[SQUEAKING]

[PAPER RUSTLING]

[CLICKING]

YUFEI ZHAO: Last time we started talking about Roth's theorem, and we showed a Fourier analytic proof of Roth's theorem in the finite field model. So Roth's theorem in F3 to the N. And I want to today show you how to modify that proof to work in integers. And this will be basically Roth's original proof of his theorem.

OK. So what we'll prove today is the statement that the size of the largest 3AP-free subset of 1 through N is, at most, N divided by log log N. OK, so we'll prove a bound of this form.

The strategy of this proof will be very similar to the one that we had from last time. So let me review for you what is the strategy.

So from last time, the proof had three main steps. In the first step, we observed that if you are in the 3AP-free set then there exists a large Fourier coefficient. From this Fourier coefficient, we were able to extract a large subspace where there is a density increment. I want to modify that strategy so that we can work in the integers. Unlike an F2 to the N, where things were fairly nice and clean, because you have subspaces, you can take a Fourier coefficient, pass it down to a subspace. There is no subspaces, right? There are no subspaces in the integers. So we have to do something slightly different, but in the same spirit.

So you'll find a large Fourier coefficient. And we will find that there is density increment when you restrict not to subspaces, but what could play the role of subspaces when it comes to the integers? So I want something which looks like a smaller version of the original space. So instead of it being integers, if we restrict to a subprogression, so to a smaller arithmetic progression. I will show that you can restrict to a subprogression where you can obtain density increment.

So we'll restrict integers to something smaller. And then, same as last time, we can iterate this increment to obtain the conclusion that you have an upper bound on the size of this 3AP-free set.

OK, so that's the strategy. So you see the same strategy as the one we did last time, and many of the ingredients will have parallels, but the execution will be slightly different, especially in the second step where, because we no longer have sub spaces, which are nice and clean, so that's why we started with a finite fuel model, just to show how things work in a slightly easier setting. And today, we'll see how to do a same kind of strategy here, where there is going to be a bit more work. Not too much more, but a bit more work.

OK, so before we start, any questions?

All right. So last time I used the proof Roth's theorem as an excuse to introduce Fourier analysis. And we're going to see basically the same kind of Fourier analysis, but it's going to take on a slightly different form, because we're not working at F3 to the N. We're working inside the integers.

And there's a general theory of Fourier analysis on the group, on a billion groups. I don't want to go into that theory, because that's-- I want to focus on the specific case, but the point is that given the billion groups, you always have a dual group of characters. And they play the role of Fourier transform. Specifically in the case of Z, we have the following Fourier transform.

So the dual group of Z turns out to be the torus. So real numbers mod one. And the Fourier transform is defined as follows, starting with a function on the integers, OK. If you'd like, let's say it's finitely supported, just to make our lives a bit easier. Don't have to deal with technicalities. But in general, the following formula holds. We have this Fourier transform defined by setting f hat of theta to be the following sum. OK, where this e is actually somewhat standard notation, additive combinatorics. It's e to the 2 pi i t, all right? So it goes a fraction, t, around the complex unit circle.

OK. So that's the Fourier transform on the integers. OK, so you might have seen this before under a different name. This is usually called Fourier series. All right. You know, the notation may be slightly different. OK, so that's what we'll see today. And this Fourier transform plays the same role as the Fourier transform from last time, which was on the group F3 to the N.

And just us in-- so last time, we had a number of important identities, and we'll have the same kinds of identities here. So let me remind you what they are. And the proofs are all basically the same, so I won't show you the proofs. f hat of 0 is simply the sum of f over the domain. We have this Plancherel Parseval identity, which tells us that if you look at the inner product by linear form in the physical space, it equals to the inner product in the Fourier space.

OK. So in the physical space now, you sum. In the frequency space, you take integral over the torus, or the circle, in this case. It's a one-dimensional torus.

There is also the Fourier inversion formula, which now says that f of x is equal to f hat of theta. E of x theta, you integrate theta from 0 to 1. Again, on the torus, on the circle. And third-- and finally, there was this identity last time that related three-term arithmetic progressions to the Fourier transform, OK? So this last one was slightly not as-- I mean, it's not as standard as the first several, which are standard Fourier identities. But this one will be useful to us. So the identity relating the Fourier transform, the 3AP, now has the following form.

OK, so if we define lambda of f, g, and h to be the following sum, which sums over all 3APs in the integers, then one can write this expression in terms of the Fourier transform as follows. OK. All right. So comparing this formula to the one that we saw from last time, it's the same formula, where different domains, where you're summing or integrating, but it's the same formula.

And the proof is the same. So go look at the proof. It's the same proof. OK. So these are the key Fourier things that we'll use. And then we'll try to follow on those with-- the same as the proof as last time, and see where we can get.

So let me introduce one more notation. So I'll write lambda sub 3 of f to be lambda of f, f, f, three times. OK. So at this point, if you understood the lecture from last time, none of anything I've said so far should be surprising. We are working integers, so we should look at the corresponding Fourier transform in integers. And if you follow your notes, this is all the things that we're going to use.

OK, so what was one of the first things we mentioned regarding the Fourier transform from last time after this point?

OK.

AUDIENCE: The counting lemma.

YUFEI ZHAO: OK. So let's do a counting lemma. So what should the counting lemma say? Well, the spirit of the counting lemma is that if you have two functions that are close to each other-- and now, "close" means close in Fourier-- then their corresponding number of 3APs should be similar, OK? So that's what we want to say. And indeed, right, so the counting lemma for us will say that if f and g are functions on Z, and-- such that their L2 norms are both bounded by this M. OK, the sum of the squared absolute value entries are both bounded.

Then the difference of their 3AP counts should not be so different from each other if f and g are close in Fourier, OK? And that means that if all the Fourier coefficients of f minus g are small, then lambda 3ff, which considers 3AP counts in f, is close to that of g.

OK. Same kind of 3AP counting lemma from last time. OK, so let's prove it, OK? As with the counting lemma proofs you've seen several times already in this course, we will prove it by first writing this difference as a telescoping sum. The first term being f minus g f f, and then g of f minus g f and lambda g, g, f, f minus g. OK, and we will like to show that each of these terms is small if f minus g has small Fourier coefficients.

OK. So let's bound the first term. OK, so let me bound this first term using the 3AP identity, relating 3AP to Fourier coefficients, we can write this lambda as the following integral over Fourier coefficients.

And now, let me-- OK, so what was the trick last time? So we said let's pull out one of these guys and then use triangle inequality on the remaining factors. OK, so we'll do that. So far, so good. And now you see this integral. Apply Cauchy-Schwartz. OK, so apply Cauchy-Schwartz to the first factor, you got this l2 sum, this l2 integral. And then you apply Cauchy-Schwartz to the second factor. You get that integral.

OK, now what do we do? Yep?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: OK, so yeah. So you see an l2 of Fourier, the first automatic reaction should be to use a Plancherel or Parseval, OK? So apply Plancherel identity to each of these factors. We find that each of those factors is equal to this l2 sum in the physical space. OK, so this square root, the same thing. Square root again.

OK. And then we find that because there was an hypothesis-- in the hypothesis, there was a bound M on this sum of squares. You have that down there. And similarly, with the other two terms. OK. So that proves the counting lemma.

Question?

AUDIENCE: Last time, the term on the right-hand side was the maximum over non-zero frequency?

YUFEI ZHAO: OK. OK. So the question is, last time we had a counting lemma that looked slightly different. But I claimed they're all really the same counting lemma. They're all the same proofs. If you run this proof, it won't work. If you take what we did last time, it's the same kind of proofs. So last time we had a counting lemma where we had the same f, f, f essentially. We know have-- I allow you to essentially take three different things, and-- OK, so both-- in both cases, you're running through this calculation, but they look slightly different.

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: So, yeah. So I agree. It doesn't look exactly the same, but if you think about what's involved in the proof, they're the same proofs. OK. Any more questions? All right. So now, we have this counting lemma, so let's start our proof of Roth's theorem in the integers. As with last time, there will be three steps, as mentioned up there, OK? In the first step, let us show that if you are 3AP-free, then we can obtain a density-- a large Fourier coefficient.

Yeah, so in this course, this counting lemma, we actually solved this-- basically this kind of proof for the first time when we discussed graph counting lemma, back in the chapter on Szemerédi's regularity lemma. And sure, they all look literally-- not exactly the same, but they're all really the same kind of proofs, right? So I want-- I'm showing you the same thing in many different guises. But they're all the same proofs.

So if you are a set that is 3AP-free-- and as with last time, I'm going to call alpha the density of A now inside this progression, this length N progression. And suppose N is large enough. OK, so the conclusion now is that there exists some theta such that if you look at this sum over here as a sum over both integers-- actually, let me do the sum only from 1 to uppercase N. Claim that-- OK, so-- so it's saying what this title says. If you are 3AP-free and this N is large enough relative to the density, you think of this density alpha is a constant, then I can find a large Fourier coefficient.

Now, there's a small difference, and this is related to what you were asking earlier, between how we set things up now versus what happened last time. So last time, we just looked for a Fourier coefficient corresponding to a non-zero r. Now, I'm not restricting non-zero, but I don't start with an indicator function. I start with the demeaned indicator function. I take out the mean so that the zeroth coefficient, so to speak, which corresponds to the mean, is already 0.

So you don't get to use that for your coefficient.

So if you didn't do this, if you just tried to do this last time, I mean, you can also do exactly the same setup. But if you don't demean it, then-- if you don't have this term, then this statement is trivially true, because I can take theta equal to 0, OK? But I don't want. I want an actual significant Fourier improvement. So I take-- I do this demean, and then I consider its Fourier coefficient.

OK. Any questions about the statement?

Yeah, so this demeaning is really important, right? So that's something that's a very common technique whenever you do these kind of analysis. So make sure you're-- so that you're-- yeah, so you're looking at functions with mean 0.

Let's see the proof. We have the following information about all 3AP counts in A. Because A is 3AP-free, OK, so what is the value of lambda sub 3 of the indicator of A? Lambda of 3, if you look at the expression, it basically sums over all 3APs, but A has no 3APs, except for the trivial ones. So we'll only consider the trivial 3APs, which has size exactly the size of A, which is alpha N from trivial 3APs.

On the other hand, what do we know about lambda 3 of this interval from 1 to N? OK, so how many 3APs are there? OK, so roughly, it's going to be about N squared over 2. And in fact, it will be at least N squared over 2, because to generate a 3AP, I just have to pick a first term and a third term, and I'm OK as long as they're the same parity. And then you have a 3AP. So the same parity cuts you down by half, so you have at least N squared over 2 3APs from 1 through N.

So now, let's look at how to apply the counting lemma all to the setting. So we have the counting lemma up there, where I now want to apply it-- so apply counting to, on one hand, the indicator function of A so we get the count 3APs in A, but also compared to the normalized indicator on the interval.

OK, so maybe this is a good point for me to pause and remind you that the spirit of this whole proof is understanding structure versus pseudorandomness, OK? So as was the case last time. So we want to understand, in what ways is A pseudorandom? And here, "pseudorandom," just as with last time, means having small Fourier coefficients, being Fourier uniform. If A is pseudorandom, which here, means f and g are close to each other. That's what being pseudorandom means, then the counting lemma will tell us that f and g should have similar AP counts.

But A has basically no AP count, so they should not be close to each other. So that's the strategy, to show that A is not pseudorandom in this sense, and thereby extracting a large Fourier coefficient. So we apply counting to these two functions, and we obtain that.

OK. So this quantity, which corresponds to lambda 3 of g, minus alpha N. So these were lambda 3 of g, lambda 3 of F. So it is upper-bounded. The difference is up rebounded by the-- using the counting lemma, we find that their difference is upper-bounded by the following quantity. Namely, you look at the difference between f and g and evaluate its maximum Fourier coefficient. OK. So if A is pseudorandom, meaning that Fourier uniform-- this l infinity norm is small, then I should expect lots and lots of 3APs in A, but because that is not the case, we should be able to conclude that there is some large Fourier coefficient.

All right, so thus-- so rearranging the equation above, we have that-- so this should be a square. OK. So we have this expression here. And now we are-- OK, so let me simplify this expression slightly. And now we're using that N is sufficiently large, OK? So we're using N is sufficiently large. So this quantity is at least a tenth of alpha squared N.

OK, and that's the conclusion, all right? So that's the conclusion of this step here. What does this mean? This means there exists some theta so that the Fourier coefficient at theta is at least the claimed quantity.

Any questions? All right. So that finishes step 1. So now let me go on step 2. In step 2, we wish to show that if you have a large Fourier coefficient, then one can obtain a density increment.

So last time, we were working in a finite field vector space. A Fourier coefficient, OK, so which is a dual vector, corresponds to some hyperplane. And having a large Fourier coefficient then implies that the density of A on the co-sets of those hyperplanes must be not all close to each other. All right, so one of the hyperplanes must have significantly higher density than the rest.

OK, so we want to do something similar here, except we run into this technical difficulty where there are no subspaces anymore. So the Fourier character, namely corresponding to this theta, is just a real number. It doesn't divide up your space. It doesn't divide up your 1 through N very nicely into sub chunks. But we still want to use this theta to chop up 1 through N into smaller spaces so that we can iterate and do density increment.

All right. So let's see what we can do. So given this theta, what we would like to do is to partition this 1 through N into subprogressions. OK, so chop up 1 through N into sub APs such that if you evaluate for-- so this theta is fixed. So on each sub AP, this function here is roughly constant on each of your parts.

Last time, we had this Fourier character, and then we chopped it up using these three hyperplanes. And each hyperplane, the Fourier character is literally constant, OK? So you have-- and so that's what we work with. And now, you cannot get them to be exactly constant, but the next best thing we can hope for is to get this Fourier character to be roughly constant.

OK, so we're going to do some positioning that allows us to achieve this characteristic. And let me give you some intuition about why this is true. And this is not exactly a surprising fact. The intuition is just that if you look at what this function behaves like-- all right, so what's going on here? You are on the unit circle, and you are jumping by theta. OK, so you just keep jumping by theta and so on.

And I want to show that I can sharp up my progression into a bunch of almost periodic pieces, where in each part, I'm staying inside a small arc. So in the extreme case of this where it is very easy to see is if x is some rational number, a over b, with b fairly small, then we can-- so then, this character is actually constant on APs with common difference b. Yep?

AUDIENCE: Is theta supposed to be [INAUDIBLE]?

YUFEI ZHAO: Ah, so theta, yes. so theta-- thank you. So theta 2 pi.

AUDIENCE: Like, is x equal to your theta?

YUFEI ZHAO: Yeah. Thank you. So theta equals-- yeah. So if theta is some rational with some small denominator-- so then you are literally jumping in periodic steps on the unit circle. So if you partition N according to the exact same periods, you have that this character is exactly constant in each of your progressions.

Now, in general, the theta you get out of that proof might not have this very nice form, but we can at least approximately achieve the desired effect. OK. Any questions?

OK. So to achieve approximately the desired effect, what we'll do is to find something so that b times theta is not quite an integer, but very close to an integer. OK, so this, probably many of you have seen before. It's a classic pigeonhole-type result. It's usually attributed to Dirichlet.

So if you have theta, a real number, and a delta, kind of a tolerance, then there exists a positive integer d at most 1 over delta such that d times theta is very close to an integer. OK, so this norm here is distance to the closest integer. All right, so the proof is by pigeonhole principle.

So if we let N be 1 over delta rounded down and consider the numbers 0, theta, 2 theta, 3 theta, and so on, to N theta-- so by pigeonhole, there exists i theta and j theta, so two different terms of the sequence such that they differ by less than-- at most delta in their fractional parts. OK, so now take d to be difference between i and j. OK, and that works.

OK. So even though you don't have exactly rational, you have approximately rational. So this is a-- it's a simple rational approximation statement. And using this rational approximation, we can now try to do the intuition here, pretending that we're working with rational numbers, indeed. OK, so if we take eta between 0 and 1 and theta irrational and suppose N is large enough-- OK, so here, C means there exists some sufficiently large-- some constant C such that the statement is true, OK? So suppose you think a million here. That should be fine.

So then there exists-- so then one can partition 1 through N into sub-APs, which we'll call P i. And each having length between cube root of N and twice the cube root of N such that this character that we want to stay roughly constant indeed does not change very much. If you look at two terms in the same AP, in the sub-AP, then the value of this character on each P sub i is roughly the same. So they don't vary by more than eta on each P i. So here, we're partitioning this 1 through N into a sub-A piece so that this guy here stays roughly constant.

OK. Any questions? All right. So think about how you might prove this. Let's take a quick break.

So you see, we are basically following the same strategy as the proof from last time, but this second step, which we're on right now, needs to be somewhat modified because you cannot cut this space up into pieces where your character is constant. Well, if they're roughly constant then we're go to go, so that's what we're doing now. So let's prove the statement up there.

All right. So let's prove this statement over here. So using Dirichlet's lemma, we find that there exists some d. OK, so I'll write down some number for now. Don't worry about it. It will come up shortly why I write this specific quantity. So there exists some d which is not too big, such that d theta is very close to an integer. So now, I'm literally applying Dirichlet's lemma.

OK. So given such d-- so how big is this d? You see that because I assumed that N is sufficiently large, if we choose that C large enough, d is at most root N. So given such d, which is at most root N, you can partition 1 through N into subprogressions with common difference d. Essentially, look at-- let's do classes mod d. So they're all going to have length by basically N over d. And I chop them up a little bit further to get-- so a piece of length between cube root of N and twice cube root of N.

OK. So I'm going to make sure that all of my APs are roughly the same length. And now, inside each subprogression-- let me call this subprogression P prime, subprogression P, let's look at how much this character value can vary inside this progression.

All right. OK, so how much can this vary? Well, because theta is such that d times theta is very close to an integer and the length of each progression is not too large-- so here's-- I want some control on the length. So we find that the maximum variation is, at most, the size of P, the length of P, times-- so this-- that difference over there. So all of these are exponential, so I can shift them.

Well, the length of P is at most twice cube root of N. And-- OK, so what is this quantity? So the point is that if this fractional part here is very close to an integer, then e to that, e to the 2 pi times that i times some number should be very close to 1, because what is happening here? This is the distance between those two points on the circle, which is at most bounded by the length of the arc. OK, so cord length, at most of the arc.

So now, you put everything here together, and apply the bound that we got on d theta. So this is the reason for choosing that weird number up there. We find that the variation within each progression is at most eta, right? So the variation of this character within each progression is not very large, OK? And that's the claim.

Any questions?

All right, so this is the analogous claim to the one that we had-- the one that we used last time, where we said that the character is constant on each coset of the hyperplane. They're not exactly constant, but almost good enough.

All right. So the goal of step 2 is to show an energy-- show a density increment, that if you have a large Fourier coefficient, then we want to claim that the density goes up significantly on some subprogression. And the next part, the next lemma, will get us to that goal. And this part is very similar to the one that we saw from last time, but with this new partition in mind, like I said.

If you have A that is 3AP-free with density alpha, and N is large enough, then there exists some subprogression P. So by subprogression, I just mean that I'm starting with original progression 1 through N, and I'm zooming into some subprogression, with the length of P fairly long, so the length of P is at least cube root of N, and such that A, when restricted to this subprogression, has a density increment.

OK, so originally, the density of A is alpha, so we're zooming into some subprogression P, which is a pretty long subprogression, where the density goes up significantly from A to essentially A-- from alpha to roughly alpha plus alpha squared. OK. So we start with A, a 3AP-free set. So from step 1, there exists some theta with large-- so that corresponds to a large Fourier coefficient. So this sum here is large.

OK, and now we use-- OK, so-- so step 1 obtains us, you know, this consequence. And from this theta, now we apply the lemma up there to-- so we apply lemma with, let's say, eta being alpha squared over 30. OK, so the exact constants are not so important. But when we apply the lemma to partition, and into a bunch of subprogressions, which we'll call P1 through Pk. And each of these progressions have length between cube root of N and twice cube root of N.

And I want to understand what happens to the density of A when restricted to these progressions. So starting with this inequality over here, which suggests to us that there must be some deviation. OK, so starting with what we saw. And now, inside each progression this e x theta is roughly constant. So if you pretend them as actually constant, I can break up the sum, depending on where the x's lie. So i from 1 to k. And let me sum inside each progression.

So by triangle inequality, I can upper bound the first sum by where I now cut the sum into progression by progression. And on each progression, this character is roughly constant. So let me take out the maximum possible deviations from them being constant. So upper bound-- again, you'll find that we can essentially pretend-- all right, so if each exponential is constant on each subprogression, then I might as well just have this sum here.

But I lose a little bit, because it's not exactly constant. It's almost constant. So I loose a little bit. And that little bit is this eta. So you lose that little bit of eta. And so on each progression, P i, you lose at most something that's essentially of alpha squared times the length of P i.

OK. Now, you see, I've chosen the error parameter so that everything I've lost is not so much more than the initial bound I began with. So in particular, we see that even if we had pretended that the characters were constant, on each progression we would have still obtained some lower bound of the total deviation.

OK. And what is this quantity over here? Oh, you see, I'm restricting each sum to each subprogression, but the sum here, even though it's the sum, but it's really counting how many elements of A are in that progression. So this sum over here is the same thing. OK, so let me write it in a new board. Oh, we don't need step 1 anymore.

All right. So what we have-- OK, so left-hand side over there is this quantity here, all right? We see that the right-hand side, even though you have that sum, it is really just counting how many elements of A are in each progression versus how many you should expect based on the overall density of A. OK, so that should look similar to what we got last time.

I know the intuition should be that, well, if the average deviation is large, then one of them, one of these terms, should have the density increment. If you try to do the next step somewhat naively, you run into an issue, because it could be-- now, here you have k terms. It could be that you have all the densities except for one going up only slightly, and one density dropping dramatically, in which case you may not have a significant density increment, all right?

So we want to show that on some progression the density increases significantly. So far, from this inequality, we just know that there is some subprogression where the density changes significantly. But of course, the overall density, the average density, should remain constant. So if some goes up, others must go down. But if you just try to do an averaging argument, you have to be careful, OK?

So there was a trick last time, which we didn't really need last time, but now, it's much more useful, where I want to show that if this holds, then some P i sees a large energy-- sees a large density increment. And to do that, let me rewrite the sum as the following, so I keep the same expression. And I add a term, which is the same thing, but without the absolute value.

OK, so you see these guys, they total to 0, so adding that term doesn't change my expression. But now, the summand is always non-negative. So it's either 0 or twice this number, depending on the sign of that number.

OK. So comparing left-hand side and right-hand side, we see that there must be some i. So hence, there exists some eye such that the left-hand side-- the i-th term on the left hand side is less than or equal to the i-th term on the right-hand side. And in particular, that term should be positive, so it implies-- OK, so how can you get this inequality? It implies simply that the restriction of a to this P i is at least alpha plus alpha squared over 40 times P i. So this claim here just says that on the i-th progression, there's a significant energy increment. If it's more decrement, that term would have been 0. So remember that.

OK. So this achieves what we were looking for in step 2, namely to find that there is a density increment on some long subprogression. OK, so now we can go to step 3, which is basically the same as what we saw last time, where now we want to iterate this density increment.

OK, so it's basically the same argument as last time, but you start with density alpha, and each step in the iteration, the density goes up quite a bit. And we want to control the total number of steps, knowing that the final density is at most 1, always. OK.

So how many steps can you take? Right, so this was the same argument that we saw last time. We see that starting with alpha, alpha 0 being alpha, it doubles after a certain number of steps, right? So we double after-- OK, so how many steps do you need? Well, I want to get from alpha to 2 alpha. So I need at most alpha over 40 steps. OK, so last time I was slightly sloppy. And so there's basically a floor, upper floor down-- rounding up or down situation. But I should add a plus 1. Yeah.

AUDIENCE: Shouldn't it be 40 over alpha?

YUFEI ZHAO: 40-- thank you. 40 over alpha, yeah. So you double after at most that many steps. And then now, you add density at least 2 alpha. So we double after at most 20 over alpha steps, and so on. And we double at most-- well, basically, log sub 2 of 1 over alpha times. OK, so anyway, putting everything together, we see that the total number of steps is at most on the order of 1 over alpha.

When you stop, you can only stop for one reason, because in the-- yeah. So, yeah. So in step 1, remember, the iteration said that the process terminates. The process can always go on, and then terminates if the length-- so you're now at step i, so let N i be the length of the progression, that step i is at most C times alpha i to the minus 12th. So we have a-- right, so provided that N is large enough, you can always pass to a subprogression. And here, when you pass to subprogression, of course, you can re-label that subprogression. And it's now, you know, 1 through N i. Right, so I can-- it's-- all the progressions are basically the same as the first set of positive integers. I'm sorry, prefix of the positive integers.

So when we stop a step i, you must have N sub i being at most this quantity over here, which is at most C times the initial density raised to this minus 12th. So therefore, the initial length N of the space is bounded by-- well, each time we went down by a cube root at most. Right, so the fine-- if you stop a step i, then the initial length is at most N sub i to the 3 times 3 to the power of i each time you're doing a cube root.

OK, so you put everything together. At most that many iterations when you stop the length is at most this. So you put them together, and then you find that the N must be at most double exponential in 1 over the density. In other words, the density is at most 1 over log log N, which is what we claimed in Roth's theorem, so what we claimed up there.

OK. So that finishes the proof. Any questions?

So the message here is that it's the same proof as last time, but we need to do a bit more work. And none of this work is difficult, but there are more technical. And that's often the theme that you see in additive combinatorics. This is part of the reason why the finite field model is a really nice playground, because there, things tend to be often cleaner, but the idea's often similar, or the same ideas. Not always. Next lecture, we'll see one technique where there's a dramatic difference between the finite field vector space and over the integers. But for many things in additive combinatorics, the finite field vector space is just a nicer place to be in to try all your ideas and techniques.

Let me comment on some analogies between these two approaches and compare the bounds. So on one hand-- OK, so we saw last time this proof in F3 to the N, and now in the integers inside this interval of length N. So let me write uppercase N in both cases to be the size of the overall ambient space. OK, so what kind of bounds do we get in both situations? So last time, for-- in F3 to the N, we got a bound which is of the order N over log N, whereas today, the bound is somewhat worse. It's a little bit worse. Now we lose an extra log.

So where do we lose an extra log in this argument? So where does these two argument-- where do these two arguments differ, quantitatively? Yep?

AUDIENCE: When you're dividing by 3 versus [INAUDIBLE]?

YUFEI ZHAO: OK, so you're dividing by-- so here, in each iteration, over here, your size of the iteration-- I mean, each iteration the size of the space goes down by a factor of 3, whereas over here, it could go down by a cube root. And that's precisely right. So that explains for this extra log in the balance.

So while this is a great analogy, it's not a perfect analogy. You see there is this divergence here between the two situations. And so then you might ask, is there some way to avoid the loss, this extra log factor loss over here? Is there some way to carry out the strategy that we did last time in a way that is much more faithful to that strategy of passing down to subspaces. So here, we pass to progressions. And because we have to do this extra pigeonhole-type argument, it was somewhat lost-- we lost a power, which translated into this extra log.

So it turns out there is some way to do this. So let me just briefly mention what's the idea that is involved, all right? So last time, we went down from-- so the main objects that we were passing would start with a vector space and pass down to a subspace, which is also a vector space, right? So you can define subspaces in F3 to the N by the following. So I can start with some set of characters U, and I define-- some set of characters S, and I define U sub S to be basically the orthogonal complement of S.

OK, so this is a subspace. And these were the kind of subspace that we saw last time, because the S's or the R's that came out of the proof last time, every time we saw one, we threw it in. We cut down to a smaller subspace, and we repeat. But the progressions, they don't really look like this. So the question is, is there some way to do this argument so that you end up with progressions, and looked like that?

And it turns out there is a way. And there are these objects, which we'll see more later in this course, called Bohr sets. OK, so they were used by Bourgain to mimic this Machoulin argument that we saw last time more faithfully into the integers, where we're going to come up with some set of integers that resemble-- much more closely resemble this notion of subspaces in the finite field setting.

And for this, it's much easier to work inside a group. So instead of working in the integers, let's work inside and Z mod nZ. So we can do Fourier transform in Z mod nZ, so the discrete Fourier analysis here. So in Z mod nZ, we define-- so given a S, let's define this Bohr set to be the set of elements of Z mod nZ such that if you look at what really is supposed to resemble this thing over here, OK? If this quantity is small for all S-- OK, so we put that element into this Bohr set.

OK, so these sets, they function much more like subspaces. So there are the analog of subspaces inside Z mod nZ, which, you know, n is prime, has no subgroups. It has no natural subspace structure. But by looking at these Bohr sets, they provide a natural way to set up this argument so that you can-- but with much more technicalities, repeat these kind of arguments more similar to last time, but passing not to subspaces but to Bohr sets. And then with quite a bit of extra work, one can obtain bounds of the quantity N over a poly log N. So the current best bound I mentioned last time is of this type, which is through further refinements of this technique.

The last thing I want to mention today is, so far we've been talking about 3APs. So what about four term arithmetic progressions? OK, do any of the things that we talk about here work for 4APs? And there's an analogy to be made here compared to what we discussed with graphs. So in graphs, we had a triangle counting lemma and a triangle removal lemma. And then we said that to prove 4APs, we would need the hypergraph version, the simplex removal lemma, hypergraph regularity lemma. And that was much more difficult.

And that analogy carries through, and the same kind of difficulties that come up. So it can be done, but you need something more. And the main message I want you to take away is that 4APs, while we had a counting lemma that says that the Fourier coefficients, so the Fourier transform, controls 3AP counts, it turns out the same is not true for 4APs. So the Fourier does not control 4AP counts.

Let me give you some-- OK. So in fact, in the homework for this week, there's a specific example of a set where it has uniformly small Fourier coefficients. But that's the wrong number of 4APs. So the following-- it is true-- OK, so it is true that you have Szemerédi's term in-- let's just talk about the finite field setting, where things are a bit easier to discuss. So it is true that the size of the biggest subset of F5 to the N is a tiny fraction-- it's a little, one fraction of the entire space. OK, I use F5 here, because if I set F3, it doesn't make sense to talk about 4APs. So F5, but it doesn't really matter which specific field.

So you can prove this using hypergraph removal, same proof, verbatim, that we saw earlier, if you have hypergraph removal. But if you want to try to prove it using Fourier analysis, well, it doesn't work quite using the same strategy. But in fact, there is a modification that would allow you to make it work. But you need an extension of Fourier analysis. And it is known as higher order Fourier analysis, which was an important development in modern additive combinatorics that initially arose in Gowers' work where he gave a new proof of similarities theorem. So

Gowers didn't work in this setting. He worked in integers. But many of the ideas originated from his paper, and then subsequently developed by a lot of people in various settings.

I just want to give you one specific statement, what this high-order Fourier analysis looks like. So it's a fancy term, and the statements often get very technical. But I just want to give you one concrete thing to take away. All right, so for a Fourier piece, higher-order Fourier analysis, roughly-- OK, so it also goes by the name quadratic Fourier analysis.

OK, so let me give you a very specific instance of the theorem. And this can be sometimes called an inverse theorem for quadratic Fourier analysis. OK, so for every delta, there exists some c such that the following is true. If A is a subset of a F5 to the N with density alpha and such that it's-- OK, so now lambda sub 4, so this is the 4AP density, so similar to 3AP, but now you write four terms. The 4AP density of A differs from alpha to the fourth by a significant amount. OK, so for 3APs, then we said that now A has a large Fourier coefficient, right?

So for-- OK. For 4APs, that may not be true, but the following is true, right? So then there exists a non-zero quadratic polynomial. F and N variables over F5 such that the indicator function of A correlates with this quadratic exponential face.

So Fourier analysis, the conclusion that we got from counting lemma is that you have some linear function F, such that this quantity is large, this large Fourier coefficient. OK, so that is not true 4APs. But what is true is that now you can look at quadratic exponential faces, and then it is true.

So that's the content of higher order Fourier. I mean, that's the example of higher-order Fourier analysis. And you can imagine with this type of result, and with quite a bit more work, you can try to follow a similar density increment strategy to prove similarities term for 4APs.