[SQUEAKING]

[RUSTLING]

[CLICKING]

LAWRENCE GUTH: OK. So today we're going to talk about the large sieve. So last week, we introduced the large sieve. And we explained how it's analogous to the projections theorems that we proved in finite fields, and the projection theorems that we proved in Euclidean space. And today, I wanted to explain to you some of the applications of the large sieve in number theory.

So the original application of the large sieve was by Linnik. And he was interested in how quadratic residues are distributed. And I wrote that one up on the homework. So the homework problem is to retrace Linnik's steps proving about how quadratic residues are distributed. And then a second important influential application has to do with how prime numbers are distributed modulo q. So that's what I'm going to tell you all about today.

So first, we'll do some background. First, we'll talk about what is known about how prime numbers are distributed modulo q. All right. So let's say pi of N is the number of primes up to N. And then, we can look at the modulo q for some choice q and see how many of those primes are in each moduli class.

So pi N, q, a is the number of p prime, p less than or equal to N. And p equals a mod q. Now some of these moduli classes will be quite-- some of these will be quite small. If a and q have a common factor, then there will be only finitely many primes in this moduli class. So the main focus is on the ones where the GCD of a and q is 1. So focus on the case.

So I'm going to write it that a is in Zq star. Zq star is the set of a and Z mod q, where a is relatively prime to q. So numerically and theoretically, it looks like the primes are pretty evenly distributed among all of these residue classes. And it is relevant to say how many there are, so that's called phi of q. It's the cardinality of Zq star, the number of different a that are relatively prime to q.

So if the primes were really well evenly distributed among the moduli classes, then you would expect-- so I'll write this way. So we see even distribution, and that means that the actual number of primes in the moduli class is close to the total number of primes divided by the number of moduli classes. So how would we measure this precisely?

We define delta q of N, the defect of equidistribution, to be the maximum over all the residue classes of the difference between the actual number of primes and this naive estimate. So when we ask, how evenly are primes distributed modulo q, quantitatively, that means, how big is this thing? So this question was first addressed by Dirichlet, who showed that, in a certain sense, that this is close to this. Here's what he proved.

He proved that for any q, if you take the limit as N goes to infinity of delta q of N, it's asymptotically smaller than N/q. So this limit is 0. So actually, let me say, when you're talking about these being equidistributed, we're going to mean that this is small. Small relative to what? Well, this is a good approximation of this if this error term is small compared to the main term. So equidistribution would mean that delta q of N is much smaller than the mair term, which is on the order of N/q, or it could be N over log N q.

So Dirichlet's arguments, you can make them quantitative. Instead of having a limit, you could say, if N is sufficiently big. And they were made quantitative by Siegel and Walfisz, which involves some additional ideas. And so this line of thinking led to the following theorem. So for any q, delta q of N is bounded by-- OK, so let me say it this way. So for any A, there's a constant C sub A. And delta q of N is bounded by C sub A, N, log of N to the minus A.

Now when is this useful? When does this tell us that delta q of N is much smaller than N/q? Well, q cannot be too big. So delta q N is much smaller than N/q if q is smaller than a power of log N. I'm not going to talk about the proofs of these theorems in the class. So Dirichlet invented L-functions, you may have heard of Dirichlet L-functions, in order to prove this theorem. And all of this work is based on L-functions.

So those are nice theorems, but they only work when q is pretty small compared to N. What happens if q is bigger? Well, experimentally what seems to happen is the following. So this is conjectured by Montgomery. So for every epsilon bigger than 0, delta q of N is bounded by C epsilon N to the epsilon. It's like a fudge factor. And then, there's N/q to the 1/2.

So this is a pseudorandom behavior. So a good reference point, sometimes people ask, are primes distributed? Does it look random the way primes are distributed? So a good reference point is there are about N over log N primes up to N. So instead, pick N over log N random numbers. Reduce the mod q and see what happens. So this is what you would see if you had random numbers. So this is a pseudorandom conjecture.

So this is much stronger than we could prove. And it implies, incidentally, that delta q of N is much smaller than N/q, so we have some pretty good equidistribution, as long as q is smaller than N to the 1 minus epsilon. So there is a huge gap between this and this.

So this result, I think, was proven in the 1930s. And it is still, today, the best result of its kind. So we cannot increase this to say N to the epsilon. If we are willing to assume the Riemann hypothesis, or this would be the generalized Riemann hypothesis, then we get much stronger information, although still not this conjecture.

So it says, for every epsilon bigger than 0, delta q of N is bounded by C epsilon, N to the epsilon, times N to the 1/2. Looks like this, but with no q in the denominator. And that's good enough to imply that we get some amount of equidistribution if q is smaller than N to the 1/2 minus epsilon. So it gives some idea of how difficult this conjecture is that, even with the Riemann hypothesis, we're only halfway there.

So this is everything that is known if you want to have a theorem that definitely works for a particular q. And having run into a barrier there that has been difficult to overcome, people started to think about, well, maybe we won't prove a theorem for every single q. But we could prove a theorem for most of them. So that's the main topic of the class today, so theorems for most q.

So this theorem, again, I want to credit Rényi and Bombieri-Vinogradov. I read about the history yesterday. So this is usually called the Bombieri-Vinogradov theorem. The first person to prove a theorem of this flavor seems to be Rényi. And then, various people improved the parameters in the theorem, building up to Bombieri and Vinogradov.

So it says the following thing for every epsilon bigger than 0. So if I take the sum, epsilon q, up to N to the 1/2 minus epsilon of delta q of N, check the exact phrasing. A, then, this is bounded by constant depending on epsilon and A, and log of N to the minus A. So let's digest this a little bit.

So the job of epsilon is that we can almost go up to N to the 1/2, but not quite. And then, the job of A is that-- so this tells us that for most, not all, but most of the q up to N to the 1/2 minus epsilon, delta q of N is bounded by, well, I divide both sides by the number of terms. And so I could put here, N/q log of N to the minus A, so which is indeed much smaller than N/q. So informally, for most q, we get good equidistribution of the primes modulo q.

And it so happens that the size of the q's we can get up to here matches the size of the q's you can get up to if you knew the Riemann hypothesis. And so this is sometimes summarized as being, it's definitely not as good as the Riemann hypothesis, but it's for most version of what you would learn from the Riemann hypothesis. So my goal today is to tell you the ideas of the proof of this theorem.

There are certain things that are a bit messy. And I won't do all of the messy things, so we won't do a complete proof. But I'd like to address a couple of questions, like, what does this have to do with sieves and projection theory? And another question, what does it have to do with primes, or how will we use it there, in primes?

Since we're only trying to prove things for most q, we're looking at all of the different q's. So we have one set of primes and we're projecting it modulo q for lots of different q's. So that's the setting of projection theory where we can bring in things like the large sieve. So I wanted to recall, so the large sieve allows us to prove the following kind of thing.

So we have a set x contained in the integers up to N. And I look at pi p of 1 x. And I'll take the high part of that. And then this thing is small for most p in Pm. So this Pm, maybe I'll put it-- so Pm is the set of p prime, p around m. So this looks, on the surface, like a similar thing.

We have a set x. We project it modulo p for many different primes p. And we throw out the constant part and just look at the variation from being constant. And this is saying that this projection is pretty evenly distributed for most of the primes. So it sounds similar to this. And indeed, this will be crucially relevant.

Now this is a statement about every set x. In particular, you could put the primes here so you get a theorem about primes, but it's true for every set. This theorem is definitely not true for every set. So I'll make a remark that the theorem is not true if you replace the primes by just some set x of numbers up to N, with maybe roughly the same size as the primes. So there's a simple example, but I think it's instructive to see.

So the examples like this, I'm going to try to build a set where I can reduce it modulo every prime. And whichever prime I reduce it by, I will see one moduli class which is extremely overrepresented. So here's how we build such a set. So for every p in primes of size around m, m is just a free parameter here, I choose some a and Z mod p. And then, I'll make my set x is going to be the union over all the primes.

And then, I take the set of N up to N so that N is equal to a mod p. So for each prime I pick a moduli class. I take all the integers in that moduli class, and I take a union over all the primes in my set of primes. So if I were to take pi p1x of a, well, x contains all the integers in this moduli class up to N. So this is basically N/p, which is around N/m. And that's much bigger than average.

The average would be the size of-- yes. So actually, let's figure out the size of x. So the size of x is, at most, there will be about N/m numbers in this moduli class. And then, there are about M over log m primes, so something like this. So that is a smallish fraction. It's much less than all of them.

So this is way smaller than the size of x over p. So this guy is seriously overrepresented, and it happens for every p. OK, great. So this theorem is, we will see, the large sieve is absolutely crucial thing in the proof. It's the main idea. But this is not a theorem about arbitrary sets. It's a theorem about prime numbers, and so the proof needs to use something about prime numbers.

So the main character is a concept called multiplicative convolution. And I think this is a nice concept to know about. It's a cousin of regular additive convolution. We'll see that it interacts with all the things in our story. So multiplicative convolution interacts really nicely with prime numbers, describing what our prime numbers, it comes up naturally. And it also interacts nicely with taking projections. And so it interacts nicely with the large sieve, and fitting all those things together is how we prove this.

So what is multiplicative convolution? So suppose f and g are functions on the natural numbers. If I take the multiplicative convolution of them, it means I add up over all the pairs n1 and n2 whose product is n. And then, I take f of n1, g of n2. So let's compare that with regular convolution.

So regular convolution, say, on the integers, would be the same thing. But I would have the sum over all the pairs n1 n2 who add up to n. And the multiplicative version is, we replace the sum here by a product. So that's multiplicative convolution. So we have to see how multiplicative convolution connects with primes. And we have to see how multiplicative convolution interacts with projections. We'll put that together and that's how we prove the theorem. So let's start with primes.

So the large sieve inequality, we use this word, sieve. And I will explain to you now, where does the word sieve come from? And it comes out naturally when you think about primes. So how do you find the prime numbers? Suppose I wanted to make a list of prime numbers up to n. What would I do?

I would start with all the natural numbers. Then, I cross off the multiples of two. Then, I cross off the multiples of 3, and so on. And then, what's left are the prime numbers. So that process is called sieving. I think sieving physically is where you take a thing full of sand that somewhere in there has a little gold, or diamond, or something. And you shake this thing and the sand falls out, until eventually you're just left with the diamonds.

So the image is you have a thing full of sand, which are all the natural numbers. And you shake it, and you take out the multiples of 2, and the multiples of 3, or whatever. And then, eventually you just have the prime numbers left. That's the image. To find the primes, we start with the natural numbers. Then, we cross out multiples of 2. Then, we cross out multiples of 3, and dot, dot, dot. So we can all imagine that.

Now, let's try to describe this process using formulas. And we'll see that the formulas that naturally appear are formulas involving multiplicative convolution. So I'm going to write 1 of n to be the characteristic function of all the natural numbers. And then, the next character I'm going to put is a function called difference sub p of n. So it's 1 if n equals 1, and it's negative 1 if n is p, and it's 0 otherwise.

Now let's think what would happen if I took all the natural numbers and I did a multiplicative convolution with D2. So this would give me, so D2 has a positive delta function at 1 and a negative delta function at 2. So when I do this I'm going to get characteristic function of all the natural numbers coming from this 1, minus the characteristic function of 2 times the natural numbers coming from that 1.

Let me pause there. Does that make sense? Does that fit with the plugging in the definition of multiplicative convolution? Let's do it slowly to be sure. So what's the definition? 1 convolved with D2 of n is the sum. n1, n2 equals n of 1 of n1 times D2 of n2. Now D2 is only non-vanishing in two cases. So we actually only have to consider that n2 is 1 or n2 is 2.

So the 1 is always there. So that is 1 of n times D2 of 1. And then, if D is even, so this is what happens if n is odd. And then, if n is even, then n2 could be 2. So we would get 1 of n, 1 of n over 2. Sorry, we get the-- so this is both cases. 1 of n, D2 of 1, minus 1 of plus 1 of n over 2, D2 of 2. So this is just 1. And this is 1 minus 1, which is 0. So we're left with the characteristic function of the odd numbers.

Let's go one further. If I take the characteristic function of the naturals, I do a multiplicative convolution with D2. Then, I do a multiplicative convolution with D3. What happens? Well, this 1 gives me the characteristic function of the odd numbers. And then I'm convolving with D3. Now what happens? I get the characteristic function of the odd numbers minus the characteristic function of 3 times the odd numbers. Then, I do this.

So I have the odd numbers, but now I'm going to cross out the ones that are multiples of 3. And so this is the characteristic function of, n is relatively prime to 2 and 3. So this shows by example that this process, starting with the natural numbers, crossing out multiples of 2, crossing out multiples of 3, it corresponds to convolving by these D's over and over.

So now let me make a little notation. So we're going to do sieve for primes, general version. So suppose that S is a set of primes. And then, RP sub s of n, that stands for relatively prime to s. This is 1 if p and n are relatively prime for all the primes in my set. So it's 1 if my number n is relatively prime to everybody in s, and 0 otherwise. So it's 0 if there exists a prime in my set that divides it.

So let's make our remark. So if S is the primes that go up to the square root of N, and little n is, let's say, bigger than the square root of N, but at most N, then either n has a prime factor from this list or it is prime. So in that case, our relatively prime s of n is just P of n whether n is prime.

So this thing, if I'm interested in the primes in a certain range, say, from n over 2 to n, I can detect them using this function. And this function we've just seen is a big convolution. So lemma, so if my set of primes is p1 up to pr, then relatively prime sub s of n is 1 convolved with Dp1, convolved with Dpr. These are all multiplicative convolutions.

AUDIENCE: Can I ask you, for the examples there, I wondered why you have p of 1x of a is over p, like why it really becomes contributions from other teams?

LAWRENCE GUTH: Yeah, OK. So the question is coming back to this example. We built a set whose projection onto every prime has one place that's really big. And x is made as a union of arithmetic progressions. Now this x is a set. It's not a function. So these arithmetic progressions might overlap.

But if they do, the point where they overlap in this formulation only counts once, although it's not super important. So if you look at pi p of 1x of a, that's how many elements of x are congruent to a mod p. And so all of them are already. And some of the other p might also contain some of these elements, but it doesn't count anymore.

AUDIENCE: Oh, OK.

LAWRENCE GUTH: Although if it did count anymore, it would just make this phenomenon even more extreme than it already is. OK, cool. Any other questions or comments? So the characteristic function of the primes is a special function, probably in lots of ways. And the way that's relevant for us today is that the characteristic function of the primes is basically this function, which is a big convolution, a big multiplicative convolution.

And so the Bombieri-Vinogradov theorem is mostly a theorem about functions that have the structure that they're convolutions. So that is how multiplicative convolution is related to primes. Next, let's think about how multiplicative convolution is related to projections.

So here there are a bunch of different fairly simple lemmas. And all these lemmas have the flavor that doing multiplicative convolution and doing projections fits together as well as you could hope. And I think, basically, the reason for that is that the projection operation is a ring homomorphism from the ring Z to the ring Z mod q. And we'll be talking about multiplication.

So being a ring homomorphism, it means that the projection respects the product operation. And so the multiplicative convolution is defined in terms of the product operation, so it all fits together. So lemma 1 says that, if f and g natural numbers go to C, then if I take the projection of f convolved with g, I get the projection of f convolved with the projection of g.

This one requires the definition. If I had been savvy in general, I might have defined multiplicative convolution in the first place for any multiplicative monoid or something. But anyway, let's just say what it is for z mod q. It is what we think. If I have maybe capital F and capital G functions on Z mod q, then multiplicative convolution of a class a is the sum over all pairs a1 a2 that multiply to a of capital F of a1, capital G of a2.

So this is the first lemma. And the proof idea is you just start on either side and write down the definition and unwind everything. And you use the fact that that reduction mod q is a ring homomorphism and it all fits together. So I think that my saying that is clearer than if I were to write it out. Yeah.

AUDIENCE: Do you not get the issues from double counting them at the left side because you have to make the kernel over them?

LAWRENCE GUTH: OK. So this is a homomorphism but with a kernel. Let's write it out and make sure that this actually is OK. I'm going to put the proofs on this board so that eventually we'll just have this list of lemmas that will be handy. So here's how I like to think about the proof. I'm going to think of f as a sum over n1 of a delta function at n1 weighted by f of n1. And I can do the same with g. Sum over n2 delta function of n2, g of n2.

What happens when I take their convolution? The multiplicative convolution is the sum over n1 and n2. I have a delta function at n1, n2 weighted by f of n1, g of n2. Let's check that this is the same definition as before. So if I want to evaluate this at a point n, well, when I evaluate this at n it only appears if n1 times n2 is n. So I'm summing over pairs n1 times n2 that multiply to n. And I'm summing up f of n1, g of n2.

Does this look OK to people? Maybe let me add, when I write delta n1, that's a function of n. And it means it's 1 if n is equal to n1, and it's 0 otherwise. So this delta is a delta function. So f is this, and g is this, and f star g is this. So now, what is pi qf of a? It's the sum over n1 of a delta function at n1 mod q of a, f of n1.

So I like the image of f and g as some kind of density. So we think of f as a description of some stuff. And there's f of n1 stuff located at n1, and same for g. So what does the projection do? It just takes the stuff at n1 and it moves it to n1 modulo q. And stuff is coming to n1 modulo q from several different places and it just adds up.

And what does convolution do? It takes every bit of stuff here and every bit of stuff here. It takes a bit that's at n1 here and a bit that's n2 there. Takes those two bits and it makes a new bit at n1 times n2. And the weight of the new bit is the product. And there may be several new bits arriving at the same place, then we add them up. So that's what the projection does.

So now, if I take the projection of f convolved with the projection of g at a, it's the sum over n1 and n2, the delta function n1 times n2 mod q of a, f of n, 1, g of n2. So then that's the same as pi q of f convolved with g.

So all it boils down to is, I have this bit of f and this bit of g. And I could first reduce them each mod q and then multiply them together, or I could first multiply them together and then reduce mod q. And those are the same as each other because reduction mod q is a homomorphism. So that's the first lemma.

Second lemma says, so eventually, we would like some L-infinity bounds. But so far in projection theory, we've mostly seen L2 bounds. And they're related to each other by basically Cauchy-Schwarz. So it says that if I have f and g on Z, q, then a multiplicative convolution of them, if I take the L-infinity norm on Zq star, that is bounded by f L2, g L2.

So in lemma 2, I think it's important to put this star here, multiplication. So we leave out 0 and we leave out things with a common factor with q. If you included 0, there are a lot of ways to have a product that makes 0, and I felt a little worried about how that would go. So that's why we have this star here.

Let's prove lemma 2. Proof of lemma 2. So if a is in Zq star, then f convolved g of a is going to be the sum b in Zq star, f of b, g of ab inverse. Why? Well, we're adding up over pairs that multiply to a. And since a is in Zq star, a is invertible. Each of these has to be invertible. So b had better also be in Zq star, and so we get this. And then, we use Cauchy-Schwarz, and we bound that by f L2, g L2.

There is a small technical issue about Zq or Zq star. So we're going to break something up into a constant part and the other part, constant part and the high frequency part. And it's slightly different whether you do this on Zq or Zq star. It's just a minor technical nuisance. q is even usually going to be prime, so I'm just going to be leaving out 0. But let me just write a lemma that says that this isn't such a big deal.

So a little notation. So suppose I have a function f on Zq. I might be interested in f just on Zq star. So I'm going to call that f star. It's just the restriction of f to Zq star. Now I could split either one of these up into a constant part plus a leftover, but the constants would be slightly different. So f is f0 plus f high, and f star is f0 star plus f high star. So f0 and f0 star are constant. And the other pieces have mean 0, but in slightly different sense.

So if I add up fh of a over Z mod q, I'd get 0. But if I add up f star h over Zq star, I get-- OK. So I felt like I had to say this to be honest at some point, but it's not really a big deal. And lemma 3 says that the L2 norm of the high star part is at most the L2 norm of the regular high part. So this one lives on Zq star, and this one lives on Zq.

And then, the last lemma says that the high part commutes well with multiplicative convolution. So if f star and g star are on Zq star, then-- there's going to be a lot of stars in this formula-- the high part of the convolution is the convolution of the high parts. We could prove all these things if we feel like it. They're not difficult. I think I want to leave it for now, because I've put an excessive number of details on the board, and try to step back, and have a bigger picture, and then pause and check in.

So all of these lemmas imply the following proposition. So maybe I have some convolution, some multiplicative convolution, f star g. And I would like to reduce it mod q for different q's. So I look at it projected mod q. And then, I'm only going to be interested in the star part. I'm going to restrict it to Zq star.

And I want to know whether that's almost constant, so I'll look at the high part of that. And I want to know that that's small everywhere. This is the kind of estimate that we are asking about in the Bombieri-Vinogradov theorem because our set of primes is morally a multiplicative convolution.

And I'm taking the set of primes and reducing it mod q, and looking at the different classes that I have in Zq star. And I think that that's going to be almost constant, so I'm going to subtract off the constant part. And now I want to know that what's left is very small everywhere, so that's an L infinity norm. So the Bombieri-Vinogradov theorem is about estimating this kind of thing. And the proposition says that that is smaller than pi q of f high L2, pi q of g high and L2.

And it's just assembling a bunch of those lemmas. So the projection of the convolution high part is the-- so proof of proposition. We have the projection of the convolution. Restrict it. Take the high part. This is the projection, star high convolved with the projection of g star high. So that's lemma 4, high parts.

No, sorry. This is pi q of f convolved with pi q of g, restricted to star high. So that's lemma 1. Convolution plays well with projections. And then, lemma 4 says that that's that. And so now the L-infinity norm is bounded by pi q of f star high L2, pi q of g star high L2. That's lemma 3. Sorry, that's lemma 2. And then, that's bounded by just pi q of f high L2, pi q of g L2, and that's lemma.

So let me summarize a little bit and then pause to check in. So an issue that we're dealing with is L2 norms versus L-infinity norms. Because we're trying to prove something that's true for every moduli class, so it's an L-infinity norm. But our techniques that come from Fourier analysis, and so on, it usually tells us about L2 norms.

So if you have a bound like this, it tells you that pi q of f is pretty evenly distributed, at most a. And we're trying to prove a bound like this that says that the projection of this thing is pretty evenly distributed at every a. And the point of this proposition, it says that if pi q of f and pi q of g are pretty evenly distributed, just at most a, then the projection of their convolution is pretty evenly distributed at every a. And so that's why being a convolution is helpful.

So a little pause for questions and comments. OK, cool. So now let's go on. And these two things are controlled by the large sieve. The large sieve tells us that if f is some function, then this L2 norm is small for most q. And so we're going to write down that bound and see how small it is. And then, we're going to plug it into this proposition. We'll get a bound for this.

So large sieve. So remember that Pm is the set of p prime, p around m. And so we proved in class, and so we'll always have a function which is defined on the integers up to N, and it goes to C. And we proved in class that if you take the sum p and Pm of-- so in class, we thought about what happens if M is smaller than the square root of N. Then, we got the sum pi q of f high L2 squared. And that is bounded by N/M times just f L2 squared.

M could be bigger than N to 1/2, and that will be relevant, and this story is often relevant. And the proof idea is the same, but then it works out a little bit differently. So if M is bigger than N to 1/2, then this same sum is bounded by M times f L2 squared. So this one was in class, and this one is on the homework.

So I posted the homework for next week, and the first problem is to do this mostly as a way of thinking through how the large sieve worked. And then the second bit is about Linnik's thing about quadratic residues. So now we are going to plug these things into here. So we get, all together, I'll call it a theorem. Let's say like this.

So suppose f is a function on N1 up to C, and g is a function on numbers up to N2. And so if I take their multiplicative convolution, that will be a function on the integers up to N, where N is N1 times N2. And then, I can take the sum p and Pm of f convolved with g. I project it mod q.

I maybe take the high, take the star part. That doesn't matter if you do it or not, but in our application we'll do it, and then subtract off the constant. So the idea is that this thing should be pretty close to constant much of the time. So we look at the non-constant part. And then, that is bounded by, there's an expression that's slightly messy. And then, we have f L2 and g L2. And the thing that goes in here when you work it out is-- I'll have to check. Yeah?

AUDIENCE: When you're defining the convolution of f and g is a function that can be integers up to N, how is that defined for the product pair 1 times N for N large? Because f and g are individually only defined up to N1 and N2, right?

LAWRENCE GUTH: Yes. That's right. So you should imagine f and g being extended as 0 beyond that. So the question was how to define in the definition of this. We'll have all the pairs, little n1 times little n2, that multiply to n. And it could happen that little n1 or little n2 are too big to lie in this region. And then we just set f to be 0.

Oh, and sorry. And this is L infinity. It's not L2. So one thing that I find a little difficult about the proof, or teaching the proof, is that the formulas are a little bit complicated. But the idea of this is to just combine the large sieve and this proposition. So if you start with the left-hand side and you just apply this proposition, then we get-- yeah, I guess we don't need that.

We get a sum, p and Pm of pi q of f high L2, pi q of g high and L2. So the proposition just said that this L-infinity norm is bounded by this product. I copied down the sum. And then this looks a lot like the input of the large sieve. So if I Cauchy-Schwarz it, I get the sum pi q of f high L2 squared to the 1/2, and then the sum pi q of g high L2 squared to the 1/2.

And then I apply the large sieve. So the large sieve gives a bound for each of these things and I plug it in. And what makes it a little bit messy is that, because there are two cases, the bound for this is a sum of two terms, and the bound for this is a sum of two terms. So then, when I multiply that together, I get this sum of two terms times that sum of two terms.

In the setting of our main theorem about primes, the theorem of Rényi, and Bombieri, and Vinogradov, the size of f and the size of g would be around 1 where they're defined. And so the L2 norms of f and g would be around N1 and N2, and squared would be around N2. So if you plug that all in, then we'd get our sum L infinity would be bounded by N/m, plus the square root of N1, N, plus the square root of N2, N, plus M root N.

And then, our goal would be that this sum is much smaller than N. So it's good as long as M is less than N to 1/2 minus epsilon. That's where that condition comes from, the range of q, and as long as N1 and N2 are much smaller than N. So in this part, I redid the computation at home and it's a little bit messy.

I did spend some time trying to think if there's a way of saying it without computing to see when it works well and when it doesn't work well. I have a couple comments about that, although I don't feel like I know how to see everything that way yet. So one comment is that it doesn't work very well if N1 or N2 is almost N. And there's a conceptual reason for that.

So suppose N2 was N. That means N1 would be 1. So my f is supported just on 1, and my g is supported all the way up to N. And when I convolve them together, I just get g times the value of f at 1. So in that situation there's no special structure of this function. It's a totally arbitrary function. And since this kind of theorem is not true for arbitrary functions, it can't work when N1 and N2 are really close to N.

So as long as N1 and N2 are kind of even, there's a meaningful convolution going on. As long as neither of them is close to 1, neither of them is close to N, there's a meaningful convolution structure going on. And in that setting, these terms are OK. This term here is OK as long as M is big. M is reasonably big. We need M to be reasonably big so that we're actually looking at many different projections. Because we're just looking at one projection, we can never say anything about it by just using projection theory

And here, we need M to be not too much bigger than the square root of N. And the reason for that is, so one important case here is when N1 and N2 are equal. So they would both be the square root of N. And if M is bigger than the square root of N, M would be bigger than both N1 and N2. And when that happens, the theorem might still be true, but the proof doesn't work very well.

You see, if M is bigger than N1 and N2, I have this function f that's defined on the numbers up to N1. And then, I tried projecting it mod q, where q is around M. So q is actually bigger than N1. So this projection operation is extremely uninteresting. The first N1 elements I copied down, and the rest I put 0. So since that projection operation is trivial, no theorem from projection theory is going to say anything useful here.

The gain that we're getting is that as long as-- so I need either N1 or N2 to be bigger than M in order to get off the ground. Once N1 or N2 is bigger than M, let's say N2 is bigger than M, then, when I look at these projections, most of them are almost constant. That's where the mileage is coming from. So these guys are almost constant.

These guys might or might not be almost constant. It's OK either way. And then, when I convolve it together, the convolution is almost constant. And in a strong sense, it's everywhere almost constant, instead of just most places almost constant. Actually, another visual thing is, so say this guy is almost constant but not everywhere. So in most places it's almost constant, but a couple of places it has a big bump.

When I take this convolution, then to evaluate the convolution at any given point, I have to look at the contribution of this at every point. And so those couple of places only contribute a little. And mostly I see the contribution of the places where this is small. So that's what's going on. And then you do have to do some algebra to see just how much we learned from that.

So now let's go back to our theorem. And this last part I'm not going to do every detail, which even gets a little bit messier. But I'll just say the skeleton and then we'll pause. So let's say that S is the primes up to N to 1/2. And then, remember that if, say, N over 2 is less than p, is less than N, or actually, just N to 1/2 is less than N, then R relatively prime to S of n is prime. So we can basically study this function by studying that function.

And then, we also know that relatively prime S of N is 1 convolved with Dp1, convolved with Dpr. And these are all multiplicative convolutions. This whole thing is a ton of multiplicative convolutions. But if I just want to see two of them, then I could say it's this guy convolved with that guy.

Now this function, as written, it goes on forever, and this function goes on forever. But I only want to evaluate this up to capital N. So this guy at N1 and this guy at N2, after they come together in the convolution, they'll only be relevant if N1 times N2 is less than capital N. So what I'll do is I'll write this as a sum on I1 of f times the characteristic function of I1. So I'll just divide up the numbers into some intervals.

And I'm going to convolve that with the sum over I2 of g characteristic function of I2. And I can open up the parentheses. It's the sum, I1 and I2, of f I1, f I1 convolved with g I2. And now if I only want to capture this up to capital N, so let me say that N1 is the maximum of I1, and N2 is the maximum of I2.

So for little n less than or equal to capital N, R Ps of n will be the sum over I1, I2 of f characteristic function of I1 convolved with g characteristic function of I2. And here, I can assume that N1 times N2 is less than around N. Maybe I should make this the minimum. Yeah, because otherwise the product of these two intervals will be bigger than capital N, so it won't have any influence here.

So then for each of these we apply the theorem, theorem A, that theorem there. And it tells me that when I reduce these things, mod q, for most q's, the non-constant part is small in L infinity. So this works for most of the terms. It works as long as N1 and N2 are much bigger than 1, and much less than N, but it could happen.

A piece of this would be this guy with I2 very close to 1, or even just the 0.1, convolved with that guy. That's in there, too. How do you deal with that? Well, this function f is itself a convolution. So then I can take f and I split it up, and I keep going.

So if N1 is almost N, and N2 is almost 1, use f is f1 convolved with f2, dot, dot, dot. So this part gets a little bit messy. And to be honest, I'm not positive whether this exact thing works. This is a possible project. So what I read in analytic number theory books is a slightly fancier way of taking this function, characteristic function of prime numbers, and writing it in terms of some multiplicative convolutions.

And for the slightly fancier way, there end up being less terms here, and it works out more cleanly. But it also, at least I had a little bit more trouble motivating it and seeing how people thought of it. Anyway, so I'm not positive. I think that this may work in all details, or maybe not quite. Maybe it requires more fidgeting. But nevertheless, I think it shows the main ideas that go into proving the Bombieri-Vinogradov theorem.

AUDIENCE: Sorry, I'm still really confused. So what are those intervals, I1 and I2?

LAWRENCE GUTH: What are the intervals I1 and I2? So you might first try dyadic intervals. And so modulo technical details, that's good. But actually, there's a little problem with taking dyadic intervals that wherever you stop, this function will be exactly the function that we want up to capital N. And then it'll go a little bit beyond capital N. And there'll be some stuff that's not exactly this function anymore, and then it'll die off.

And we'd like to prove that you just take this function and sharply cut it off at capital N. We'd like that to be evenly distributed. But we have this extra junk at the end. OK, so that's a problem. And the solution to that problem is we can greatly reduce the amount of junk by replacing dyadic intervals by intervals that are somewhat narrower. If we make I1 and I2 narrow, then whenever I1 times I2 crosses over N, it will only cross over a little bit, so we'll have only a little bit of junk. So that's who I1 and I2 are.

AUDIENCE: OK.

LAWRENCE GUTH: Yeah?

AUDIENCE: Can you recap what properties of R Ps we used?

LAWRENCE GUTH: Yeah. The question was, can we recap what properties of R Ps we used? So the first property is that R Ps detects primes. And that property is written here. So s is the primes that are less than N to 1/2. And on this big region, R Ps of n is the same as just characteristic function of primes. It tells us who's prime.

But this function is nicer than this function on the face of it because as a second good property, which is that it is a multiplicative convolution of a whole bunch of things together. And so we, in particular, can split it up like this. And we see it has this multiplicative convolution structure.

I guess another thing that I used without saying it as clearly as I should have is that we use that, pointwise, f and g are bounded by around 1. And that comes about for the following reason. Pointwise, all of these functions are bounded by 1 from the definitions. And then, if you look at f at a point N, well, there are different ways of factoring N. And so they contribute different amounts.

And there aren't very many ways to factor a number, and so f of N is pretty small. And I think that's it. I think that's all that we used about R Ps and about primes. Proving things about primes is a funny business because we don't know that much about primes. And so you have a theorem, something about the primes, you could ask, well, what input about the primes did we put into this theorem? And there's not that many different things that we know how to use as input anyway. And so for today it was this thing. Yeah?

AUDIENCE: In the Cauchy-Schwarz, what about, do we always lose something, or are there other situations like [INAUDIBLE]?

LAWRENCE GUTH: Good. It's a great question. And there is a specific question, but the big picture question was, at what steps in this argument did we lose something? So we proved a theorem here. It's a nice theorem, but it's not the whole truth. We put at the beginning the conjectures of what we believe are true, which are way stronger than this in several respects. Therefore, we must have lost some things in this theorem. Where did we lose them? Let's think through the proof.

OK, so this is an equality. The primes really are this thing. And this thing really is a convolution, so we haven't lost anything yet. We can now break the convolution up into pieces, like this. I don't think we lost too much here. There is this very annoying business about the I1, I2 going over the edge, and having to make I1 and I2 small. So there is some loss associated to that annoying business. So that's one piece of loss.

Then let's just focus on one of these pieces. That piece got fed into theorem A, so let's switch the two boards. So now we fed our piece into this theorem. This theorem is not usually sharp. So f and g are two particular functions. And for those two functions, probably something much stronger is true. Let's think about where we lost something.

So we wrote down the left-hand side. So far, so good. Then, we applied the proposition at, basically, we said this is a convolution. And you can bound the L-infinity norm of the convolution by the product of the L2 norms by Cauchy-Schwarz. This step here is a big loser. I didn't bring my red chalk, but I would put a big red star here. This is a big lossy step.

So remember that this guy is a high part, so it has average 0, and this guy has average 0. When we convolve them together there are a lot of positive and negative signs that appear. Usually there will be cancellation there, but we didn't take advantage of any of that cancellation. We just used the triangle inequality and so we got this upper bound. That's probably very lossy. So for our particular functions f and g, that step was probably very lossy.

Now this more general theorem holds for any functions f and g, which is nice. If you could choose any f and g, then for one particular q that may well have been correct. You could rig f and g so that for one particular q this Cauchy-Schwarz was sharp. But I don't know how to rig f and g so that for most of these P's or q's something like that happens. It sounds even unlikely to me that that is true.

So it's plausible that without any more input about prime numbers, it's plausible that there's a much stronger version of this theorem. But I have no idea how to prove it. OK, so that was our big lossy step. Then, we went on to the next step. We did this Cauchy-Schwarz. I don't think this one is so bad.

For this Cauchy-Schwarz to be sharp, we just means that it would be sharp if, for the different P's, all these L2 norms were about the same as each other. I think that usually happens. And then, we applied the large sieve. And the large sieve itself is usually not lossy in the key ranges. So this is the really important loss here and would require a big, new idea to do better. Cool. Nice questions, everybody. Let's break there, but I'll also hang out for office hours if there are more things that people want to talk about.