f is equal to f times psi N. So if the support of f is contained in integers f to N. And then when we take the Fourier transform, we get that f hat is f hat convolved with psi N hat. And so in particular, so f hat at a particular frequency of ksi is f hat convolved with psi N hat. Psi N hat looks like that. So the norm of f hat is bounded by the norm of f hat convolved with psi N hat, and also the norm of f hat squared is bounded by the norm of f hat squared convolved with psi N hat by Cauchy-Schwarz. So actually one feature of this formula is that the integral of psi N hat ksi is bounded by 1. It has height 1 and it has width-- height n and width 1 over n.

AUDIENCE: Sorry. So psi n hat is the Fourier transform of cosine in the sense of real function? Or it is a reference point in the sense of--

LAWRENCE GUTH: OK. Yeah. That's a good question. That's a good question. OK. All right. So phi, I had originally said, is a function on the integers. But in answer to the question, what does it mean to be smooth? I said, we should also think of it as a function on the real line. So I'll call them psi n integers. That's our function on the integers. Psi N real line. That's a function on the real line.

And they each have a Fourier transform. Psi n on the integers hat of ksi is the sum over integers psi n of n, e to the minus 2 pi i n ksi. And psi n, if you think of it as a real valued function, it has a Fourier transform. And ksi, that's defined to be the integral over the real line. Psi n of x, e to the minus 2 pi i x ksi dx.

OK. So the question was, when I write psi n hat, which one of these two things do I mean? OK. So I mean this one, but they're equal to each other. So the Poisson summation theorem says that psi n z hat is equal to-- oops. Is equal to. Is that right? Wait, wait, wait. So these are closely related. Let me think for a second. Yeah. So this guy is the sum of z in z of psi n R hat of ksi plus z.

OK. So we should make two pictures. In yellow, I will draw psi n R hat of z. And let's extend this. It looks basically the same. And then it gets really small and it stays really small. And what does-- so in blue, I'll do psi n z hat ksi.

It looks really almost the same except this one is periodic. This Fourier transform is defined on R modulo z, so it's going to be periodic. Z periodic. So then here at minus 1, it's going to go up again. There's going to be an identical copy. The blue one is z periodic.

OK. So if you compare. So then if you look at this formula, also, it's taking this function and turning it into a z periodic function by periodizing it. And since this function originally was extremely localized to between minus 1/2 and 1/2, it almost looks like I took the yellow function and I just took copies of it and put them here. So the copy over here can-- yeah. Sorry. OK.

Anyway, so that's what this one-- that's what this one looks like and what this one looks like. OK. So to prove this estimate, first you prove it for psi n R. And you can do that using Fourier analysis over R. And you use the smoothness and you integrate by parts many times. And then to study the 1 over z, you use this formula.

And because this guy is decaying so rapidly, the contribution from the-- so if this ksi is between minus one half and one half, then the contribution coming from integers z other than 0 is incredibly small. Anyway. So using the bounds for this one you use to give bounds for this one. OK. Thanks for those questions, everybody. That was helpful.

So now we're going to do a slightly more rigorous version of this argument. So rigorous proof. So the left-hand side, we decided, was around 1 over M times the sum over the frequencies in QM of f high hat ksi norm squared. That part was rigorous before. We didn't need to make it rigorous. But now we need to relate this sum to an integral.

And we use this fact here. This fact encodes the idea that this quantity is locally constant at the scale 1 over N. So this is bounded by 1 over M sum ksi in QM of the integral f hat squared of omega times psi n hat of ksi minus omega d omega. So this is just writing out the convolution as an integral.

Now we can bring this sum inside. So this is 1 over M integral f high hat of omega squared times the sum over ksi in QM of psi n hat of ksi minus omega d omega. OK. And this sum we can bound by N. Why? Because each term in the sum has size at least around N. This term is a bump, essentially a bump of height N and width 1 over N around ksi. And the distance between any two different ksi's is at least 1 over N, so those bumps don't overlap with each other.

Now, the bumps aren't quite compactly supported. They have the tails that you can see in the picture, but the tails are incredibly tiny. And so if you just plug in the rapidly decaying bound, the tails don't contribute anything and you get this bound. OK. Then if you take out this n, you have what we had before. N over M, integral from 0 to 1, f high hat squared ksi, ksi, N over M sum f high of N norm squared. And that's the proof of Linnik's large sieve.

Cool. So we only have a little bit of time left, about five minutes. We could either, if people have things you want to talk about we could just talk for five minutes. Or I could set up the next thing and try to give a feeling of how something in real variable setting is analogous to this thing in a number theoretic setting.

OK. So I could give a little-- try to give a little teaser, preview, foreshadowing of how this same idea occurs in real analysis. I want to keep this picture. OK. The thing that we proved is if you have a function on the integers from 1 to N, then when you project it mod p, most of the projections look almost constant. They look smoother than the original function.

And there's a totally analogous phenomenon in R to the D. If you have a function on R to the D and you consider projections onto lower dimensional subspaces, then almost all of them look smoother than the original function. And like we mentioned on the first day, if you're in high enough dimension, if you take a function and you project it onto a typical line, even if the initial function is only in L2 and is nowhere continuous, on a typical line, the projection will be continuous and even C1 and even C2. So the projections are much smoother than the original function.

OK. So here is a setup for this that's kind of analogous to the large sieve. So we have a function on R to the D, and then we'll have v contained in R to the D which is a subspace. And if you have a function-- so pi V of f is the projection of f onto this plane. So that's a function from V to C.

OK. So now I have a remark that if you have any function of this vector space V, then its Fourier transform is also defined on the vector space V. OK. And there's a dictionary lemma. How is the Fourier transform of the projection related to the Fourier transform of the original function? And it says that pi V of f hat is just f hat restricted to V.

So this is a function on the vector space V. Its Fourier transform is supposed to be a function on the vector space V. f hat is a function on R to the D. But this Fourier transform is just f hat restricted to that, to V. OK. So now make a picture. So here is R to the D, and inside of there, there's the space V1 and there's another space V2. Whoops. OK. And I let you imagine other ones. They're all subspaces, so they all go through zero, although my picture was not perfect.

And I want you to imagine this as Fourier space. So we have a function f. f is just in L2. Say it's an L2 function on the unit ball. So its Fourier transform could be anywhere here. And all we know about it is say we know the L2 norm of f, so we know the L2 norm of f hat. Could be anywhere. If you happen to put some L2 norm of f hat near 0, then that will be part of the Fourier transform of pi V for all the different V's.

But if you put some of the L2 mass of f hat over here, that will contribute to pi V1, but it won't contribute to most of the other pi V's. So therefore, when you take a random projection, the part of the Fourier transform at low frequencies is more highly represented and the part at high frequencies is less highly represented. So a typical projection. So initially the Fourier transform could have been anywhere, but then a typical projection will now have its Fourier transform concentrated at low frequencies, which makes it look smooth.

OK. So here's a statement like that that we will prove next time by the analogous reasoning theorem. If f is an L2 function on R to the D supported in the unit ball, and if D is big enough, then for almost every-- so then if I take the average over theta in the D minus 1 sphere of pi theta of f. So this means I take f and I project it onto the line in the theta direction. CK norm of that is bounded by the L2 norm of f.

OK. So we'll prove that next time. But to end, I just wanted to have everyone look at the analogy between these two pictures. So in Fourier space when we project our function onto different subspaces, then we have these different pieces of the Fourier transform that we're seeing. And in Fourier space in the large sieve, when we take our function and we project it mod p for different p, we have these different color pieces of the Fourier space that we're seeing.

And the different color pieces intersect at 0, and they don't intersect anywhere else, or they don't intersect very much anywhere else. And that has the effect that when you take the projections, it emphasizes the zero frequency or the low frequency. And that plays out slightly in slightly different ways in the different settings but morally very similar. Cool. OK. Cool. So let's stop there for today. I will hang out afterwards, make office hours talk, and I will see you next week.