

1. INTRODUCTION

This document is my personal attempt to understand properly Bourgain’s first proof of Roth’s theorem. As I write these first sentences, I do so knowing that there is a good chance that I will finish the document before I understand the proof properly – where by “properly” I mean well enough to be able to think seriously about how the argument might be modified. I shall assume that the reader (and it may well be that “the reader” means exactly what it says) is familiar with the Roth/Meshulam argument in \mathbb{F}_3^n and the argument of Roth that gives a $1/\log \log N$ bound in $[N]$.

2. PRELIMINARY OVERVIEW

At one level, the basic idea of Bourgain’s argument is very simple. The weakness in Roth’s argument is that when one has obtained bias with respect to a trigonometric function, one then passes to a subprogression where there is also bias, in order to be back in the situation one was in when the iteration started. The trouble with this is that the length of the subprogression is roughly the square root of the length of the original progression, which is quite expensive when one iterates.

This problem is far less severe in \mathbb{F}_3^n , since then bias with respect to a character implies bias on a codimension-one affine subspace, so that all one loses at each iteration is one dimension, or in terms of cardinality just a factor of 3.

Bourgain’s idea, or rather his preliminary idea, was the natural one of avoiding Roth’s passage to a subprogression, and instead taking on board the structure on which one has bias. Let me make that remark slightly clearer.

At the first iteration, one shows that if a set $A \subset \mathbb{Z}_N$ of density δ_0 contains no 3APs, then there must exist r such that $\mathbb{E}_x A(x)\omega^{-rx} \geq c\delta_0^2$. From this one can deduce by a straightforward averaging argument that there is a mod- N arithmetic progression P_1 of length $N/2$ such that $|A \cap P_1| \geq (\delta + c'\delta^2)|P_1|$. At this stage, Roth would say that we must pass to a “genuine” arithmetic progression, which he does by partitioning P_1 into not too many subprogressions that do not wrap around and applying another averaging argument. Bourgain, by contrast, just sticks with P_1 itself.

We already have a problem here, which is that if we eventually manage to find a triple $(x, x + d, x + 2d)$ in P_1 , we do not appear to have any guarantee that it will be a genuine progression. This can be dealt with in many ways, of which the simplest is probably just to assume that A is contained in the first half of \mathbb{Z}_N . Of course, this instantly implies that we have a huge amount of bias on the progression $\{1, 2, \dots, N/2\}$, so P_1 may well

be precisely this progression. But it could be some other mod- N progression on which A happens to be biased, so all we really need to say is that we are free to assume that any mod- N progression of length 3 that we happen to find in A will still be a progression if we regard A as a subset of $\{1, 2, \dots, N\}$.

What happens at the next iteration? The rough idea is this. Either $A \cap P_1$ looks very much like a random subset of P_1 of the same density, or it doesn't. In the first case, if $|A \cap P_1| = \delta_1 |P_1|$, then the number of 3APs in A will be approximately δ_1^3 times the number of 3APs in P_1 . Since P_1 is a mod- N progression, it contains many 3APs, and therefore so does A . In the second case, we find that the Fourier behaviour of A differs significantly from the Fourier behaviour of $\delta_1 P_1$. This gives us trigonometric bias as before. More precisely, it provides us with a mod- N progression P_2 such that $P_1 \cap P_2$ is reasonably large and $|A \cap P_1 \cap P_2| / |P_1 \cap P_2|$ is $\delta_2 \geq \delta_1 + c' \delta_1^2$.

If we continue this line of argument, we find that we are passing not to subprogressions but to intersections of long mod- N progressions, which are more or less the same as translates of Bohr sets. We can think of these as playing the role in \mathbb{Z}_N that subspaces play in \mathbb{F}_3^n . (Indeed, one can easily define subspaces and Bohr sets in a unified way.)

So now we appear to be in a very good situation: we have a proof that works very nicely in \mathbb{F}_3^n and at least the beginnings of a dictionary for translating the proof into a \mathbb{Z}_N context. In particular, subspaces translate into Bohr sets.

At this point, it would be reasonable to expect that the $C/n = C'/\log(3^n)$ bound that comes out of the Roth/Meshulam argument would carry over to a $C/\log N$ bound in Bourgain's argument. But it doesn't. It gives a bound of $1/\sqrt{\log N}$ up to $\log \log N$ factors. My two main aims in writing this are to understand properly why things don't work quite as well in \mathbb{Z}_N , and to think about whether there might be a modification of Bourgain's argument that would give a $1/\log N$ bound, again up to $\log \log N$ factors.

That might seem a strange thing to write, since the bound has since been improved to $(\log N)^{-2/3}$, then $(\log N)^{-3/4}$, then $(\log N)^{-1}$ (all up to $\log \log N$ factors) by Bourgain, Sanders and Sanders, respectively. In a moment I will explain why what I would like to do (but which may well be impossible) differs from what they have done.

Before I do that, let me explain why Bohr sets are not as easy to work with as subspaces. The reason is simple: an intersection of k mod- N arithmetic progressions P_1, \dots, P_k resembles a k -dimensional lattice convex body (that is, the intersection of a k -dimensional lattice with a convex body in \mathbb{R}^k). If you take two points $x, x+d$ in a typical k -dimensional lattice convex body B , then the probability that $x+2d$ belongs to B is exponentially small in k .

This is for concentration-of-measure reasons: the measure of B is mostly on the boundary, and if x and $x + d$ are near the boundary then it is hard for $x + 2d$ to belong to B . By contrast, if x and $x + d$ belong to an affine subspace of \mathbb{F}_3^n , then $x + 2d$ is guaranteed to belong to that subspace as well.

Bourgain’s approach to this difficulty is to take a much smaller Bohr set B' , which we can think of as “close to the centre” of B . If we then take a random pair $(x, x + d)$ such that $x \in B$ and $x + d \in B'$, it is much more likely that $x + 2d$ will be in B (since it is close to the point “opposite” x). Indeed, if you choose the numbers correctly, you can make the probability close to 1. This allows one to write down analytic expressions similar to the ones that one writes down in \mathbb{F}_3^n , but of the three points in the 3AP one of them lives in B' instead of B .

Having done this, he obtains a density increase on a Bohr set of dimension greater by 1. So far, this is just like what happens in \mathbb{F}_3^n . However, the difference is that the *width* of the Bohr set is comparable to that of B' rather than to that of B . So he loses width as well as gaining dimension, and this has a significant impact on his final bound. The improvements to $(\log N)^{-2/3}$ and $(\log N)^{-3/4}$ are obtained by reducing the number of iterations: typically, this has little effect on the dimension but enables one to make a width sacrifice fewer times. As for Sanders’s $1/\log N$ bound, that is obtained by means of a substantially different argument.

So the question I am interested in becomes more precise: is there a way of iterating using Bohr sets, and *not* making a big width sacrifice at each iteration? That is, is there an argument in \mathbb{Z}_N that resembles much more closely the Roth/Meshulam argument? A priori the answer would seem to be no – after all, surely if it existed then it would have been discovered by now. I find that argument pretty plausible, but not a good enough reason not to think about the question, especially as Tom Sanders didn’t completely dismiss the idea.

The motivation for asking this question is obvious. A serious obstacle to carrying over the Bateman-Katz approach to \mathbb{Z}_N is that it takes as its starting point the assumption that we have a set $A \subset \mathbb{F}_3^n$ for which all the steps of the Roth/Meshulam argument are close to best possible. If the natural analogue of the Roth/Meshulam argument is Bourgain’s argument, then by importing the Bateman-Katz ideas, we can expect to obtain an argument that gives a bound of $(\log N)^{-1/2-\epsilon}$ for some very small ϵ . Perhaps if we use some of the known techniques for reducing the number of iterations, we can improve this to something like $(\log N)^{-3/4-\epsilon}$ (Tom Sanders thinks that this is a realistic aim), but obtaining a bound of

$(\log N)^{-1-\epsilon}$ looks much harder, since Sanders's $(\log N)^{-1+o(1)}$ bound uses physical-space arguments in a crucial way rather than analysing the structure of the spectrum. If one could find a modification of Bourgain's original argument that gave a bound more like $(\log N)^{-1}$, then this picture would change, and importing the Bateman-Katz ideas would suddenly look very feasible indeed.

3. WHAT COULD CONCEIVABLY BE VARIED IN BOURGAIN'S ARGUMENT?

If one is looking for a modification of an existing argument, a first step is to try to understand which features of the argument are essential and which are arbitrary choices made by the author of the argument. Unfortunately, when writing up a proof one is not obliged, or even particularly encouraged, to specify this, so if one wishes to decide whether a feature of an argument is essential, there is often nothing for it but to vary that feature and see whether one gets stuck later. Given that this is not always easy, it would be much more efficient if it were standard practice to include a few sentences of the type, "Another thing one might think of trying to do here is X, but if you do that then you run into difficulties because Y."

3.1. Possible changes. I want to focus on two changes that one might consider making to Bourgain's argument. The first is that if you have a large Bohr set B and a smaller Bohr set B' inside it, then there is a different way that one might naturally imagine using the fact that B is approximately closed under addition of elements of B' . Recall that Bourgain insists that the middle terms of his APs lie in B' , but he might instead have insisted that the *common differences* of the APs lie in B' . Indeed, that is the first thing that would have occurred to me. The fact that it is the first thing that would have occurred to me suggests to me that Bourgain thought of it and decided against it for some reason. But I would like to know what that reason was.

The second change one might consider making is this. At each iteration, Bourgain has a small Bohr set B' contained in a large Bohr set B . But for the argument, it seems to be much more important that B' is a Bohr set than that B is a Bohr set. All one really needs of B (I think – this is an unchecked statement) is that it is approximately closed under addition of elements of B' , and there are many examples of such sets. Or one could generalize from sets to non-negative functions. For instance, perhaps a Riesz product would be more convenient to deal with than the larger Bohr set.

3.2. Do we have to pass to a *small* Bohr set after iterating? Let me play devil's advocate for a moment and try to give a heuristic argument in favour of the proposition

that if we use a small Bohr set B' to obtain approximate closure under addition, then at the end of the iteration the best that we can hope for is a density increase on a Bohr set B'' that is contained in a translate of B' , which then has to be the “large set” at the next iteration. If that is correct, then the difficulty I am talking about is a fundamental difficulty of this whole approach.

The heuristic argument is something like this. Let B' be a Bohr set and let B be some other set (or weighted set) that is approximately closed under addition of elements of B' . I don't insist that B is itself a Bohr set, but it could be. Let me begin by imagining that it *is* possible to produce a variant of Bourgain's argument where it is the common difference that belongs to B' rather than the central term. (If we went for the central term then we would want B to be symmetric about the centre of B .)

Now suppose that we have a set $A \subset B$ with relative density α and the wrong number of 3APs with common difference in B' . That is, the expression

$$\mathbb{E}_{x \in B} \mathbb{E}_{d \in B'} 1_A(x) 1_A(x+d) 1_A(x+2d)$$

is not approximately equal to α^3 . What Fourier information can we hope to deduce from that?

Well, one way of producing a set A with the wrong local 3AP count (by which I mean the count of 3APs with “small” common differences) is to take a random looking function that is approximately constant on translates of B' but not approximately constant globally. Then there will be some translates of B' where the density is bigger than α and some where it is smaller. The average density will be α and on each translate the contribution to the 3AP count will be roughly proportional to the cube of the density (this is the local randomness hypothesis). Since the expectation of the cube is substantially greater than the cube of the expectation when a function isn't roughly constant, we have the wrong 3AP count. However, since the function is fairly random when viewed from a distance, it would appear that any density increases would have to take place inside translates of B' and would not be visible from the more global perspective of the large set B .

That argument was of course not precise. Now let me try to criticize it. (Secretly, or rather not so secretly, I want it to fail.)

And to my secret, or not so secret, pleasure it does fail, or at least it does not obviously succeed. The reason is the very phenomenon that required us to produce a small Bohr set in the first place: measure concentration. How might we create a function that is roughly constant on translates of B' ? The sort of mental picture one might have is this: we tile B

with translates of B' and make the function constant in each of those translates. But that doesn't work at all: the measure of each translate chosen is concentrated on the boundary, so a typical progression $(x, x + d, x + 2d)$ with $x \in B$ and $d \in B'$ will jump from translate to translate and we will not get a high 3AP count after all.

OK, you might say, we should make the function a bit more continuous. But how much more continuous does it need to be? (By "the function" I mean something like a function f that determines the probability that a point belongs to A . Or one could simply generalize and talk about functions instead of sets.) We want it to be sufficiently continuous that if you add some $d \in B'$ then you (normally) don't change the value by very much. But that is more or less exactly the reverse of the condition we imposed when choosing B' in the first place! That is, it seems that in order to produce a function that is sufficiently smooth to give an increased 3AP count, it has to be *globally* smooth rather than just *locally* smooth.

An amusing feature of this argument is that it breaks down when the problem of high-dimensionality does not occur. Then we no longer have measure concentration, but we also no longer have the difficulties that caused us to take a much smaller Bohr set. Thus, when I talked about reversing the condition imposed on B' , that is exactly what I mean: in \mathbb{F}_3^n , for instance, the measure-concentration argument breaks down completely, but we don't have to take a small Bohr set. Could we prove that there is always an "exact balance" of this kind?

Let me try to formulate a semi-precise statement that would be very useful and that does not seem to me to be obviously false (though it could be non-obviously false, or obviously false for a reason that I have not yet thought of). The statement is this: if B is a regular Bohr set, or something of a similar kind, then there exists a Bohr set B' such that for any set $A \subset B$ of density α , either

$$\mathbb{E}_{x \in B} \mathbb{E}_{d \in B'} 1_A(x) 1_A(x + d) 1_A(x + 2d) \approx \alpha^3$$

or there is a translate of a Bohr set B_1 , of dimension one higher than that of B and of comparable cardinality, such that $|A \cap B_1| \geq (\alpha + c\alpha^2)|B_1|$.

What I mean by that is that we would start with B and would intersect it with a mod- N arithmetic progression P to form B_1 , and that we would have some estimate such as that $|B_1| \geq |B|/100$, though even an estimate like $|B_1| \geq \alpha^2|B|$ would be good enough (in that it would cost just a log log factor).

That formulation is just a guess at this stage. A more general hope is that *something* along the above lines is true. And I think the very informal arguments of this section are,

at least for now, convincing enough that I must continue to investigate this possibility. (I would say “we must” if it were not for the fact that I still half expect this line of enquiry to fizzle out in the light of some simple observation that I have not yet made.)

4. SOME SIMPLE CALCULATIONS

Let λ be a non-negative function defined on \mathbb{Z}_N and suppose that $\mathbb{E}_x \lambda(x) = 1$. The main example I have in mind is the characteristic measure of the “large” Bohr set, but I am using the letter λ to avoid making any assumptions about the function until they are forced on me. Since I’m doing that, let me write B instead of B' for the small Bohr set. Thus, λ is a non-negative function that averages 1, and λ is approximately closed under addition of elements of B , where the meaning of “approximately closed” is something else that we are free to decide later. (For now we can think of it as saying that $\lambda * \mu_B - \lambda$ is small in some as yet unspecified norm.)

I would sometimes like to think of λ as a measure. So if A is a set, then $\lambda(A)$ means $\mathbb{E}_x \lambda(x) 1_A(x)$, and if f is a function defined on \mathbb{Z}_N , then $\int f(x) d\lambda(x)$ means $\mathbb{E}_x \lambda(x) f(x)$. In some sense, λ is playing the role of what some people like to call the “ambient group”. Thus, $\lambda(A)$ is “the density of A inside λ ”.

Let us now try to mimic the usual proof that if A contains no 3APs then there is a density increment on a subspace. The slight twist is that we shall assume that our common differences belong to B . Thus, the expression we write down for the 3AP count is

$$\int \mathbb{E}_{d \in B} 1_A(x) 1_A(x+d) 1_A(x+2d) d\lambda(x),$$

which equals (for those who don’t like this notation)

$$\mathbb{E}_x \mathbb{E}_{d \in B} \lambda(x) 1_A(x) 1_A(x+d) 1_A(x+2d).$$

Remark. Why do the above expressions not also involve $\lambda(x+d)$ and $\lambda(x+2d)$? The reason is that we are measuring the probability that the 3AP $(x, x+d, x+2d)$ is entirely within A if you choose x randomly according to the measure λ and d as a random element of B . It may help to observe that most of the time $\lambda(x) \approx \lambda(x+d) \approx \lambda(x+2d)$.

For the purposes of taking the Fourier transform, the second expression is slightly easier (to me at least). Let us rewrite it as

$$\mathbb{E}_x \mathbb{E}_d \lambda(x) 1_A(x) 1_A(x+d) 1_A(x+2d) \mu_B(d)$$

where μ_B is the characteristic measure of B . Writing ω for $\exp(2\pi i/N)$, we can rewrite this as

$$\mathbb{E}_x \mathbb{E}_y \mathbb{E}_z \mathbb{E}_u \mathbb{E}_d \lambda(u) 1_A(x) 1_A(y) 1_A(z) \mu_B(d) \sum_r \omega^{r(x-u)} \sum_s \omega^{s(x-2y+z)} \sum_t \omega^{t(y-x-d)},$$

which equals

$$\sum_r \sum_s \sum_t \hat{\lambda}(-r) \hat{1}_A(r+s-t) \hat{1}_A(-2s+t) \hat{1}_A(s) \hat{\mu}_B(-t)$$

We can make this a bit more transparent by changing variables, to obtain the expression

$$\sum_r \sum_s \sum_t \hat{1}_A(r) \hat{1}_A(s) \hat{1}_A(t) \hat{\lambda}(-r-s-t) \hat{\mu}_B(-s-2t).$$

To see that this makes some kind of sense, imagine that λ is the characteristic function of \mathbb{Z}_N and $B = \mathbb{Z}_N$. Then both $\hat{\lambda}$ and $\hat{\mu}_B$ are delta functions at zero, so on r, s and t we end up imposing the conditions that $r+s+t=0$ and $s=-2t$. This implies that $s=-2r$ as well, so $(r, s, t) = (r, -2r, r)$ and we recover the usual Fourier expression for the 3AP count.

One other important remark is that the approximate B -translation invariance of λ implies (under suitable interpretations of what the invariance actually is) that $\hat{\lambda}$ is essentially supported in a set where $\hat{\mu}_B$ is approximately equal to 1. (To sketch the argument, if λ is approximately unchanged when you translate by d , then $\hat{\lambda}(r)$ is approximately equal to $(1 - \omega^{rd})\hat{\lambda}(r)$. If that is true for every $d \in B$, then $\hat{\mu}_B(r) = \mathbb{E}_{d \in B} \omega^{rd}$ is approximately 1.)

I'm finding this expression a bit ugly, so I'm going to make a little change. Let us write ν for $\lambda^{1/3}$ and let $f(x) = \nu(x)1_A(x)$. Then the approximate invariance of λ implies that the initial expression we wrote down for the 3AP count is approximately equal to

$$\mathbb{E}_x \mathbb{E}_{d \in B} f(x) f(x+d) f(x+2d).$$

This time if we do some similar calculations to move to the Fourier side we obtain the much nicer expression

$$\sum_{r+s+t=0} \hat{f}(r) \hat{f}(s) \hat{f}(t) \hat{\mu}_B(r-t).$$

Note that if $r+s+t=0$ and $r=t$ then $(r, s, t) = (r, -2r, r)$, so we are replacing the condition $r=t$ by the condition that $r-t$ is small (in the sense that it belongs to B).

Now we are intending to compare this last expression with what we would have got if f had been a constant multiple of ν . (This corresponds to the situation where A is a quasirandom "subset of λ ".) So just to see where we might be heading let us write out the

corresponding expression in the case where $f = \nu$. We get

$$\sum_{r+s+t=0} \hat{\nu}(r)\hat{\nu}(s)\hat{\nu}(t)\hat{\mu}_B(r-t).$$

To see that this makes sense, recall that $\hat{\nu}$ is essentially supported where $\hat{\mu}_B$ is close to 1. Therefore, the above sum is approximately equal to $\sum_{r+s+t=0} \hat{\nu}(r)\hat{\nu}(s)\hat{\nu}(t)$, which is equal to $\hat{\nu} * \hat{\nu} * \hat{\nu}(0)$, which is equal to $\sum_x \nu(x)^3 = \sum_x \lambda(x)$. Since $\nu(x) \approx \nu(x+d) \approx \nu(x+2d)$ for most x and $d \in B$ (if we choose the right assumptions), this is just what we would expect.

We still haven't really touched the main question, which is what sort of trigonometric bias we can hope to obtain. Our starting assumptions are that $\int f(x) = \nu(x)1_A(x)$ for some set A , that $\lambda(A) = \alpha$, and that

$$\mathbb{E}_x \mathbb{E}_d f(x)f(x+d)f(x+2d)\mu_B(d) \not\approx \alpha^3 \mathbb{E}_x \mathbb{E}_d \nu(x)\nu(x+d)\nu(x+2d)\mu_B(d).$$

or equivalently that

$$\sum_{r+s+t=0} \hat{f}(r)\hat{f}(s)\hat{f}(t)\hat{\mu}_B(r-t) \not\approx \alpha^3 \sum_{r+s+t=0} \hat{\nu}(r)\hat{\nu}(s)\hat{\nu}(t)\hat{\mu}_B(r-t).$$

It is not immediately obvious what to do with this information, but we know quite a lot about the function $\hat{\mu}_B$, so there seems to be a reasonable chance of being able to do something. Nevertheless, an advantage of Bourgain's approach is that we end up with a sum of the conventional kind – an AP count with different functions – rather than having four functions in play. As I say, when one of the four functions is highly structured and “linear” it is not necessarily a problem. However, at the very least one has to think a bit before deciding what to do next.

5. EQUAL-SHELLS MEASURE

Before I do that, I want to float another idea, inspired by the idea of equal-slices measure, which played a prominent part in the combinatorial proof of the density Hales-Jewett theorem that was discovered in Polymath1.

In that problem, concentration of measure caused difficulties. Roughly speaking, the main difficulty was that a random point on a random combinatorial line looked very different from a random point. Here we have a rather similar difficulty: a high-dimensional Bohr set is similar to a high-dimensional convex body. A typical progression of length 3 in a high-dimensional convex body will (I'm pretty sure) have the two outer points close to the boundary and the inner point a bit more central. But the more central area has

exponentially small measure (as a function of the dimension), so if you choose a random 3AP $(x, x + d, x + 2d)$ then $x + d$ will be distributed in a completely different way from a random point in B .

How did we deal with that problem in Polymath1? Well, the problem there was that a typical point in $[k]^n$ has roughly n/k coordinates of each value, whereas a typical point in a typical line had about $2n/(k+1)$ coordinates taking one value and $n/(k+1)$ coordinates for the other $k-1$ values. To get round this, we created a measure that did not concentrate at points with any particular distribution of the numbers of the various coordinates. We did that in the most obvious way: we simply chose uniformly at random how many coordinates would take each value (of course, our choices had to add up to n) and then, given our choice, we chose our point uniformly at random.

In a very similar spirit, one might consider creating a measure on a convex body K that is not concentrated at any particular radius, and in particular not concentrated near the boundary of K , by first choosing a random λ between 0 and 1 and then choosing a random point on the boundary of λK . Then the size of a random point (as measured by the smallest multiple of K that the point belongs to) would be uniformly distributed on $[0, 1]$ rather than concentrated near 1.

When it comes to Bohr sets, it is not quite so clear what it means to choose a random point on the boundary. But we can do something pretty similar. Given a Bohr set $B(K, \delta)$, we could choose a random point by first picking λ uniformly at random from $[0, 1]$ and then picking a point uniformly at random from $B(K, \lambda\delta)$. This gives us a measure $\mu_{K, \delta}$ that is closely related to the Bohr set but that is not concentrated near the boundary.

There are several variants of this basic idea, and the one I have chosen may well not be the best one. But before we try to optimize the definition it seems better to think about whether the definition itself is likely to be of any use. Let me try to address this question without doing any complicated calculations.

The problem with a high-dimensional Bohr set B was that the obvious quantity

$$\mathbb{E}_x \mathbb{E}_d \mu_B(x) \mu_B(x + d) \mu_B(x + 2d)$$

is very small. Indeed, it measures the probability that $x + 2d \in B$ given that $x \in B$ and $x + d \in B$, and this probability is exponentially small in the dimension of B (because $x + d$ has to occupy a rather central position in B).

Contrast this with the corresponding quantity with μ_B replaced by $\mu_{K, \delta}$. For the sake of discussion, I'll pretend that B is a centrally symmetric convex body in \mathbb{R}^k and write ν_B

for the measure you obtain if you pick a random $\lambda \in [0, 1]$ and then a random point in λB . What can we say about

$$\mathbb{E}_x \mathbb{E}_d \nu_B(x) \nu_B(x+d) \nu_B(x+2d)?$$

This is asking for the following. Choose λ_1 , λ_2 and λ_3 uniformly at random and then choose x and y randomly from $\lambda_1 B$ and $\lambda_2 B$, respectively. What is the probability that $2y - x \in \lambda_3 B$?

It is easy to see that this probability is at least some absolute constant. Indeed, if λ_1 and λ_2 are at most $1/4$ and λ_3 is at least $3/4$, then $2y - x$ is guaranteed to lie in $\lambda_3 B$, and the probability of this is at least $1/64$.

Now let A be a set with no 3APs such that $\nu_B(A) = \alpha$. (That is, if you choose a point at random according to the measure ν_B , then with probability α it belongs to A .) We would like to obtain a density increase on a measure that is of the same general type as ν_B but where the dimension of the Bohr set has gone up by 1. And we would like to do this without losing any width in the process. Is this feasible?

Here is a potentially serious problem. Suppose A consists of all points in $B \setminus (1 - \alpha)B$. (Again, I'm pretending that B is a symmetric convex body in \mathbb{R}^k and just assuming that it will be possible to say something similar about Bohr sets.) Then the measure of A will be precisely α . So do we get a density increase?

I can't quite get my head round this. One idea might be to take a wedge defined by a hyperplane tangent to B and intersect it with a translate of B , in an effort to pick up as much of A as one can.