Bounds for 3-progressions

Talk by Zander Kelley

Notes by Sanjana Das

April 3, 2025

This is joint work with Raghu Meka.

§1 Introduction

We'll talk about a question due to Erdős–Turán 1936:

> **Question 1.1.** If we have $A \subseteq [N]$ and A does not contain any 3 numbers that are equally spaced (i.e., that form a 3-AP), how large can A be?

To be super clear about the object we're talking about, you can test if three points are in 3-AP with a single linear equation $x + y = 2z$. We're not interested in the case $x = y = z$ ('trivial' 3-APs); we want to ask this question for *nontrivial* 3-APs, i.e., three different points which are evenly spaced.

§1.1 Lower bounds

To start off, let's try to construct some sets that are as large as possible with no 3-APs.

One thing we can try is to greedily include integers starting from 1 and going to N, including one whenever you can without breaking the constraint. It turns out that you can explicitly describe this set using base 3; you'll find that it gives a set of size δN where $\delta \approx N^{-0.42}$ is some polynomial in N. So this construction gets polynomial density.

How about a random construction? Imagine we pick a random set A of size δN. We need it to contain no 3-APs; a specific 3-AP appears with probability roughly δ^3, and there are roughly N^2 possible 3-APs. So we can make this work with $\delta \approx N^{-2/3}$, which is some slightly different polynomial in N.

It turns out that neither of these is close to optimal. There's a nontrivial construction due to Behrend from the 1950s which gets a substantially larger density $\delta \approx 2^{-O(\sqrt{\log N})}$, which is *quasipolynomial* in N. This separates this problem from lots of forbidden substructure problems in combinatorics — it's pretty often in graph theory or coding theory that a random construction is basically the best you can do, but that's not the case here.

The idea is to start with a sphere full of integer points in $\mathbb{Z}^r$, where you pick r carefully so that you get a lot of points (we'll have $r \approx \sqrt{\log N}$). The point is that generalizing to larger dimensions lets you think about an even stronger property — a sphere doesn't even have three points on a line (that's a property that wouldn't make sense in one dimension, but it does in higher dimensions). You start with this set and carefully project it down to the integers (in such a way that you don't create any solutions to $x + y = 2z$). If you do this and balance parameters (setting $r \approx \sqrt{\log N}$), you get this bound.

§1.2 Upper bounds

For the rest of this talk, we'll discuss upper bounds. The first nontrivial upper bound was due to Roth:

Theorem 1.2 (Roth)
We must have $\delta = o(1)$ as $N \to \infty$.

This means if you have any set which is dense enough, it's guaranteed to have a nontrivial 3-AP somewhere.

As some history of the upper bounds for this problem: Roth proved an upper bound of $\delta \approx 1/\log\log N$. Afterwards, there were improvements due to Heath-Brown and Szemerédi, who got $1/(\log N)^c$; Bourgain; Sanders; and then Bloom–Sisask, who got $1/(\log N)^{1+c}$. This was exciting because it beat $1/(\log N)$, which is a barrier for both technical and aesthetic reasons. For example, it answers a question of Erdős asking whether any set as dense as the primes has nontrivial k-APs (this handles the case $k = 3$). There were also technical reasons $1/(\log N)$ was a barrier, so it's interesting they were able to overcome it.

We get the first quasipolynomial type bound for this problem.

Theorem 1.3 (Kelley–Meka 2023)
If $\delta \geq 2^{-(\log N)^{1/12}}$, then A contains a nontrivial 3-AP.

This has the same shape as Behrend's lower bound, just with a different constant for 'quasipolynomial' (we have $\frac{1}{12}$ instead of $\frac{1}{2}$).

For the rest of the talk, we'll switch to a slightly different formulation, where we find not just one 3-AP but *many*. (This isn't just our work; most other works do this too.)

Theorem 1.4 (Kelley–Meka 2023)
If $|A| \geq 2^{-d}N$, then A has at least $2^{-d^{12}}N^2$ solutions to $x + y = 2z$.

So we'll try to find many solutions without worrying about whether they're trivial or not. If there's enough solutions, they can't all be trivial (there's only N trivial solutions), so we actually get some nontrivial ones. (This bound says we get at least a $2^{-d^{12}}$ fraction of all the solutions you could potentially see.)

Another thing is that it'll be convenient to talk about other settings. You can also ask Question 1.1 in other abelian groups, particularly finite ones. In $\mathbb{F}_q^n$, we proved a slightly better bound with 2^{-d^9}. Follow-up works due to Bloom–Sisask showed how to really cleanly do the whole argument in the *general* finite abelian group setting and get 2^{-d^9}. (The case $G = \mathbb{Z}_N$ is roughly equivalent to the original formulation — it's easy to reduce back and forth between them — so this really captures the case we started with, where we work over the integers.)

§2 The high-level analytic approach

Next, we're not going to talk about all the details of the analytic approaches discussed earlier, but we'll talk about some high-level points that are common to all of them. Then we're going to instantiate this abstract approach with some specific details and see how that works out.

§2.1 Structure vs. randomness

The philosophy behind the approach is an attempt at a reduction to one of two extreme easy cases — one where you have complete additive structure, and one where you have no additive structure.

Example 2.1 (No additive structure)

If A is a *random* set of size $\delta |G|$ (in some finite abelian group G), then we'll get $\delta^3 |G|^2$ solutions.

This is plenty of solutions; it's better than we can hope for in general, as shown by Behrend's example.

Example 2.2 (Complete additive structure)

If A is a subgroup of size $\delta |G|$, then there are $\delta^2 |G|^2$ solutions. (This is also true for any coset of a subgroup, because the constraint $x + y = 2z$ is translation-invariant.)

Lastly (this is important because it's what we'll actually reduce to), we can combine these two cases: if we take a large subgroup of G and then take a random subset, we'll also see lots of 3-APs there.

In some sense, this can't work out — we know these cases are not the actual extreme ones (we can only hope to prove a quasipolynomial bound, so these are too good in some sense). But we'll try to reduce to these cases in a 'loose' sense.

One other thing we want to do in the analytic approach is to identify some nice analytic condition on a set that guarantees it has as many solutions as a random set would. Then we're going to try to find some $A' \subseteq A$ which has this property that makes it easy to estimate the number of solutions in A', and we're just going to zoom in on A' and count solutions there, ignoring everything else. And of course we want A' to be as large as we can manage.

How are we going to locate this nice subset A'? A natural thing to try is to take $A' = A \cap V$ where V has lots of additive structure. What this means depends on what ambient group you're in (the general plan looks mostly the same in any group, but this part splits off depending on what specific group you have) — if you're in a group with lots of subgroups (e.g., $\mathbb{F}_p^n$) then you can take V to be a subgroup. But in $\mathbb{Z}_N$ you have no subgroups (e.g., if N is a big prime), so you have to do someting else. There are things like *Bohr sets* which can play the role of an 'approximate subgroup,' as well as *generalized arithmetic progressions*.

Example 2.3

An interval $I = [-m, m]$ is an example of both a low-rank Bohr set and a low-rank GAP. For some justification that it's reasonable to think of it as an 'approximate subgroup' — for a generic set S, you'd expect $|S + S| \approx |S|^2$. For an interval I, you see more than I sums, but much fewer than the general case — you have $|I + I| \approx 2|I|$. So I is 'approximately closed under addition' in some quantitative way.

Lots of previous work focused on taking Roth's original argument for $1/(\log N)$ in the simple case $A \subseteq \mathbb{F}_q^n$, and refining it by getting good ways of understanding approximate subgroups so that you get almost as good of a bound for the harder case $A \subseteq [N]$. The exception is that Bateman–Katz 2012 work with $A \subseteq \mathbb{F}_q^n$ and do something fancy that improves Roth's original argument from $1/(\log N)$ to $1/(\log N)^{1+c}$. It takes a lot of work to move this to $[N]$, and Bloom–Sisask managed to do so to get past the $1/(\log N)$ barrier.

§2.2 Density increments

To ensure that A' has the right number of solutions, we'll need to pick some notion of *pseudorandomness* that guarantees this.

Then how are we going to find A'? We'll try to zoom in on it step by step. We start with our entire set A and ambient space G, so we initialize $A_0 = A$ and $V_0 = G$. Then at the ith step, if A_i is not pseudorandom, we want to zoom in on some further subgroup or coset — so we find some V_{i+1} and set

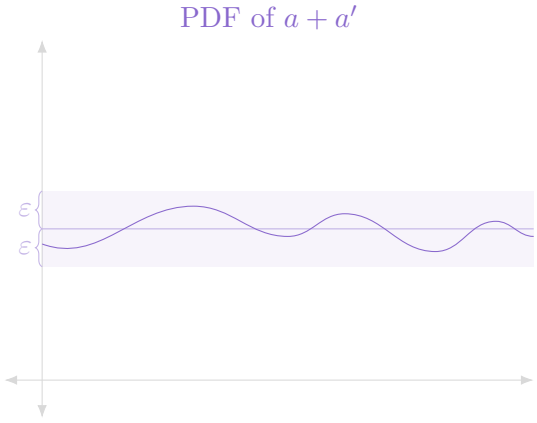
$$A_{i+1} = A_i \cap V_{i+1}.$$

And we want to do this so that our density increases, by some constant multiplicative factor. We'll stop when A_t is 'pseudorandom' with respect to V_t, where this notion of pseudorandomness needs to be designed so that this means A_t has the 'expected' number of 3-APs based on a random subset of its container set.

§3 A notion of pseudorandomness

The next thing we'll talk about is instantiating a notion of pseudorandomness. Lots of prior works (inspired by Roth) used a notion that looked at the largest Fourier coefficient; we'll use a different notion, which is describable in physical space.

Imagine we take A and draw $a, a' \in A$ uniformly at random. Our test for when to call A pseudorandom will essentially be that $a + a'$ should be near-uniform over G — so we're looking at the probability density function of the random variable $a + a'$, and we want it to be within a constant tolerance ε of the uniform density function over G. (If you're a number theorist, we're looking at convolutions of $\mathbf{1}_A$ with itself.)



First let's make sure that this is sufficient for our purposes. (It's going to be too much to ask for, so we'll weaken it soon, but let's first talk about why it would be good enough.)

If we have this kind of situation, where the random variable $a + a'$ is near-uniform over the whole group G, then we can control the number of solutions to $a + a' = c$ for *any* set C (where we draw $c \in C$) — it'll be

$$(1 \pm \varepsilon) \cdot \frac{|A|^2 |C|}{|G|}.$$

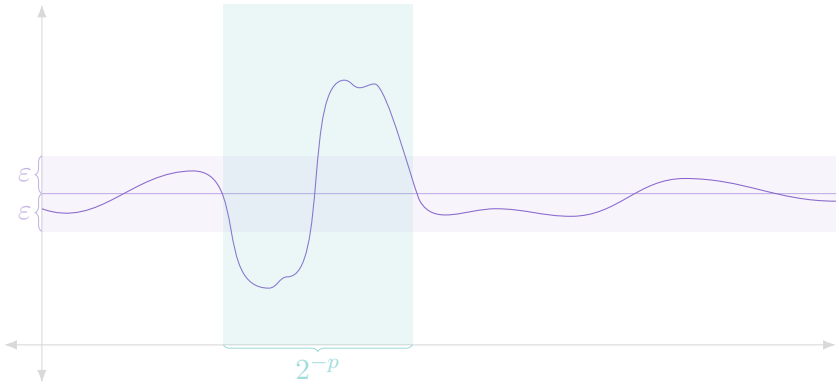
Then in particular, we can apply this to $C = 2 \cdot A$.

However, the version of near-uniformity we just stated is too much to ask for, so we need to allow some error — we can't get this control to happen *everywhere*. So we'll allow deviations outside the $\pm\varepsilon$ band around the uniform distribution *somewhere*; and we'd like the fraction of the group G where these deviations occur to be exponentially small. So we'll have some parameter p floating around, and we want the set of bad points S to satisfy

$$\frac{|S|}{|G|} \leq 2^{-p}.$$

For example, if we set $p = d + 1$ (where 2^{-d} is the density of A), this would be enough to get within a constant factor of the right answer — the number of solutions to $a + a' = 2a''$ would be at least

$$\frac{1}{4} \cdot \frac{|A|^3}{|G|}.$$



Definition 3.1. We say that A is (ε, p)-near-uniform if letting D be the density function of $a + a'$, we have that D is within $1 \pm \varepsilon$ of the uniform distribution outside a set S of (fractional) size 2^{-p}.

You can also quantify this by

$$\|D - 1\|_p \leq \varepsilon;$$

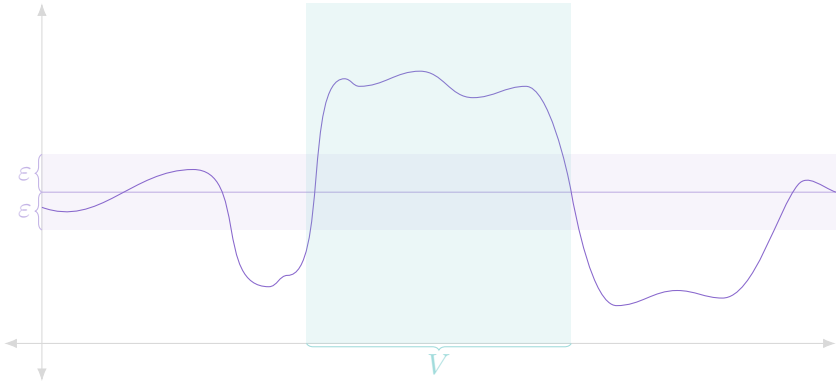
the two statements are roughly equivalent, and we'll jump back and forth between them.

Our main lemma, stated for general abelian groups:

Lemma 3.2

Either the PDF of $a + a'$ is near-uniform, or we can find a density increment onto some approximate subgroup V which is not too small.

The above picture illustrates the first case (where D is near-uniform). And as a picture of the second case, where we find a large structured set V and get a substantial density increment:



This picture actually shows us getting a density increment on D, not A. But a density increment on D implies a density increment on A by an averaging argument (we can write $\langle V, A * A\rangle = \mathbb{E}_{a'}\langle V, a' + A\rangle$, so if the left-hand side is large then $\langle V, a' + A\rangle$ is large for some a', which means $\langle V - a', A\rangle$ is large). So this is the way to think about it (and we're actually going to get a density increment on D).

Here's the main lemma restated for $\mathbb{F}_q^n$ with specific parameters:

Lemma 3.3

In $G = \mathbb{F}_q^n$, either D is near-uniform, or we can find an affine subspace V with codimension at most d^4p^4 such that

$$\frac{|A \cap V|}{|V|} \geq (1+\varepsilon) \cdot \frac{|A|}{|G|}.$$

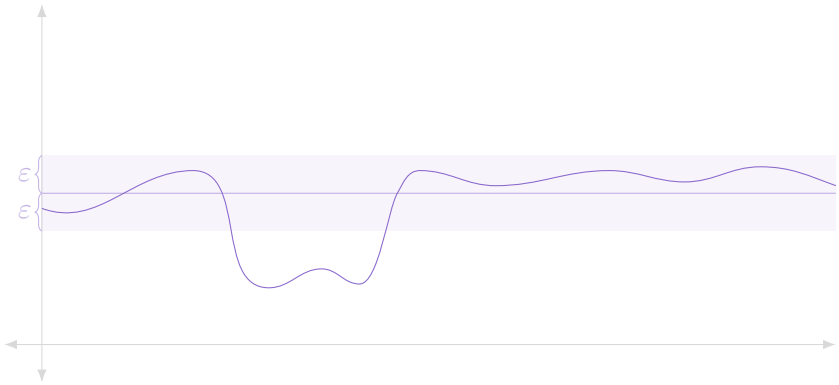
The place where d^9 comes from in the final bound is that we set $p \approx d$, and that gives us d^8 loss in dimension (equivalently, 2^{d^8} loss in size) at every step; and we need to run for about d steps (since we start with density 2^{-d} and multiply by a constant at each step).

For the rest of the talk, we'll give an overview of the proof of this key lemma — this dichotomy that you can't have some third option (where you have some density function D which is not near-uniform, but it's also not structured enough to find a large subset V with a density increment).

So we're going to assume that D is not near-uniform, i.e., that $\|D-1\|_p \geq \varepsilon$. And we want to find a large set V where $\mathbb{E}_V[\mathbf{1}_A] \geq (1+\varepsilon)\mathbb{E}[\mathbf{1}_A]$. As in the above picture, we're actually going to prove a stronger statement, where we get a density increment on D itself, not just A.

§4 Step 1: Spectral positivity

There's a really bad thing that could happen at the beginning which would kill all your dreams. Even forgetting about the structure, what if it's the case that D only has substantial *downwards* deviations?



We've normalized D to be a density function, so if it has downwards deviations it must have *some* upwards deviations, but those might be exponentially small (which isn't good enough for us). So if we were using 1-norms instead this wouldn't be an issue; but with p-norms it's a bad possibility that we need to rule out.

The first idea is a strange idea showing that this bad case is just not possible. We actually want to trade the density function D for a different function F, the PDF of $a - a'$. For the same reason we talked about with D, it's good enough to find a density increment for F instead (this would give us a density increment for A); so that's what we'll actually look for.

The first point is that we have

$$\|D-1\|_p \leq \|F-1\|_p,$$

so the deviation only goes up when we switch from D to F; this can be justified with some Fourier analysis.

The second point is that we have

$$\|(F-1)_-\|_p \leq \|(F-1)_+\|_p,$$

i.e., the deviation downwards from 1 can't be larger (as measured by the p-norm) than the deviation upwards from 1. So we can't have the above bad picture for F.

Why is this? This can be reformulated as saying that F has nonnegative odd central moments — we want to say that for any odd p, we have

$$\mathbb{E}_x(F(x)-1)^p \geq 0.$$

And the reason for this is that F is spectrally positive — it only has nonnegative Fourier coefficients. Then you can just Fourier-expand $F(x)$ inside the parentheses; subtracting 1 corresponds to taking away the constant term, so you're still sitting there with some function that only has nonnegative Fourier coefficients. Then you take that function to the pth power and take some expectation; and you can check that you only have nonnegative things sitting around. So that proves $\mathbb{E}_x(F(x)-1)^p \geq 0$, which is exactly what we need.

This means we're only in the good case — if things are going wrong (meaning that F is not near-uniform), then they must be going wrong in the positive direction, which is necessary for this density increment approach.

§5 Step 2: Sifting

So that's the first big main idea that we needed to take care of. Now we know there are actually upwards deviations to look for, and we're hoping to find upward deviations on a *structured* set.

§5.1 A hard example: a planted subspace

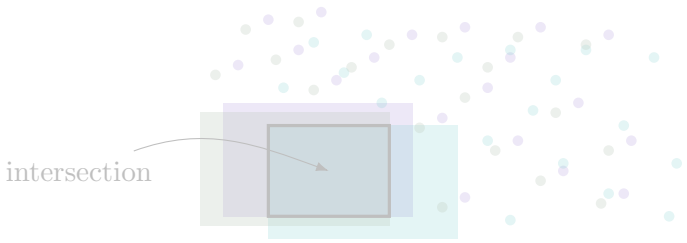
As some motivation, we'll talk about a potential 'hard example.' We considered two extreme cases — a random set and one with complete additive structure — so you could imagine the worst case is a mixture of the two. Imagine that A is made of a large random set R and a medium-sized subspace V.

Note that this is an example that you really have to deal with — it might be the case that $\|F\|_p \geq 1+\varepsilon$ because of V.

> **Question 5.1.** If we knew that A looked like this but didn't know where the subspace V was, how could we try to find it?

Thinking about this is what leads to the next idea. Here's the idea — we can take A and translate it randomly in a couple of directions, and take the intersection. So for example, we could look at

$$A \cap (A+s_1) \cap (A+s_2).$$



The random part of A (all the 'dust') gets lost when we take this intersection. And we haven't found the *whole* subspace V, but we've found a reasonable-sized part of it, which we're happy with as well.

§5.2 Sifting

We can actually take this procedure and run it for general sets A. Now it's unclear what we'll get. We're hoping that when we do this we find something that's useful — specifically, something that has substantial additive structure.

So suppose we have $F = \mathsf{PDF}(a - a')$, and we assume that $\|F\|_p \geq 1 + \varepsilon$. (Once you know the deviation upwards from 1 is bigger than the deviation downwards, you can make this translation and quantify the deviations in this way, up to slightly changing the value of p.)

What sifting does is it finds a set B of the form $B = \bigcap_i (A + s_i)$ where $i = 1, \ldots, p$, and where B is witnessing a density increment of F in some sense — we'll have $|B| \geq 2^{-dp} |A|$ (roughly), and

$$\mathbb{E}_{b,b' \in B}[F(b - b')] \geq (1 + \varepsilon)\mathbb{E}[F].$$

How does sifting fit into the big picture? We wanted to get a density increment on a structured set V. Sifting gives us a set B where the convolution of B with itself locates a density increment — we get a density increment on some distribution, namely $\mathsf{PDF}(b - b')$, which has more additive structure than nothing.

§6 Step 3: Bootstrapping

The last step, which was already known in the literature, is how to bootstrap this. Once you have a witness which is a convolution of two things (here $\mathsf{PDF}(b - b')$), somehow that's enough to get a witness which is a convolution of *three* things, then four things, and so on. And once you have a convolution of p things, you can approximate it by a subspace.

This was known before. For example, there's versions of the 3-AP problem where instead of a 3-variable equation like $x + y = 2z$ you look at a 4-variable equation. And there people already knew how to get quasipolynomial bounds, using this Croot–Sisask lemma in the same way we do here. You can imagine that once we've gotten to this point we've transformed the 3-variable problem into a 4-variable one, and this can get us the rest of the way.

§7 Summary

The three steps of the argument are spectral positivity, sifting, and Croot–Sisask. It's interesting to track the witnesses of deviation from uniformity as we go down the argument.

- At the start, by assumption we have some witness $|C| \geq 2^{-p} |G|$ with no structure whatsoever; all that we know is that it's large. And we have $\mathbb{E}_C[|F - 1|] \geq \varepsilon$.
- The first thing we do is find some other set S, *also* with no structure, but which specifically witnesses deviations *upwards* from uniform — we have $\mathbb{E}_S[F] \geq 1 + \varepsilon/2$.
- From this we find another set B where we trade size for additive structure — we get a density increment of the form $\mathbb{E}_{b,b'}[F(b - b')] \geq 1 + \varepsilon/4$ with $|B| \geq 2^{-O(pd)} |G|$. (So instead of having a density increment on an arbitrary set, now we have one on the form $\mathsf{PDF}(b - b')$.)
- And then we trade even more size for even more additive structure to get all the way to a subspace — we find V with $\mathbb{E}_V[F] \geq 1 + \varepsilon/8$ and $|V| \geq 2^{-O(p^4 d^4)} |G|$ — which gives the density increment we needed.