1. Roth’s Theorem, Classically

I’ll be presenting a paper of Bloom and Sisask, [2] which provides a new proof of Roth’s theorem on 3-term arithmetic progressions. Their proof uses an almost periodicity argument in physical space, rather than relying on Fourier analysis, as many previous proofs have done. Crucially, it also gives a very good bound, decreasing the minimum density of a subset of $[1, N]$ in order to see arithmetic progressions to $(\log N)^{-1+o(1)}$.

Let’s start by stating (a version of) Roth’s theorem and outlining the proof, largely following [4].

**Theorem 1.1 (Roth, 1953).** There exists a positive constant $C$ so that if $A \subset [1, N]$ with $|A| \geq CN / \log \log N$, then $A$ has a non-trivial three term arithmetic progression.

In other words, if $A$ has no nontrivial three-term arithmetic progressions, then $|A| \ll N / \log \log N$.

Let $A \subset [1, N]$ with $|A| = \alpha N$. Broadly, the proof will proceed along these lines. Either $A$ is in some sense unstructured, in which case there will be many non-trivial 3APs, or $A$ doesn’t. In the latter case we’ll identify some structure of $A$ which will allow us to find a subset of $N$ on which $A$ has a bit higher density; this step is called a density increment. Iterating the density increment enough times will ultimately yield a subset on which $A$ has very high density, and then it will be easy to find a 3AP.

Many things in that outline were vague, but let’s start with the question of “having structure.” Historically, this has been done using Fourier analysis.

Let $B$ be the set of either odd or even terms in $A$, whichever is larger. Let $1_A$ be the characteristic function of $A$, and $1_B$ that of $B$. With

$$\hat{f}(r) = \sum_n f(n)e\left(-\frac{rn}{N}\right),$$

we have

$$\frac{1}{N} \sum_{r \equiv 0 \pmod{N}} \hat{1}_B(r)^2 \hat{1}_A(-2r) = \#\{x + y = 2z \pmod{N} : x, y \in B, z \in A\}.$$ 

Some of these will be trivial, i.e. with $x = y = z$, so the number of non-trivial 3APs is

$$\frac{1}{N} \sum_{r \equiv 0 \pmod{N}} \hat{1}_B(r)^2 \hat{1}_A(-2r) - |B| = \frac{|A||B|^2}{N} - |B| + \frac{1}{N} \sum_{r \neq 0} \hat{1}_B(r)^2 \hat{1}_A(-2r).$$
If $1_A$ has no large Fourier coefficients, i.e. for all $r \neq 0$ we have $|\hat{1}_A(r)| \leq a^2 N/4$, then this can be used to directly bound

$$\frac{1}{N} \left| \sum_{r \neq 0} \hat{1}_B(r)^2 \hat{1}_A(-2r) \right| \leq \frac{a^2}{4} \sum_r |\hat{1}_B(r)|^2 = \frac{a^2}{4} N|B| \leq \frac{|A||B|^2}{2N}.$$ 

Thus, using the triangle inequality with our formula for the number of non-trivial 3APs, we can see that there will be many non-trivial 3APs.

The “structured” case is then the case when $|\hat{1}_A(r)| \geq a^2 N/4$ for some $r$. In this case the goal is to perform a density increment. We’ll fix two parameters $M$ and $Q$, which will depend on $N$. By Dirichlet’s theorem on rational approximation, there exists some $b/q$ with $q \leq Q$, $(b, q) = 1$, such that $|r/N - b/q| \leq \frac{1}{4Q}$. We divide $[1, N]$ into progressions $\pmod{q}$, and subdivide each progression into $M$ intervals. These $qM$ intervals, each with $N/(qm) + O(1)$ elements, are the subsets we’ll consider; we’ll show that $A$ has high density on one of these intervals.

The benefit of the intervals as we’ve chosen them is that $e(ar/N)$ changes very little on a typical interval. In particular, $e(ar/N) = e(ab/q + a\theta)$ with $|\theta| \leq 1/qQ$. Since elements of an interval lie in the same progression $\pmod{q}$, $e(ab/q)$ is constant. The variation in $e(a\theta)$ is at most $O(N|\theta|/M) = O(N/(qQM))$.

Since $|\hat{1}_A(r)| \geq a^2 N/2$,

$$\left| \sum_{\alpha=1}^N (1_A(a) - \alpha)e(ar/N) \right| \geq \frac{a^2}{2} N.$$ 

After some computation with splitting this sum up in terms of the intervals $I$ above, this implies

$$\frac{a^2 N}{2} \leq \sum_I \left| \sum_{\alpha \in I} (1_A(a) - \alpha) \right| + O\left( \frac{N^2}{qQM} \right).$$ 

Since

$$0 = \sum_I \sum_{\alpha \in I} (1_A(a) - \alpha),$$ 

there must be an interval $I$ with

$$\sum_{\alpha \in I} (1_A(a) - \alpha) \geq \frac{a^2 N}{8qM},$$ 

and appropriate choice of $Q$ and $M$ here, specifically $Q = \sqrt{N}$ and $M = C\sqrt{N}/(qa^2)$ for large $C$, the relative density of $A$ within $I$ is at least $\alpha + a^2/16$.

The idea is then to dilate and translate $I$, which preserves 3APs, and then iterate the argument applied to $I$. In the end for this to work, we need $\alpha > C/\log \log N$.

2. Historical Improvements and Bloom and Sisask’s Result

The main area of improvement has been to decrease the lower bound on the density $\alpha$. If $R(N)$ is the size of the largest subset of $\{1, \ldots, N\}$ with no non-trivial 3AP, we’d like a better upper bound for $R(N)$. The history of the best known upper bounds is below [1]:
ROTH’S THEOREM: LOGARITHMIC BOUNDS VIA ALMOST-PERIODICITY

<table>
<thead>
<tr>
<th>Result</th>
<th>$R(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roth [1953]</td>
<td>$N / \log \log N$</td>
</tr>
<tr>
<td>Szemerédi [1990], Heath-Brown [1987]</td>
<td>$N / (\log N)^c$ for some $c &gt; 0$</td>
</tr>
<tr>
<td>Bourgain [1999]</td>
<td>$(\log \log N)^{1/2}N / (\log N)^{1/2}$</td>
</tr>
<tr>
<td>Bourgain [2008]</td>
<td>$(\log \log N)^{2}N / (\log N)^{2/3}$</td>
</tr>
<tr>
<td>Sanders [2012]</td>
<td>$N / (\log N)^{3/4-o(1)}$</td>
</tr>
<tr>
<td>Sanders [2011]</td>
<td>$(\log \log N)^{6}N / \log N$</td>
</tr>
<tr>
<td>Bloom [2016]</td>
<td>$(\log \log N)^{4}N / \log N$</td>
</tr>
</tbody>
</table>

Our goal here is to prove that $R(N) \ll N / (\log N)^{1-o(1)}$. The approach will be using an almost-periodicity result, with very little Fourier analysis. We will not worry about optimizing the precise power of $\log \log N$, but it is worth noting that this technique can give $(\log \log N)^7N / \log N$ but does not directly give a result better than Bloom [2016].

The main theorem is the following, somewhat more general result.

**Theorem 2.1.** Let $G$ be a finite abelian group of odd order, and let $A \subseteq G$ be a set of density $\alpha > 0$. Let $T(A)$ be the number of 3APs in $A$; then

$$T(A) \geq \exp(-C\alpha^{-1}(\log 2/\alpha)^C)|A|^2,$$

for $C > 0$ an absolute constant.

In this case setting $\alpha \geq (C+1)(\log \log |G|)^C / \log |G|$, say, gives that $T(A) > |A|$. Note also that this subsumes our goal by embedding $A \subseteq \{1, \ldots, N\}$ into $\mathbb{Z}/(2N+1)\mathbb{Z}$, say.

We’ll start by looking at the finite field case in a fair amount of detail to see how these arguments work, and then talk about how to generalize.

### 3. Notation and Normalization

For a subset $A \subseteq G$, we will write $1_A$ for the indicator function of $A$, and $\mu_A$ for the function $1_A / |A|$. We will use a discretely normalized Haar measure on $G$, so that

$$f \ast g(x) = \sum_{y \in G} f(y)g(x-y),$$

and

$$\langle f, g \rangle = \sum_{y \in G} f(y)\overline{g(y)}.$$

The $L^p$ norm is defined as usual, with

$$||f||_p^p = \frac{1}{|G|} \sum_{y \in G} |f(y)|^p.$$

We will also make use of Hölder’s inequality for convolutions, specifically that if $\frac{1}{p} + \frac{1}{q} = 1$, then

$$||f \ast g||_\infty \leq |G|||f||_p||g||_q.$$

Note that for $A, B \subseteq G$,

$$1_A \ast \mu_B(x) = E_{t \in B} 1_A(x-t) = \frac{1}{|B|} \sum_{t \in B} 1_A(x-t),$$
and the number of 3APs in $A$ is 
\[ T(A) = \sum_{x+z=2y} \mathbb{1}_A(x) \mathbb{1}_A(y) \mathbb{1}_A(z) = \sum_{x \in G} \mathbb{1}_A \ast \mathbb{1}_A(x) \mathbb{1}_{2,A}(x) = (\mathbb{1}_A \ast \mathbb{1}_A, \mathbb{1}_{2,A}). \]

4. A New Kind of Density Increment: Finite Field Case

For the following section, we will set $G = \mathbb{F}_q^n$, for $\mathbb{F}_q$ a finite field. We’ll get the following theorem with relatively few technical hurdles; in the next section, we’ll see how this argument needs to be adjusted to apply to other cases.

**Theorem 4.1.** Let $A \subseteq \mathbb{F}_q^n$ be a subset with density $\alpha$ and $T(A) \leq \frac{\alpha}{2} |A|^2$. Then there is a subspace $V$ with codimension $\ll (\log(2/\alpha))^C \alpha^{-1}$ such that $||\mathbb{1}_A \ast \mu_V||_\infty \geq \frac{5}{4} \alpha$.

The conclusion is saying that there exists some $x$ with $(x+A) \cap V$ having density $\geq \frac{5}{4} \alpha$ in $V$, which gives us a subspace that we can pass to and iterate. In other words, this is precisely a density increment.

We’ve said that we’ll rely on almost-periodicity, so let’s state the almost-periodicity result that we use.

**Theorem 4.2 (L$^p$ almost periodicity).** Let $p \geq 2$ and $\varepsilon \in (0, 1)$. Let $G = \mathbb{F}_q^n$ be a vector space over a finite field, with $A \subseteq G$ a subset with $|A| \geq \alpha |G|$. Then there is a subspace $V \leq G$ of codimension $d \ll p\varepsilon^{-2} \log(2/\varepsilon)^2 \log(2/\alpha)$ so that 

\[ ||\mu_A \ast \mathbb{1}_A \ast \mu_V - \mu_A \ast \mathbb{1}_A||_p \leq \varepsilon ||\mu_A \ast \mathbb{1}_A||_p^{1/2} + \varepsilon^2. \]

To unpack this just a bit, note that $\mu_A \ast \mathbb{1}_A \ast \mu_V$ is the average over elements $t \in V$ of $\mu_A \ast \mathbb{1}_A (\cdot + t)$. The proof shows that $\mu_A \ast \mathbb{1}_A$ is “close” to translates via elements of $V$ in the sense that its $L^p$ norm is bounded, which means that the same holds for the average.

We now proceed with the proof of Theorem 4.1. We’ll split into two cases: the first, when $||\mu_A \ast \mathbb{1}_A||_2$ is small for some large $m$, and the second where $||\mu_A \ast \mathbb{1}_A||_2$ is large for some large $m$.

4.1. Case 1: $||\mu_A \ast \mathbb{1}_A||_2$ is small for some $m$.

**Lemma 4.3.** Let $A \subseteq G = \mathbb{F}_q^n$ with density $\alpha$ and $T(A) \leq \frac{\alpha}{2} |A|^2$. If $m \gg \log(2/\alpha)$ with 

\[ ||\mu_A \ast \mathbb{1}_A||_2 \leq 10 \alpha, \]

then there is a subspace $V$ with codimension $\ll (\log(2/\alpha)^C m \alpha^{-1}$ with $||\mathbb{1}_A \ast \mu_V||_\infty \geq \frac{5}{4} \alpha$.

**Proof.** Apply Theorem 4.2 with $p = 4m$ and $\varepsilon = \alpha^{1/2}/100$. This yields a subspace $V$ of codimension $d \ll 400m/\alpha \log(200/\alpha^{1/2})^2 \log(2/\alpha) \ll (\log(2/\alpha))^{C \alpha^{-1}}$ with 

\[ ||\mu_A \ast \mathbb{1}_A \ast \mu_V - \mu_A \ast \mathbb{1}_A||_4 \leq \varepsilon ||\mu_A \ast \mathbb{1}_A||_2^{1/2} + \varepsilon^2 \leq \frac{\alpha}{100} \left( \alpha^{-1/2} ||\mu_A \ast \mathbb{1}_A||_2^{1/2} + 1 \right) \leq \alpha/8. \]

Let $r$ be such that $1/r + 1/4m = 1$; by Hölder’s inequality, 

\[ ||\mu_A \ast \mathbb{1}_A \ast \mathbb{1}_{-2,A} \ast \mu_V - \mu_A \ast \mathbb{1}_A \ast \mathbb{1}_{-2,A}||_\infty \leq |G||\mathbb{1}_{-2,A}||_r ||\mu_A \ast \mathbb{1}_A \ast \mu_V - \mu_A \ast \mathbb{1}_A||_4 m \leq |G|\alpha^{1/r} (\alpha/8) = |G|\alpha^{2-1/4m}/8. \]
Let's compare the values at 0, which by the above differ by at most $|G|\alpha^2/4$. We assumed that $T(A) \leq \frac{\alpha}{2}|A|^2$. Since $T(A) = \langle 1_A * 1_A, 1_{2,A} \rangle$, we have:

\[
\langle 1_A * 1_A, 1_{2,A} \rangle \leq \frac{\alpha}{2}|A|^2
\]

\[
\Rightarrow 1_A * 1_A * 1_{-2,A}(0) \leq \frac{\alpha}{2}|A|^2
\]

\[
\Rightarrow \mu_A * 1_A * 1_{-2,A}(0) \leq \frac{\alpha}{2}|A| = \frac{\alpha^2}{2}|G|.
\]

Using this with our $L^\infty$ bound and the triangle inequality gives

\[
\mu_A * 1_A * 1_{-2,A} * \mu_V(0) \leq \frac{\alpha^2}{4}|G| + \frac{\alpha^2}{2}|G|
\]

\[
\Rightarrow 1_A * 1_A * 1_{-2,A} * \mu_V(0) \leq |A||G| \left( \frac{3\alpha^2}{4} = \frac{3}{4}\alpha^3|G|^2. \right)
\]

We'd still like to convert this upper bound into a lower bound for $\|\|1_A * \mu_V\|_\infty$. Assume that $\|1_A * \mu_V\|_\infty \leq (1+c)\alpha$, and let $f(x) = (1+c)^{-1}\alpha^{-1}1_A * \mu_V(x)$. Note that $0 \leq f(x) \leq 1$, and that

\[
\|f\|_1 = \frac{(1+c)^{-1}\alpha^{-1}}{|G|} \sum_{y \in G} 1_A * \mu_V(y)
\]

\[
= \frac{(1+c)^{-1}\alpha^{-1}}{|G|} \sum_{z \in G} 1_A(z) \left( \sum_{y \in G} \mu_V(y - z) \right)
\]

\[
= \frac{(1+c)^{-1}\alpha^{-1}}{|G|} \sum_{z \in G} 1_A(z)
\]

\[
= \frac{(1+c)^{-1}\alpha^{-1}}{|G|} |A| = (1+c)^{-1}.
\]

Thus considering $(1-f) * (1-f)$, we get

\[
0 \leq (1-f) * (1-f) = f * f - 2|G\|\|f\|_1 + |G| = (1+c)^{-2}\alpha^{-2}1_A * 1_A * \mu_V - \frac{1-c}{1+c}|G|.
\]

In particular, this implies that

\[
(1-c^2)\alpha^2|G| \leq 1_A * 1_A * \mu_V(x)
\]

for all $x$, so taking the inner product with $1_{2,A}$ implies

\[
(1-c^2)\alpha^2|G||A| = (1-c^2)\alpha^3|G|^2 \leq \langle 1_A * 1_A * \mu_V, 1_{2,A} \rangle \leq \frac{3}{4}\alpha^3|G|^2,
\]

so choosing $c = 1/4$ gives a contradiction, which in turn implies that $\|1_A * \mu_V\|_\infty > \frac{5}{4}\alpha$. \hfill $\square$

4.2. Case 2: $\|\mu_A * 1_A\|_{2m}$ is large for some $m$. We'll now turn to address the case when one of the $L^{2m}$ norms is large; this case is in fact a more direct application of Theorem 4.2.

**Lemma 4.4.** Assume that $\|\mu_A * 1_A\|_{2m} \geq 10\alpha$. Then there is a subspace $V$ of codimension $\ll (\log (2/\alpha))^C m\alpha^{-1}$ such that $\|1_A * \mu_V\|_\infty \geq 5\alpha$. 

Proof. Again, we’ll start by applying Theorem 4.2, but in this case with \( p = 2m \). Again we use \( \epsilon = \alpha^{1/2}/100 \). Theorem 4.2 yields a subspace \( V \) of codimension
\[
d \ll (200m/\alpha) \log(200/\alpha^{1/2})^2 \log(2/\alpha) \ll (\log(2/\alpha))^C m \alpha^{-1}.
\]
The subspace \( V \) satisfies
\[
||\mu_A * 1_A * \mu_V - \mu_A * 1_A||_{2m} \leq \frac{\alpha}{100} \left( \alpha^{-1/2} ||\mu_A * 1_A||_{m}^{1/2} + 1 \right).
\]
By the triangle inequality,
\[
||\mu_A * 1_A * \mu_V||_{2m} \geq ||\mu_A * 1_A||_{2m} - \frac{\alpha}{100} \left( \alpha^{-1/2} ||\mu_A * 1_A||_{m}^{1/2} + 1 \right).
\]
Since for \( f \geq 0 \) we have \( ||f||_p \leq ||f||_q \) whenever \( p \leq q \), we can replace \( ||\mu_A * 1_A||_{m} \) above with \( ||\mu_A * 1_A||_{2m} \) to get
\[
||\mu_A * 1_A * \mu_V||_{2m} \geq ||\mu_A * 1_A||_{2m} - \frac{\alpha}{100} \left( \alpha^{-1/2} ||\mu_A * 1_A||_{2m}^{1/2} + 1 \right).
\]
However, \( ||\mu_A * 1_A||_{2m} \geq 10\alpha \). Considering the above as a function of \( x = ||\mu_A * 1_A||_{2m}^{1/2} \), specifically \( f(x) = x^2 - \frac{\alpha}{100} x - \frac{\alpha}{100} \), the minimum of \( f(x) \) is at \( x = \frac{\sqrt{\alpha}}{200} < \sqrt{10\alpha} \), so the smallest value of \( f(x) \) among \( x \geq \sqrt{10\alpha} \) is when \( x = \sqrt{10\alpha} \). Plugging this in shows that
\[
||\mu_A * 1_A * \mu_V||_{\infty} \geq 5\alpha,
\]
say, where 5 is not chosen particularly carefully.
Thus
\[
||1_A * \mu_V||_{\infty} \geq ||\mu_A * 1_A * \mu_V||_{\infty} \geq ||\mu_A * 1_A * \mu_V||_{2m} \geq 5\alpha,
\]
which is the desired density increment. \( \square \)

So now we have the density increment that we wanted; these two cases imply Theorem 4.1.

These lower bounds on \( ||1_A * \mu_V||_{\infty} \) show that some translate of \( A \) has higher density, since
\[
||1_A * \mu_V||_{\infty} = \max_{t \in G} \sum_{y \in G} 1_A(y) \mu_V(t - y) = \max_{t \in G} \frac{1}{|V|}(|(t - A) \cap V|).
\]

Let’s briefly see how this gives the precise statement of Theorem 2.1. Translating \( A \) still preserves three-term arithmetic progressions, so at every step we either have a subspace \( V \) so that some translate \( t + A \) of \( A \) has \( \geq \frac{\alpha}{2}|(t + A) \cap V|^2 \), or we can find a further subspace of \( V \) with increased density. The first question is, how many subspaces do we need to take?

If \( k \geq \frac{\log(1/\alpha)}{\log(5/4)} \), then \( 1 < (\frac{5}{4})^k \alpha \), so the number of iterations can’t be more than \( \ll C(\log(1/\alpha)) \). At that point, we have a subspace of \( \mathbb{F}_q^n \) of codimension \( \ll k \log(2/\alpha)^C \alpha^{-1} \ll (\log(2/\alpha))^C \alpha^{-1} \), where the Cs are not necessarily equal but are each absolute constants.

Thus we must have a subspace \( V \) of codimension \( \ll (\log(2/\alpha))^C \alpha^{-1} \) with
\[
T((t + A) \cap V) \geq \frac{\alpha}{2}|(t + A) \cap V|^2,
\]
which is the precise statement of Theorem 2.1.
where \(|(t + A) \cap V| \geq \alpha|V|\). Thus

\[
T(A) \geq T((t + A) \cap V) \\
\geq \frac{\alpha}{2} \alpha|V|^2 \\
= \frac{\alpha}{2} |A|^2 q^{-\text{codim}(V)} \\
= \frac{\alpha}{2} |A|^2 \exp(-C(\log(2/\alpha))^C \alpha^{-1}) \\
= |A|^2 \exp(-C(\log(2/\alpha))^C \alpha^{-1} - \log(2/\alpha)),
\]

but the \(\log(2/\alpha)\) is of smaller order, so for appropriate choice of constants it can be omitted.

This is exactly the desired statement!

4.3. A few notes about the transition. I won’t go into detail about the general case (or even the integer case), but I do want to mention an important ingredient that allows these same ideas to work in greater generality. Specifically, we frequently and crucially passed to subspaces in the vector space case; in general, we need a different kind of structure that we can pass to. This is accomplished by defining Bohr sets.

**Definition 4.5.** Let \(G\) be a finite abelian group and let \(\hat{G} = \{\gamma : G \to \mathbb{C}^\times\}\) be the dual group of \(G\). For a subset \(\Gamma \subseteq \hat{G}\) and a constant \(\rho \geq 0\), the Bohr set corresponding to \(\Gamma\) and \(\rho\) is defined as

\[
\text{Bohr}(\Gamma, \rho) = \{x \in G : |\gamma(x) - 1| \leq \rho \ \forall \gamma \in \Gamma\}.
\]

In the vector space case, the dual group is the group of linear functionals, and subspaces and their translates are Bohr sets with \(\rho = 0\). For arbitrary \(G\), one can prove \(L^p\)-almost-periodicity results relative to Bohr sets instead of to subspaces, and then follow a similar argument to the above to yield a density increment.

5. Background on almost-periodicity

At various times we crucially used Proposition 4.2, so let’s talk a bit about what goes into proving it. We will prove Proposition 3.1 from [3], which has a somewhat different statement; the biggest difference being that it only addresses \(L^2\) almost-periodicity, rather than \(L^p\). However, the proof still contains many of the same ideas.

**Proposition 5.1 (\(L^2\)-almost-periodicity, left-translates).** Let \(G\) be an abelian group, let \(A, B \subseteq G\) be finite subsets, and fix a parameter \(\varepsilon \in (0, 1)\). Let \(S \subseteq G\) be a subset such that \(|S + A| \leq K|A|\). Then there is a set \(T \subseteq -S\) of size

\[
|T| \geq \frac{|S|}{(2K)^{9/\varepsilon^2}}
\]

such that for all \(t \in T - T\),

\[
||1_A * 1_B(\cdot + t) - 1_A * 1_B||_2^2 \leq \varepsilon^2|A|^2|B|.
\]

**Proof.** Let \(k\) be an integer with \(1 \leq k \leq |A|/2\); we will fix \(k\) later. Let \(C \subseteq A\) be a subset of size \(|C| = k\), which we choose uniformly randomly out of all such sets. All
We also consider the variance
\[ \text{Var}(v_C * 1_B(x)) = \mathbb{E}_C|v_C * 1_B(x) - 1_A * 1_B(x)|^2, \]
where again the expectation is taken over the choice of set C. The variance satisfies
\[ \text{Var}(v_C * 1_B(x)) \leq \frac{|A|}{k} 1_A * 1_B(x). \]

We can then sum this inequality over all \( x \in A + B \), since \( A + B \) is the support of \( 1_A * 1_B \). This gives
\[ \mathbb{E}_C|v_C * 1_B - 1_A * 1_B|^2 \leq |A|^2|B|/k. \]

We say that \( C \) approximates \( A \) if
\[ ||v_C * 1_B - 1_A * 1_B||_2 \leq 2|A|^2|B|/k. \]

By the expectation bound and Markov’s inequality,
\[ \mathbb{P}_C(C \text{ approximates } A) \geq 1/2. \]

Now let \( Y = S + A \) and let \( t \in -S \), so that \( A \subseteq tY \). Then
\[
\mathbb{P}_{C \in \binom{\nu}{k}}(tC \text{ approximates } A) = \mathbb{P}_{C \in \binom{\nu}{k}}(C \text{ approximates } A) \\
\geq \mathbb{P}_{C \in \binom{\nu}{k}}(C \subseteq A) \mathbb{P}_{C \in \binom{\nu}{k}}(C \text{ approximates } A) \\
\geq \left( \frac{|A|}{k} \right) \left( \frac{|S + A|}{k} \right)^{-1} \frac{1}{2} \\
\geq \frac{1}{(2K)^k},
\]
the last step using the hypothesis that \( |S + A| \leq K|A| \). Summing this over all \( t \in -S \) gives
\[ \mathbb{E}_{C \in \binom{\nu}{k}}\{ |t \in -S : tC \text{ approximates } A \} \geq \frac{|S|}{(2K)^k}. \]

So, there exists some set \( C \) which is above average, i.e. for which the size of \( T = \{ t \in -S : tC \text{ approximates } A \} \) is at least \( |S|/(2K)^k \). For this \( C \), we have
\[ ||\mu_C * 1_B - 1_A * 1_B(\cdot + t)||_2 \leq 2|A|^2|B|/k \]
for all \( t \in T \), so by the triangle inequality, for all \( t \in T - T \) we have
\[ ||1_A * 1_B(\cdot + t) - 1_A * 1_B||_2 \leq 8|A|^2|B|/k. \]

Fixing \( k = \lceil 8/e^2 \rceil \) completes the proof of the proposition. \( \square \)
The $L^p$ version instead relies on higher moments of random variables that look like $1_C \ast 1_B$, which follow a hypergeometric distribution.

REFERENCES