YUFEI ZHAO: For the past couple lectures, we've been talking about Roth's theorem. And we showed-- so we saw a proof of Roth's theorem using Fourier analytic methods. And we saw basically the same proof but in two different settings. So two lectures ago, we saw a proof in F3 to the M And basically the same strategy, but with a bit more work, we were able to show Roth's theorem worth roughly comparable bounds over the integers.

Today, I want to show you a very different kind of proof of Roth's theorem in the finite fieldfini setting. So first let me remind you, the bound that we saw last time for Roth's in F3 to the M gave an upper bound on the maximum number of elements in the 3-AP-free set that were of the form 3 to the n over n. And so this proof wasn't too bad. So we did it in one lecture.

And then with a lot more work-- and people tried very, very hard to improve this-- and there was a paper that got it to just a little bit more. And this was a lot of work. And this was something that people thought was very exciting at the time.

And then just a few years ago, there was a major breakthrough, a very surprising breakthrough, where-- you know, at this point, it wasn't even clear whether 3 should be the right base for this exponent. That was a big open problem. And then there was a big breakthrough where the following bound was proved, that it was exponentially less than the previous bound. So this is one that I want to talk about in the first part of today's lecture.

So this development came first-- the history is a bit interesting. So Croot, Lev, and Pach uploaded a paper to the archive May 5 of 2016, where they showed not exactly this theorem but in a slightly different setting in this group, so in Z mod 4 instead of Z mod 3. And this was already quite exciting, getting exponential improvement in this setting.

But it wasn't exactly obvious how to use their method to get F3. But that was done about a week later. So Ellenberg and Gijswijt, they managed to improve the-- use this technique to modify the Croot-Lev-Pach technique to the F2 to the n setting, which is the one that we've been interested in. So there's a small difference between these two, namely this group has elements of order 2, which makes things a bit easier to do it here with.

So this is the Croot-Lev-Pach method, as it's often called in literature. And we'll see that-- it's a very ingenious use of the so-called linear algebraic method in combinatorics, in this case the polynomial method. And it works specifically in the finite field vector space. So what we're

talking about in this part of the lecture does not translate whatsoever. At least, nobody knows how to translate this technique to the integer setting.

So how does it work? The presentation I'm going to give follows not the original paper, which is quite nice to read, by the way. It's only about four pages long. It's pleasant to read. But there's is a slightly even nicer formulation on Terry Tao's blog. And that's the one that I'm presenting.

So the idea is that if you have a subset of F3 to the n that is 3-AP-free, such a set also has a name capset, which is also used in literature in this specific setting where you have no three points on the line. In this case, then we have the following identity. So here delta is the Dirac delta. Let me write that down in a second.

So the delta of a is the Dirac delta. It's either 1 if x equals to a, and 0 if x does not equal to a. So this is simply rewriting the fact that x, y, z form a 3-AP if and only if their sum is equal to 0. And because you're 3-AP-free, the only 3-AP's are the trivial ones recorded on the right-hand side. So this is simply a recording of the statement that A is 3-AP-free.

And the idea now is that you have this expression up there, and I want to show that if A is very, very large, then I could get a contradiction by considering some notion of rank. So we will show that the left-hand side is, in some sense, low rank. Well, I haven't told you what rank means yet. But the left-hand side is somewhat low rank, and the right-hand side is a high-rank object.

So what does rank mean. So recall from linear algebra-- so the classical notion of rank corresponds to two variable functions. So you should think of F as a matrix over an arbitrary field F. So such a function or a corresponding matrix is called rank 1 if it is nonzero and it can be written in the following form-- F of x, y is f of x g of y for some functions that are one variable each. So, in matrix language, this is a column vector times a row vector.

So that's the meaning of rank 1. And to say that something is of high rank of a specific rank-- rather, the rank of F is defined to be the minimum number of rank 1 functions needed to write F as a sum or a linear combination. So this is rank 1. And if you add up r rank 1 functions, then get something that's, at most, rank r. So that's the basic definition of rank from linear algebra.

For three-variable functions, you can come up with other notions of rank. So what about three-variable functions? So how do we define a rank of such a function? So you might have seen

such objects as generalizations of matrices called tensors. And tensors have, already, a natural notion of rank, and this is called tensor rank. Just like how, here, F is-- we say rank 1 if it's decomposable like that, we say F has tensor rank 1 if this three-variable function is decomposable as a product of one-variable functions.

The tensor rank, it turns out, this is an important notion, which is actually quite mysterious. There's a lot of important problems that boil down to us not really understanding what tensor rank, how it behaves. And it turns out, this is not the right notion to use for our problem. So we're going to use a different notion of rank. Here, rank 1 is decomposing this three-variable function into a product of three one-variable functions.

But, instead, I can define a different notion. We say that F has slice rank 1-- so this is a definition that's introduced in the context of this problem, although it's also quite a natural definition-- if it has one of the following forms. So I can write it as a product of a one-variable function and a two-variable function. So one variable and the remaining two variables. But this definition should also be symmetric in the variables, so the other combinations are OK as well.

So this is the definition of a rank one function, a slice rank 1. And, also, if nonzero. If it's nonzero and can be written in one of these forms. And, just like earlier, we define the slice rank of F to be the minimum number of slice rank 1 functions. Same as before, that you need to write F as a sum. So I can decompose this F into a sum of slice rank 1 functions. What's the most efficient way to do so?

So that's the definition of slice rank. And, you see, you can come up with this definition for any number of variables, where slice rank 1 means decompose into two functions, where one function takes one variable, and the other function takes all the remaining variables. And, therefore, two variables, slice rank and rank correspond to the same notion. Any questions so far? All right.

So let's look at the function on the right. So think of it as a matrix, a tensor. So what is it? Well, it's kind of like a diagonal matrix. So that's what it is. It's a diagonal matrix. So what is the rank of a diagonal matrix, in this case a diagonal function? Well, you know from linear algebra that if you have a matrix, then the rank of a diagonal matrix is the number of nonzero entries. So something similar is true for slice rank, although it's less obvious. It will require a proof.

So if I have this three-variable function defined by the following formula. So, in other words, it's a diagonal function where the entries on the diagonals are the Ca's. So what is the rank of this

function? So the slice rank of F. In the matrix case, it will be the number of nonzero entries, and it's exactly the same here. So number of nonzero diagonal entries. That turns out to be the slice rank.

Let's see a proof. So we go back to the definition of slice rank. And we see that one of the directions is easy. So this less than or equal to, greater than or equal to-- so which one is easy? So, you see, the right-hand side is a sum of r-- of a-- well, this many rank 1 functions. So this direction is-- so this direction is clear, just looking at the definition. I can write F explicitly as that many rank 1, slice rank 1 functions.

So the tricky part is greater than or equal to. And for the greater than or equal to, let's assume that all the diagonal entries are nonzero. So why can we do this? If it's not nonzero, I claim that we can remove this element from A. If the Ca is not 0, then I remove a from the set. And doing so cannot increase the rank.

A priori, the rank might go down if you get rid of an entry. Because if you add an entry, even though the function doesn't change on the original set, if you increase your set, maybe you have more space, maybe you have more flexibility to work with. But, certainly, if you remove an element, the rank cannot go up.

Now, so suppose the slice rank of F is strictly less than the size of A. So all these Ca's are nonzero. So suppose, for contradiction, that there is some different way to write function F that uses fewer terms. So what would such a sum look like? So I would be able to write this function F in a different way. Like that. And then, now, I look at these-- the other types of functions using different combination of the variables.

So suppose there were a different way to write this function F that uses fewer terms. So I assume it uses exactly the size of A minus 1 terms, and always putting zero functions if you like. So now I claim that there exists a function h on the set A whose support-- so the support is the number of entries that give nonzero values. The support of F is bigger than m, such that the following sum is 0.

So I claim that we can find a function F-- h such that I think of it as in the kernel of some of these f's. So this is a linear algebraic statement. Yes.

AUDIENCE: What is h sub [INAUDIBLE]?

YUFEI ZHAO: Ah, sorry. It's just h. Thank you. It's a single function h such that this equation is true for all x.

AUDIENCE: [INAUDIBLE] h of x minus the sum of all [INAUDIBLE].

YUFEI ZHAO: You are right. So what do I want to say here? So we want to find a function h such that the support of h is at least m. So what do we want to say? I want to say that-- yes, so you're right. This is not what I want to say. And, instead, it's something-- mm-hmm. Yes, good. So, let's see.

So here we have some number of functions. Here, we have some number of functions. And for each a, I have-- or for each-- let's see. Umm, hmm.

AUDIENCE: [INAUDIBLE].

YUFEI ZHAO: I'm sorry?

AUDIENCE: [INAUDIBLE].

YUFEI ZHAO: No. So I do want to show-- no, there's no induction, because I'm in three variables, and I want to get rid of-- so the point is-- so let's see where we're going eventually, and then we'll figure out what happened up there. So we want to consider-- so I would like to eventually consider the following sum. So I want to consider this sum, which comes from-- so you look at-- wait, no. That's not the sum I want to consider.

So let's look at this F of x, y, z, so F being that sum. No. So take that F up there. And let me consider, basically, taking the inner product of this function viewed as a function in z. So consider this inner product. And if I-- ah. I think-- so what I want to say is not this.

So what I want to say is, if I look at an inner product of h with the-- so take one of these f's-- take one of these f's and look at the bilinear form relating each in f. So I want to show that this sum vanishes for all i between m plus 1 and the size of A minus 1. So this row, I want it to vanish when being taken bilinear form with h. So that makes sense now. OK, good.

So the fact that such a nonzero h exists simply is a matter of counting parameters. It's a linear algebraic statement. You have some number of freedoms. You have some number of constraints. So the set of such h satisfy all of these constraints. So there are this many constraints. Well, each one of them could carry down to one dimension less, but the set of such h is a linear subspace of dimension bigger than m, because I have A dimensions, and I

have these many constraints. So the set of such h is-- there are a lot of possibilities.

And, furthermore, it is also true that-- and this is a linear algebraic statement-- that every subspace of dimension m plus one has a vector whose support has size at least m plus 1. I'll leave this as a linear algebraic exercise. It's not entirely obvious, but it is true. When you put these two things together, you find that there is some vector-- so I think of the corners of the vectors as indexed by the set A-- there is some vector whose support is large enough.

So we prove the claim. Let's go back to this lemma about this diagonal function having high rank. Take h from the claim. So let's take h from the claim. Then let's consider this sum over here. On one hand, what this sum is-- you can do the sum on the right-hand side. We see that it's like multiplying a diagonal matrix by a vector. So what you get, following the formula on the right-hand side, is the following. Let me rewrite this part. Sum over a of C sub a h of a delta sub a of x delta sub a of y. Just looking at the formula from the right hand side.

On the other hand, if you had a decomposition up there, doing this sum and noting the claim, we see that the third row is gone. So what you would have is a sum over these z's of-- so let me write that like this. So you would have a sum that is of the form f1 of x and g tilde 1 of y, where g tilde is basically the inner product of g1 as a function of z with h. So fl of x gl of y. And then, also, functions like that. So there exists some functions g, which come from g tilde, which come from the g's up there, such that this is true.

But now we're in the world of two-variable functions. So left and right-hand side are two-variable functions. And for two-variable functions, you understand what is the rank of a diagonal function. So the left-hand side has more than m diagonal entries, because h has support. So the number of diagonal entries is just the support of h. Whereas the right-hand side has rank-- so now a linear algebraic matrix rank-- at most, m. And that's a contradiction. Yes.

AUDIENCE: So you can show a similar statement where [INAUDIBLE].

YUFEI ZHAO: Great. So we can show a similar statement for arbitrary number of variables by generalizing this proof and using induction on the number of variables. But we only need three variables for now. Any questions? Just to recap, what we proved is the generalization of the statement that a diagonal matrix has rank equal to the number of nonzero diagonal entries. But the same fact is true for these three-variable functions with respect to slice rank. So this is intuitively obvious, but the execution is slightly tricky.

All right. So now we have the statement here. Let's proceed to analyze this function which comes from-- so this relationship here coming from set A that is 3-AP-free. So suppose now I'm in-- so let me-- so everything so far was generally with any A. But now let me think about, specifically, functions on the finite field vector space, F3 to the n. So it's a function taking value F3.

And this function is defined to be the left-hand side of that equation over there. So the claim is that-- so the left-hand side claim that this function has low rank. So we claim that a slice rank of this function is, at most, 3M, where M is the sum of, essentially, this multinomial coefficient. So we'll analyze this number in a second, but this number is supposed to be small.

So we want to show that this function here has small rank. So let's rewrite this function in a form explicitly as a sum of products by expanding this function after writing it in a slightly different form. So in F3, in a three-variable-- in characteristic-- so in F3, you have this equation. You can check that it's true for x equal to 0, 1, or 2. So picked that, and plug it in over here.

So we find-- so now x, y, z are in F3 to the n. So we find that, applying this guy here coordinate-wise, you have this product. Great. Now let's pretend we're expanding everything. This is a polynomial in 3n variables, 3n variables. It's degrees is 2n. So if we expand, we get a bunch of monomials.

And the monomials will have the following form. So the x's, which-- whose exponents I call i, the y's, whose exponents I call j, and the z's, whose exponents I call k, where-- so I get a sum of monomials like that, where all of these i, j's, and k's are either 0, 1, or 2. So I get this big sum of monomials, and I want to show that it's possible to write this sum as a small number of functions that can be written as a product, where one of the factors only involves one of x, y, z.

So what we can do is to group them. So group these monomials by the-- so, for example, I'm going to group these monomials by using the following observation. So by pigeonhole, at least one of the exponents of x, or the exponents of y, or the exponents of z, at least one of these guys is, at most, 2n over 3. So I group these monomials by the-- one of x, y, z that has the smallest exponent.

So the contributions to the rank or the slice rank from monomials with the degree of x being, at most, 2n over 3, well, I can write such contributions in the form like that, where this f of x is a

monomial, and the g is a sum of whatever that could come up. This is a sum, but this is a monomial. So the number of such terms-- so the number of such terms is the number of monomials corresponding to choices of i's, the sum to 2n over 3, and individual i's coming from 0, 1, or 2. And that number is precisely M.

So M counts the number of choices of 0, 1, 2's. There are n of them. And the sums of the i's is, at most, 2n over 3. So these are contributions coming from monomials where the degree of x is, at most, 2n over 3. And, similarly, with degree of y being 2n over 3, and also degree of z being, at most, 2n over 3. So. all the monomials can be grouped in one of these three groups, and I count the contribution to the slice rank.

AUDIENCE: Do we have a good idea as to how sharp this bound is?

YUFEI ZHAO: So the question is, do we have a good idea as to how sharp this bound is? That's a really good question. I don't know. Yes. Great. So that finishes the proof of this lemma.

So now we have this lemma. I can compare-- so we have these two lemmas. One of them tells me the rank of the right-hand side, which is A. Let's compare ranks, the slice rank. So the left-hand side, we know it is, at most, this quantity. And the right-hand side is equal to A. So we automatically find this bound. So now we want to know how big this number M is. So there's actually-- this is a fairly standard problem to solve to estimate the growth of this function M. So let me show you how to do it, and this is basically the universal method.

Notice that I can-- if I look at this number here, where if-- so now x is some real number between 0 and 1. Then I claim the following is true. And this is because if you expand the right-hand side and count your monomials-- so you can just keep track of which monomials occur, and there are M of them, where you can lower bound by this quantity here. So this is kind of related to things in probability theory on large deviations, to the Cramér's theorem. But that's what you can do.

So this is true for every value of x, so you pick one that gives you the best bound. So M is, at most, the inf of this quantity here. And to show you any bound, I just have to plug in some value. So if I plug in, for example, x being 0.6, I already get a bound which is the one that I claimed.

And it turns out this step here is not lossy. As in, basically, up to 1 plus little o1 in the exponent, this is the correct bound. And that follows from general results in large deviation theory. And

that finishes the proof. Alternatively, you can also estimate M using Sterling's formula. But this, I think, is cleaner. Great. Any questions? Yes.

AUDIENCE: [INAUDIBLE].

YUFEI ZHAO: Ah, OK. So why is this step true? So if you expand the right-hand side, you see that the right-hand side is upper bounded by all these a, b, c, as in-- same as over here, x to the b plus 2c. And because how many terms-- and, also, there's a binomial coefficient term. So, basically, I'm doing the multinomial expansion, except I toss out everything which is not part of the index. And because b plus 2c is, at most, 2n over 3, I get M times x to the 2n over 3. OK?

AUDIENCE: Yes.

YUFEI ZHAO: Now I want to convey a sense of mystique about this proof. This is a really cool proof. So because you're seeing a lecture, maybe it went by very quickly. But when this proof came out, people were very shocked. They didn't expect that this problem would be tackled, would be solved using a method that is so unexpected.

And this is part of this power of the algebraic method in combinatorics, where we often end up with these short, surprising proofs that take a very long time to find. But they turn out to be very short. So this is very short. This was basically a four-page paper. But when they work, they work beautifully. They work like magic. But it's hard to predict when they work.

And, also, these methods are somewhat fragile. So, unlike the Fourier analytic methods that we saw last time, with that method, it's very analytic. It works in one situation, you can play with it, massage it, make it work in a different situation. Here, we're using something very implicit, very special about these many variables. And if you try to tweak the problem just a little bit, the method seems to break down.

So, in particular, it is open how to extend this method to other settings. It's not even clear what the results should be. So it's open to extend it to, for example, 4-AP. So we do not know if the maximum size of 4-AP-free subset of F5 to the n is less than some constant, 4.99 to the n. So that's very much open. By the way, all of this 3-AP stuff, right now I've only done it in F3, but it works for 3-AP in any finite field.

It also is open to extend it to corners. So you can define a notion of corners. So, previously, we saw corners in integer grid. If I replace integer by some other group, you can define a notion of corners there. So not clear how to extend this method to corners.

And, also, is there some way to extend some ideas from this method to the integers? It completely fails, so this method is not clear at all how you might have it work in a setting where you don't have this high dimensionality. I mean, the result will be different, because, integers, we know that there's no power saving, but maybe you can get some other bounds. Any questions? OK. great. Let's take a break.

So in the first part of today's lecture, I showed you a proof of Roth's theorem. In F3 to the n, that gave you a much better bound than what we did with Fourier. Second part, I want to show you another proof. So yet another proof of Roth in F2 to the n, and this time giving you a much worse bound. But, of course, I do this for a reason. So it will give you the new result. So it will give you some more information about 3-AP's and F3 to the n.

But the more important reason is that in this course I try to make some connections between graph theory on one hand and additive combinatorics on the other hand. And, so far, we've seen some analogies. Well, in the proof of Szemeredi's graph regularity lemma versus the proof-- the Fourier analytic proof of Roth's theorem, there was this common theme of structure versus pseudorandomness. But the actual execution of the proofs are somewhat different. Because, on one hand, in regularity lemma, you have energy increment. You have partitioning and energy increment.

And, on the other hand, with Roth, you have density increment. Or you're not partitioning. You're zooming in. Take a set, find some structure, zoom in, find some structure, zoom in. You'll get density increment. So it's similar, but differently executed.

So, today-- I mean, this second half, I want to show you how to do a different proof of Roth's theorem that is much more closely related to the regularity proof, so that has this energy increment element to it. And I show you this proof because it also gives you a stronger consequence. And, namely, we'll get that there is also not just 3-AP's but 3-AP's with popular difference.

So here's the result that we'll see today. So it's proved by Ben Green. That for every epsilon, there exists some n0 such that every A in subset of F3 to the n with density alpha, there exists some nonzero y such that the number of 3-AP's with common difference y-- so let's think about what's going on here. So if I just give you a set A and ask you how many 3-AP's are there, and compare it to what you get from random, random meaning if A were a random set

of the same density. So question is, can the number of 3-AP's be less than the random count?

And the answer is yes. So, for example, you could have-- in the integers, you can have a barren type construction that has no 3-AP's. So, certainly, that's fewer 3-AP's than random. And you can do similar things here. But what Green's theorem says is that there exists some popular common difference-- so this is a popular common difference-- such that the number of 3-AP's in A with this common difference is at least as much as what you should expect in a random setting, up to a minus epsilon. So this is the theorem.

So let me say the intuition again. It says that, given an arbitrary set A, provided the space dimension is large enough, there exists some popular common difference, where popular means that the number of 3-AP's with that common difference is at least roughly as many as random. In particular, this proves Roth's theorem, because you have at least some 3-AP's. But it tells you more. It tells you there's some common difference that has a lot of 3-AP's, even though, on average, if you just take an average, if you take a random y, this is false. Any questions about the statement?

So Green developed an arithmetic analog of Szemeredi's graph regularity lemma in order to prove this theorem. So starting with Szemeredi's graph regularity lemma, he found a way to import that technique into the arithmetic setting, in F3 to the n. So I want to show you how, roughly, how this is done. And just like in Szemeredi's graph regularity lemma, there were unavoidable bounds which are of power type, the same thing is true in the arithmetic setting.

So Green's proof shows that the theorem is true, with n0 being something like tower in-- a tower of twos. The height of the tower is a polynomial in 1 over epsilon. So just like in regularity lemma for graphs. So this was recently improved in a paper by Fox and Pham just a couple of years ago, where-- and this is the proof that I will show you today-- where you can take n0 to be slightly better but still a tower, but a tower of now height log in 1 over epsilon. So it's from a really, really big tower to slightly less big tower.

But, more importantly, it turns out-- so they also showed that this is tight. You cannot do better. There exists constructions, there exist sets A for which you-- I mean, this theorem is false if you replace the big O by less than some very small constant. So many applications of the regularity lemma. That first proof, maybe using regularity, is difficult. Well, it gives you a very poor bound. But, subsequently, there were other proofs, better proofs, that give you non-tower type bounds.

But this is the first application that we've seen where, it turns out, the regularity lemma gives you the correct bound. So it's really-- you need a tower-type bound. I mean, we know the regularity lemma itself needs tower-type bounds. But it turns out this application also needs tower-type bounds. That's quite interesting. So, here, the use of regularity is really necessary in this quantitative sense.

So let's see the proof. So let me first prove a slightly technical lemma about bounded increments. So this is-- corresponds to the statement that if you have energy increments, you can not increase too many times, but in a slightly different form. So suppose you have numbers alpha and epsilon bigger than 0. And if you have this sequence of a's between 0 and 1, and such that a0 is at least alpha, then there exists some k, at most log base 2 of 1 over epsilon, such that 2 a sub k minus a sub k plus 1 is at least alpha cubed minus epsilon.

So don't worry about this form. We'll see shorty why we want something like that. But the proof itself is very straightforward. Because, otherwise-- so you start with a0. Now, then, if this is not true for k equals to 0, then a1 is at least 2 a0 minus epsilon cubed plus epsilon. So a0 is at least alpha cubed. So if-- otherwise, you have some lower bound on alpha 1, which is at least alpha cubed plus epsilon.

And, likewise, you have some lower bound on alpha 2. You have some lower bound on-- sorry-- alpha 2, and this lower bound is plus 2 epsilon. So you keep iterating. You see the next thing is 4 epsilon, and so on. So if you get to more than this many iterations, you go more than 1. So alpha k is bigger than 1 if k is ceiling of log base 2 of 1 over epsilon. And that will be a contradiction to the hypothesis.

So this is a small variation on this fact that you cannot increment too many times. Each time, you go up by a bit. Whereas, we save a little bit because the number of iterations is now logarithmic. So you double in epsilon each time.

If I give you a function f on F3 to the n, and U is a subspace-- so this notation means subspace. Let me write f sub U to be the function obtained by averaging f on each U coset. So you have some subspace. You partition your space into translates of that subspace, and you replace the value of f on each coset by its average on that coset. So this is similar to what we did with graphons. You're stepping. So you're averaging on each block.

So now let me prove something which is kind of like an arithmetic regularity lemma. And I mean, this statement will be new to you, but it should look similar to some of the statements

we've seen before in the course. And the statement is that, for every epsilon, there exists some m which is a function of epsilon. And, in fact, it will be bounded, in terms of tower of height, at most order logarithmic in 1 over epsilon. Such that for every function f on F3 to the n that are values bounded between 0 and 1, there exists subspaces W and U, where the codimension of W is, at most, m.

So you should think of this as the course partition and the fine partition in the partition regularity lemma. And the codimension is-- corresponds to the number of pieces. So three ways to codimension is the number of cosets. So you have bounded many parts, and have two partitions.

And what I would like is that the number-- so if I-- I want f to be pseudorandom after doing this partitioning, so to speak. And this corresponds to the statement that if I look f minus fW, then the maximum Fourier coefficient is quite small, where quite small means, at most, epsilon over the size of U complement. So size of U perp.

And, also, there is this other condition which tells you that the L3 norms between f sub U and f sub W are related in this way. So we haven't seen this before. In fact, specifically, this inequality is very ad hoc to the application of popular difference in 3-AP's. But we have seen something similar, where this relationship is replaced by something that accounts for the difference between L2 norms.

So if you go back to your notes, when we discussed regularity lemma in a more analytic fashion, we have that. And you should think of this-- when we discussed strong regularity lemma, this definition here, this roughly corresponds to definition that in the fine partition versus the course partition the edge densities are roughly similar, that when you do the further partitioning, you're not changing densities up by very much. So that's the arithmetic regularity lemma.

And once you have the statement-- I mean, I think the hardest part is writing down the statement. Once you have the statement, the proof itself is kind of this follow your nose approach, where you first define the sequence of epsilons. Epsilon 0 is 1, and epsilon sub k plus 1-- and don't worry about this for now. You will see in a second why these numbers are chosen. Let me write R sub k to be the set of r's-- so there will be characters-- such that the Fourier coefficient fr is at least epsilon sub k. So the r's are supposed to identify how we're going to do the partitioning.

Now, the size of this R is bounded. So I claim that the size of R is, at most, 1 over epsilon sub k squared. And that's because there is this parsable identity, which tells you that the L2 sum of the Fourier coefficients is equal to the L2 of the function, which is at most 1. So the number of Fourier coefficients that exceed a certain quantity cannot be too many.

So let U now be the subspace defined by taking the orthogonal complement of these r's. And let's note that if we take alpha sub k to be the-- if we take alpha sub k to be the L3 norm cubed of the function derived from averaging f along the U's, and then looking at the third moment of these densities. So these alphas, we can apply the increment lemma initially to deduce that there exists-- so, in particular, this number here is at least alpha cubed by convexity.

So by the previous lemma, there exists some k, no more than on the order of 1 over-- of log 1 over epsilon, such that 2 alpha sub k minus alpha sub k plus 1 is at least the density of f cubed minus epsilon. So this alpha is supposed to be the density of f. So we find this k. And we have this bound over here from satisfying that inequality.

So this is the density increment argument, the energy increment argument. So we're doing the energy increment argument, basically the same argument as the one that we did when we discussed graph regularity lemma, but now presented in a slightly different form and a different order of logic. But it's the same argument. And what we would like to show is that you also have this pseudorandomness condition about having small Fourier coefficients.

So what's happening here with the Fourier coefficients? Now, how is the Fourier coefficient of an average f related to the original f? So that's something you want to understand up there. And that's something that's not hard to analyze. Because if you have a function U or W-- so either one-- then the Fourier coefficients of this average version is very much related to the original function.

It turns out that if you take an r which is in the orthogonal complement, then the Fourier coefficient doesn't change. And if you are not in the orthogonal complement, then the Fourier coefficient gets zeroed out. So that's something that's not too hard to check, and I urge you to think about it.

So, with that in mind, let's go back to verify this over here. So what we have now is that the-- so this quantity, which measures the largest Fourier coefficient, the difference between f and U sub k plus 1, is, at most-- and what U sub k plus 1 is doing is we're looking at possible large

Fourier coefficients, and we are getting rid of them. So we're zeroing out these large Fourier coefficients, so that the remaining Fourier coefficients are all quite small.

But we chose our R so that if-- so this big R-- so that if your little r is not in big R, then the Fourier coefficient must be small. That's how we chose the big R. So we have this bound over here. And by the definition of the epsilon, we have that bound. And, also, we're combining with this estimate, upper bound estimate on the size of R sub k. So point being we have that. So now take W to be U sub k plus 1, and U to b U sub k, and then we have everything that we want. Question, yes.

AUDIENCE: Why is the codimension of W small?

YUFEI ZHAO: Question is, why is the codimension of W small? So what is the codimension of W? So we want to know that the codimension of W is bounded. So the codimension of W is-- I mean, the codimension of any of these U sub k's is, at most, 3 raised to the number of r's that produce it. And the size of R is bounded. So if we pick m so that it uniformly bounds the size of R, then we have a bound on the codimension. So that's important. So we need to know that the codimension is small. Otherwise, if you don't have the bound on codimension you can just take the zero subspace, and, trivially, everything's true.

We have a regularity lemma, and what comes with a regularity lemma is a counting lemma. So let me write down the counting lemma, and I'll skip the proof. So the counting lemma tells you that if you have f and g both functions on F3 to the n, and U is a subspace F, then-- so let me define-- so the quantity that I'm interested in is-- so I'm interested in understanding 3-AP's where the common difference is in a particular subspace.

So we claim that the 3-AP count of f with common difference restricted to the subspace U-- so it's similar between f and g if f and g are close to each other in Fourier. Well, not quite, because-- so something like this, we saw earlier in the proof of Roth's theorem if we don't restrict the common difference. Turns out, if you restrict the common difference, you lose a little bit. So you lose a factor which is basically the size of the complement of U. So I won't prove that.

But now let me go on to the punch line. So if we start with, again, f function in your space, taking bounds between 0 and 1, and I have subspaces U and W, I claim that the-- if I look at f averaged through W, and I consider 3-AP counts with common difference restricted to U, then this quantity here is lower bounded by this difference between L3 norms. So I claim this is true.

So this is just some inequality. This is some inequality. So of all the things that I did back in high school doing math competitions, I think the one skill which, I think, I find most helpful now is being able to do inequalities. And I thought I would never see these three-variable inequalities again, but when I saw this one-- so Fox and Pham, when they first showed me a somewhat different proof of an approach that didn't go through this specific inequality, I told them, hey, there's this thing I remember from high school. It's called Schur's inequality. And I thought I would never see it again after high school, but apparently it's still useful.

So what Schur's inequality says-- this is one of those three-variable inequalities that you would know if you did math olympiads-- that you have-- so it's an inequality between non-negative-- actually, it's true for real numbers as well, but let's say it's non-negative real numbers. So that's Schur's equality. So if you look at the left-hand side, the left-hand side is-- it can be written as a sum in the following way. I mean, it can be written in the following way. So its expectation over x, y, z that are 3-AP's in the same U coset.

So I'm counting 3-AP's with common difference restricted to U. So common 3-AP's in the same U coset. And I am looking at the product of f sub W evaluated on this 3-AP. So what I would like to do now is apply Schur's inequality to a, b, and c, being these three numbers. The point is you have this a, b, c on the left. And then everything on the right involves only a subset of a, b, c, and they simplify.

So if I do this, then I lower bound this quantity by twice the expectation of x and y in the same coset, same U coset of f sub W of x squared f sub W of y. Maybe I took two other things, but they're all symmetric with respect to each other. And minus the term that corresponds to this sum of cubes. So like that. So this is a consequence of Schur's equality applied with a, b, c like this.

But now you see, over here, I can analyze this expression even further. Because if I let y vary within the same U coset, then, over here, it averages out to U cosets. So U is bigger than W. So what we have is-- so what we have over here is that it is at least twice of f of f-- f of U-- fW squared fU minus the expectation of fW squared-- fW cubed. And I can use convexity on f sub W to get that, which is what we're looking for. So the last step is convexity.

So I'm running through a little bit quick here because we're running out of time, but all of these steps are fairly simple once you observe the first thing you can do is Schur's inequality. And we're almost there. We're almost done. We're almost done. So from that lemma up there, I

claim now that, for every epsilon, there exists some m which is tower log in 1 over epsilon, such that if f is a function on F3 to the n, taking bounds between 0 and 1, then there exists a subspace U of codimension, at most, m such that the 3-AP count, 3-AP density with common difference restricted to U, is at least the random bound minus epsilon.

Why is this true? Well, we put everything together, and choose U and W as in regularity lemma. And, by counting lemma, we have that the 3-AP density of f, so it is at least-- so we're using counting lemma over here-- it is at least the 3-AP density of f sub W of U minus a small error which we can control. So this step is counting. And now we apply that inequality up there.

And finally, we chose our U and W in the regularity lemma so that this difference here is controlled. So it is controlled by the random bound minus epsilon. And that's it. So you change epsilon to 4 epsilon, but we can change it back. And that's it. So we have the statement that you have this subspace of bounded codimension where you have this popular difference result. It doesn't quite guarantee you a single common difference, because, well, you don't really want it to be the case where U is just a single point because I want a nonzero common difference.

But if U is large enough-- if n is large enough at bounded codimension, so, then, the size of U is large enough. So, then, there exists some nonzero common difference. You pick some nonzero element of U. On average, this should work out just fine. So I'll leave that detail to you.

One more thing I want to mention is that all of this machinery involving regularity and Fourier, as with things we've done before, carries over to other settings-- general Abelian groups, and also the integers. And you may ask, well, we have this for 3-AP's. What about longer arithmetic progressions? In the integers, it turns out it is also true, that Green's statement, in the integers if you replace 3-AP by 4-AP. That's a theorem of Green and Tao involving higher-order quadratic analysis-- quadratic Fourier analysis.

However, and rather surprisingly, 4-AP, it's OK. But 5-AP and longer, it is false. The corresponding statement about popular differences for 5-AP in the integers is false. There are counterexamples. So it's really a statement about 3-AP's and 4-AP's, and there's some magic cancellations that happen in 4-AP's that make it true. OK, great. So that's all for today.