

# Kelley and Meka's proof of Roth's theorem

by

MARCEL K. GOH

23 AUGUST 2023

**Note.** This exposition of Kelley and Meka's proof closely follows that of Bloom and Sisask, but with many extra details supplied. I hope that it will be useful to other students at the Ph.D.-student level.

## 1. Definitions and elementary facts

We will use  $G$  primarily to refer to a finite abelian group. For functions  $f, g : G \rightarrow \mathbf{C}$  we have the inner product

$$\langle f, g \rangle = \mathbf{E}_{x \in G} f(x) \overline{g(x)}$$

and the  $L_p$  norm

$$\|f\|_p = \left( \mathbf{E}_{x \in G} |f(x)|^p \right)^{1/p}.$$

In  $L_p$  spaces we have the useful Hölder inequality

$$|\langle f, g \rangle| \leq \|f\|_p \cdot \|g\|_q,$$

for  $p, q \in [1, \infty)$  with  $1/p + 1/q = 1$ . Assuming now that  $f$  and  $g$  are  $\mathbf{R}$ -valued, we also have the convolution

$$(f * g)(x) = \mathbf{E}_{y \in G} f(y) g(x - y)$$

and the difference convolution

$$(f \circ g)(x) = \mathbf{E}_{y \in G} f(y) g(x + y).$$

It is easy to check that for all  $x \in G$ ,  $(f * g)(x) = (g * f)(x)$ , but with the difference convolution we have  $(f \circ g)(x) = (g \circ f)(-x)$ . We also have the following adjoint property.

**Proposition 1** (*Adjoint property*). *Let  $G$  be a finite abelian group and let  $f, g, h : G \rightarrow \mathbf{R}$ . Then*

$$\langle f, g * h \rangle = \langle h \circ f, g \rangle.$$

*Proof.* First expand

$$\begin{aligned}\langle f, g * h \rangle &= \mathbf{E}_{x \in G} f(x)(g * h)(x) \\ &= \mathbf{E}_{x \in G} f(x) \mathbf{E}_{y \in G} g(y)h(x - y) \\ &= \mathbf{E}_{y \in G} g(y) \mathbf{E}_{x \in G} f(x)h(x - y).\end{aligned}$$

Then substituting  $z = x - y$  so that  $x = z + y$  yields

$$\begin{aligned}\langle f, g * h \rangle &= \mathbf{E}_{y \in G} g(y) \mathbf{E}_{z \in G} f(z + y)h(z) \\ &= \mathbf{E}_{z \in G} (h \circ f)(z)g(z) \\ &= \langle h \circ f, g \rangle. \quad \blacksquare\end{aligned}$$

For a group  $G$  the dual group  $\widehat{G}$  is the set of all homomorphisms from  $G$  to  $\mathbf{C}^\times$ . The Fourier transform of  $f : G \rightarrow \mathbf{R}$  is the function  $\widehat{f} : \widehat{G} \rightarrow \mathbf{C}$  given by

$$\widehat{f}(\chi) = \mathbf{E}_{x \in G} f(x)\chi(-x).$$

The following proposition describes how the convolution and difference convolution behave under the Fourier transform.

**Proposition 2** (*Convolution laws*). *Let  $G$  be a finite abelian group and let  $f, g : G \rightarrow \mathbf{R}$ . Then the following identities hold:*

- i)  $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$
- ii)  $\widehat{f \circ g} = \widehat{f} \cdot \widehat{g}$

In particular,  $\widehat{f \circ f} = |\widehat{f}|^2$ .

*Proof.* Expand

$$\widehat{f * g}(\chi) = \mathbf{E}_{x \in G} (f * g)(x)\chi(-x)$$

and multiply the right-hand side by  $1 = \chi(-y)\chi(y)$  to get

$$\widehat{f * g}(\chi) = \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} f(y)g(x - y)\chi(-y)\chi(y - x).$$

Then we may interchange the order of summation and substitute  $z = x - y$  to arrive at

$$\widehat{f * g}(\chi) = \mathbf{E}_{y \in G} \mathbf{E}_{z \in G} f(y)g(z)\chi(-y)\chi(-z) = \widehat{f}(\chi)\widehat{g}(\chi),$$

which proves (i). For part (ii), we expand and multiply by the same 1 to get

$$\widehat{f \circ g}(\chi) = \mathbf{E}_{x \in G} (f \circ g)(x)\chi(-x) = \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} f(y)g(x + y)\chi(y)\chi(-x - y).$$

We again interchange the order of summation; this time substituting  $z = x + y$  gives us

$$\begin{aligned}\widehat{f \circ g}(\chi) &= \mathbf{E}_{y \in G} \mathbf{E}_{z \in G} f(y)g(z)\chi(y)\chi(-z) \\ &= \overline{\mathbf{E}_{y \in G} f(y)\chi(-y)} \mathbf{E}_{z \in G} g(z)\chi(-z) \\ &= \overline{\widehat{f}(\chi)} \widehat{g}(\chi),\end{aligned}$$

which is what we wanted.  $\blacksquare$

When we convolve two functions  $\hat{f}$  and  $\hat{g}$  on the dual group, we take a sum instead of an expectation:

$$(\hat{f} \circ \hat{g})(\chi) = \sum_{\psi \in G} \hat{f}(\psi) \hat{g}(\chi\psi^{-1}).$$

The same goes in the definition of the inner product  $\langle \hat{f}, \hat{g} \rangle$ .

Let  $f^{*k}$  denote the  $k$ -fold convolution of a function  $f$ . The next proposition interprets  $k$ -norms in terms of  $k$ -fold convolutions of the Fourier transform.

**Proposition 3.** *Let  $G$  be a finite abelian group, let  $k \geq 1$  be an integer, and let  $\chi_0$  denote the identity element of the dual group  $\hat{G}$  of  $G$ . We have the identity*

$$\mathbf{E}_{x \in G} f(x)^k = \hat{f}^{*k}(\chi_0).$$

*Proof.* Expand by the Fourier inversion formula to get

$$\begin{aligned} \mathbf{E}_{x \in G} f(x)^k &= \mathbf{E}_{x \in G} \left( \sum_{\chi \in \hat{G}} \hat{f}(\chi) \chi(x) \right)^k \\ &= \mathbf{E}_{x \in G} \sum_{\chi_1 \in \hat{G}} \cdots \sum_{\chi_k \in \hat{G}} \hat{f}(\chi_1) \cdots \hat{f}(\chi_k) \chi_1(x) \cdots \chi_k(x) \\ &= \sum_{\chi_1 \in \hat{G}} \cdots \sum_{\chi_k \in \hat{G}} \hat{f}(\chi_1) \cdots \hat{f}(\chi_k) \mathbf{E}_{x \in G} \chi_1 \cdots \chi_k(x). \end{aligned}$$

By orthogonality of characters, the inner expectation is zero when  $\chi_1 \cdots \chi_k \neq \chi_0$ , so we have

$$\mathbf{E}_{x \in G} f(x)^k = \sum_{\chi_1 \cdots \chi_k = \chi_0} \hat{f}(\chi_1) \cdots \hat{f}(\chi_k) = \hat{f}^{*k}(\chi_0). \quad \blacksquare$$

For sets  $A$  and  $X$ , let  $\mu_X(A) = |A \cap X|/|X|$  denote the relative density of  $A$  in  $X$ , and if  $X$  is understood to be a subset of a larger set  $G$ , then we use  $\mu_X$  also to denote the normalised indicator function given by

$$\mu_X(x) = \begin{cases} 1/\mu_G(X), & \text{if } x \in X; \\ 0, & \text{otherwise.} \end{cases}$$

The scaling is done so that  $\|\mu_X\|_1 = 1$  for any  $X \subseteq G$ , as can easily be checked. We denote the ordinary indicator function by  $\mathbf{1}_X = \mu_G(X)\mu_X$ , and sometimes write  $\mathbf{1}_x$  for the indicator function  $\mathbf{1}_{\{x\}}$  of a singleton set. Lastly, we also sometimes use the same symbol to denote the indicator function of a statement; i.e.,  $\mathbf{1}_{[P]}$  is 1 if the statement  $P$  is true and 0 if it is false.

It is easy to check that if  $\mu$  has  $\|\mu\|_1 = 1$ , then so does  $\mu * \mu$  and  $\mu \circ \mu$ . We shall say that  $\mu : G \rightarrow \mathbf{R}_{\geq 0}$  is a *probability measure* on  $G$  if  $\|\mu\|_1 = 1$ . The following proposition concerns such measures.

**Proposition 4.** *Let  $G$  be a finite abelian group. If  $\mu : G \rightarrow \mathbf{R}_{\geq 0}$  is a probability measure, then*

$$\widehat{\mu - 1} = \widehat{\mu}(1 - \mathbf{1}_{\chi_0}).$$

*Proof.* We expand

$$\begin{aligned} \widehat{\mu - 1}(\chi) &= \mathbf{E}_{x \in G} (\mu(x) - 1) \chi(-x) \\ &= \mathbf{E}_{x \in G} \mu(x) \chi(-x) - \mathbf{E}_{x \in G} \chi(-x). \\ &= \widehat{\mu}(\chi) - \mathbf{E}_{x \in G} \chi(-x). \end{aligned}$$

If  $\chi \neq \chi_0$ , then the expectation vanishes, and if  $\chi = \chi_0$ , then the expectation clearly equals 1, and  $\mu(\chi_0) = \|\mu\|_1 = 1$ , so the whole expression is zero.  $\blacksquare$

If  $\mu : G \rightarrow \mathbf{R}_{\geq 0}$  is a probability measure and  $f, g : G \rightarrow \mathbf{C}$  we write

$$\langle f, g \rangle_\mu = \mathbf{E}_{x \in G} \mu(x) f(x) \overline{g(x)}$$

for the inner product relative to  $\mu$ , and for  $1 \leq p < \infty$  we write

$$\|f\|_{p(\mu)} = \left( \mathbf{E}_{x \in G} \mu(x) |f(x)|^p \right)^{1/p}$$

for the  $L_p$  norm relative to  $\mu$ . The following basic proposition establishes the monotonicity of  $L_p$  norms with respect to  $p$ .

**Proposition 5** (*Monotonicity of  $L_p$  norms*). *Let  $G$  be a finite abelian group. Let  $\mu : G \rightarrow \mathbf{R}_{\geq 0}$  be a probability measure and let  $f : G \rightarrow \mathbf{C}$ . For  $1 \leq p < q < \infty$ , we have*

$$\|f\|_{p(\mu)} \leq \|f\|_{q(\mu)}.$$

*Proof.* Let  $r = q/p > 1$  and let  $s = r/(r - 1)$  so that  $1/r + 1/s = 1$ . We have

$$\mathbf{E}_{x \in G} \mu(x) |f(x)|^p = \mathbf{E}_{x \in G} \mu(x) |f(x)|^{q/r} \cdot 1$$

Now by Hölder's inequality, we have

$$\begin{aligned} \mathbf{E}_{x \in G} \mu(x) |f(x)|^p &\leq \left( \mathbf{E}_{x \in G} \mu(x) |f(x)|^q \right)^{1/r} \left( \mathbf{E}_{x \in G} 1^s \right)^{1/s} \\ &= \left( \mathbf{E}_{x \in G} \mu(x) |f(x)|^q \right)^{p/q}. \end{aligned}$$

Taking  $p$ th roots of both sides now produces the inequality we wanted.  $\blacksquare$

For convenience, when  $\mu : G \rightarrow \mathbf{R}_{\geq 0}$  is a probability measure and  $X \subseteq G$ , we write

$$\mu(X) = \|\mathbf{1}_X\|_{1(\mu)} = \mathbf{E}_{x \in G} \mu(x) \mathbf{1}_X(x),$$

and refer to this quantity as the *density of  $X$  relative to  $\mu$* .

## 2. Hölder lifting and unbalancing for finite groups

With preliminaries out of the way, we begin the proof of Kelley and Meka [2], as described and reworked by Bloom and Sisask [1]. In this section we perform the first two steps of the proof, in the general setting of finite groups.

**Lemma 6** (*Hölder lifting*). *Let  $\epsilon \geq 0$  and let  $A$  and  $C$  be subsets of a finite abelian group  $G$ , where  $C$  has relative density at least  $\gamma$ . Then at least one of the following two statements holds.*

- i)  $|\langle \mu_A * \mu_A, \mu_C \rangle - 1| \leq \epsilon$
- ii)  $\|\mu_A \circ \mu_A - 1\|_p \geq \epsilon/2$  for some  $p = O(\log(1/\gamma))$ .

*Proof.* Linearity of the inner product in the first argument gives

$$\langle \mu_A * \mu_A - 1, \mu_C \rangle = \langle \mu_A * \mu_A, \mu_C \rangle + \langle -1, \mu_C \rangle = \langle \mu_A * \mu_A, \mu_C \rangle - 1,$$

so if the first statement does not hold, then for  $q = 1/(1 - 1/p)$ , we have, by Hölder's inequality,

$$\begin{aligned} \epsilon &< |\langle \mu_A * \mu_A - 1, \mu_C \rangle| \leq \|\mu_A * \mu_A - 1\|_p \left( \mathbf{E}_{x \in G} |\mu_C(x)|^q \right)^{1/q} \\ &\leq \|\mu_A * \mu_A - 1\|_p \gamma^{1/q-1} \leq \|\mu_A * \mu_A - 1\|_p \gamma^{-1/p}. \end{aligned}$$

Letting  $p$  be an even integer greater than  $\log_2(1/\gamma)$ , we have  $\log \gamma \geq p \log(1/2)$ , whence  $\gamma^{1/p} \geq 1/2$ . This gives the inequality

$$\|\mu_A * \mu_A - 1\|_p \geq \frac{\epsilon}{2}.$$

Since  $p$  is even,

$$\|\mu_A * \mu_A - 1\|_p^p = \mathbf{E}_{x \in G} |(\mu_A * \mu_A - 1)(x)|^p = \mathbf{E}_{x \in G} (\mu_A * \mu_A - 1)(x)^p,$$

and we can apply Proposition 3 to get

$$\|g\|_p^p = \widehat{g}^{*p}(\chi_0),$$

where we have put  $g = \mu_A * \mu_A - 1$ . It was noted earlier that  $\mu_A * \mu_A$  has 1-norm equal to 1, so we can apply Propositions 4 and 2 in that order to get

$$\|\mu_A * \mu_A - 1\|_p^p = (\widehat{\mu_A * \mu_A}(1 - \mathbf{1}_{\chi_0}))^{*p}(\chi_0) = (\widehat{\mu_A}^2(1 - \mathbf{1}_{\chi_0}))^{*p}(\chi_0).$$

Repeating this whole process with  $\mu_A \circ \mu_A$  in place of  $\mu_A * \mu_A$  produces the very similar identity

$$\|\mu_A \circ \mu_A - 1\|_p^p = (|\widehat{\mu_A}|^2(1 - \mathbf{1}_{\chi_0}))^{*p}(\chi_0),$$

from which we conclude that

$$\|\mu_A \circ \mu_A - 1\|_p^p \geq \|\mu_A * \mu_A - 1\|_p^p \geq \frac{\epsilon}{2}. \quad \blacksquare$$

This lemma tells us that if  $\langle \mu_A * \mu_A, \mu_C \rangle \geq 1/2$ , then  $\|\mu_A \circ \mu_A - 1\|_p \geq 1/4$  for some  $p = O(\log(1/\gamma))$ . This information can then be fed to the following general lemma.

**Lemma 7** (*Unbalancing of spectrally nonnegative functions*). *Let  $\epsilon \in (0, 1)$  and let  $\nu : G \rightarrow \mathbf{R}_{\geq 0}$  have  $\|\nu\|_1 = 1$  and  $\widehat{\nu} \geq 0$ . If  $f : G \rightarrow \mathbf{R}$  has  $\widehat{f} \geq 0$  and  $\|f\|_{p(\nu)} \geq \epsilon$  for some  $p \geq 1$ , then*

$$\|f + 1\|_{p'(\nu)} \geq 1 + \frac{\epsilon}{2}$$

for some  $p' = O(\epsilon^{-1} \log(\epsilon^{-1})p)$ .

*Proof.* Proposition 5 tells us that  $\|f\|_{p(\nu)}$  is monotonically increasing in  $p$ , so without loss of generality we can pick  $p$  odd and at least 5. As usual, we denote the identity in  $\widehat{G}$  by  $\chi_0$ . Using the Fourier inversion formula and orthogonality of characters as we did in the proof of Proposition 3, we observe that

$$\begin{aligned} \|f\|_{p(\nu)}^p &= \mathbf{E}_{x \in G} \left( \sum_{\chi \in \widehat{G}} \widehat{\nu}(\chi) \chi(x) \right) \left( \sum_{\chi \in \widehat{G}} \widehat{f}(\chi) \chi(x) \right)^p \\ &= \sum_{\chi_1 \in \widehat{G}} \cdots \sum_{\chi_{p+1} \in \widehat{G}} \widehat{f}(\chi_1) \cdots \widehat{f}(\chi_p) \widehat{\nu}(\chi_{p+1}) \mathbf{E}_{x \in G} \chi_1(x) \cdots \chi_{p+1}(x) \\ &= \sum_{\chi_1 \cdots \chi_{p+1} = \chi_0} \widehat{f}(\chi_1) \cdots \widehat{f}(\chi_p) \widehat{\nu}(\chi_{p+1}) \\ &= \widehat{\nu} * \widehat{f}^{*p}(\chi_0). \end{aligned}$$

Let  $P = \{x \in G : f(x) \geq 0\}$  and let  $g(x) = \max\{f(x), 0\}$ . It is easy to see that  $2g(x) = f(x) + |f(x)|$ , so

$$\begin{aligned} 2\langle \mathbf{1}_P, f^p \rangle_\nu &= 2 \mathbf{E}_{x \in G} \nu(x) \mathbf{1}_P(x) f(x)^p \\ &= 2 \mathbf{E}_{x \in G} \nu(x) g(x) f(x)^{p-1} \\ &= \langle 2g, f^{p-1} \rangle_\nu \\ &= \mathbf{E}_{x \in G} \nu(x) f(x)^p + \langle |f|, f^{p-1} \rangle_\nu \\ &= \widehat{\nu} * \widehat{f}^{*p}(\chi_0) + \langle |f|, |f|^{p-1} \rangle_\nu, \end{aligned}$$

where in the last line we used the fact that  $f$  is real-valued as well as evenness of  $p-1$ . Since the Fourier transforms of  $f$  and  $\nu$  are both nonnegative, the first term is nonnegative, so

$$\langle \mathbf{1}_P, f^p \rangle_\nu \geq \frac{\langle |f|, |f|^{p-1} \rangle_\nu}{2} = \frac{\|f\|_{p(\nu)}^p}{2} \geq \frac{\epsilon^p}{2}.$$

Now let  $T = \{x \in P : f(x) \geq 3\epsilon/4\}$ . Then

$$\begin{aligned} \langle \mathbf{1}_T, f^p \rangle_\nu &= \langle \mathbf{1}_P - \mathbf{1}_{P \setminus T}, f^p \rangle_\nu \\ &\geq \langle \mathbf{1}_P, f^p \rangle_\nu - \langle \mathbf{1}_{P \setminus T}, f^p \rangle_\nu \\ &\geq \frac{\epsilon^p}{2} - \mathbf{E}_{x \in G} \mathbf{1}_{P \setminus T}(x) f(x)^p \nu(x) \\ &> \frac{\epsilon^p}{2} - \mathbf{E}_{x \in G} (3\epsilon/4)^p \nu(x) \\ &\geq \frac{\epsilon^p}{4}, \end{aligned}$$

where in the last line we used the fact that  $(3/4)^p \leq (3/4)^5 < 243/1024 < 4$ . From this we deduce

$$\begin{aligned}
\frac{\epsilon^p}{4} &\leq \langle \mathbf{1}_T, f^p \rangle_\nu \\
&= \mathbf{E}_{x \in G} (\nu(x)^{1/2} \mathbf{1}_T(x)) (\nu(x)^{1/2} f(x)^p) \\
&\leq \left( \mathbf{E}_{x \in G} \nu(x) \mathbf{1}_T(x)^2 \right)^{1/2} \left( \mathbf{E}_{x \in G} \nu(x) f(x)^{2p} \right)^{1/2} \\
&= \left( \mathbf{E}_{x \in G} \nu(x) \mathbf{1}_T(x) \right)^{1/2} \left( \mathbf{E}_{x \in G} \nu(x) |f(x)|^{2p} \right)^{1/2} \\
&= \nu(T)^{1/2} \|f\|_{2p(\nu)}^p
\end{aligned}$$

by the Cauchy–Schwarz inequality.

Now if  $\|f + 1\|_{2p(\nu)} > 2$ , then we could take  $p' = 2p$ , so assume that this norm is at most 2. By the triangle inequality, we have

$$\|f\|_{2p(\nu)} \leq \| -1 \|_{2p(\nu)} + \|f + 1\|_{2p(\nu)} \leq 3,$$

hence

$$\nu(T)^{1/2} 3^p \geq \frac{\epsilon^p}{4}.$$

Once again using the fact that  $4 < 1024/243$ , we have  $4^{1/p} < 4/3$  and thus

$$\nu(T) \geq \frac{\epsilon^{2p}}{16 \cdot 3^{2p}} = \left( \frac{\epsilon}{4^{1/p} \cdot 3} \right)^{2p} > \left( \frac{\epsilon}{4} \right)^{2p}.$$

This allows us to bound

$$\begin{aligned}
\|f + 1\|_{p'(\nu)} &= \left( \mathbf{E}_{x \in G} \nu(x) |f(x) + 1|^{p'} \right)^{1/p'} \\
&\geq \left( \mathbf{E}_{x \in G} \nu(x) \mathbf{1}_T(x) |f(x) + 1|^{p'} \right)^{1/p'} \\
&\geq (\nu(T) (1 + 3\epsilon/4)^{p'})^{1/p'} \\
&> \left( \frac{\epsilon}{4} \right)^{2p/p'} \left( 1 + \frac{3}{4}\epsilon \right).
\end{aligned}$$

Now if  $p' \geq (8p/\epsilon) \log(4/\epsilon) = O(\epsilon^{-1} \log(\epsilon^{-1})p)$ , then  $\epsilon/4 \geq (2p/p') \log(4/\epsilon)$  and thus

$$-\frac{2p}{p'} \log\left(\frac{4}{\epsilon}\right) \geq -\frac{\epsilon}{4}.$$

Taking  $e$  to the power of both sides gives us

$$\left( \frac{4}{\epsilon} \right)^{-2p/p'} \geq e^{-\epsilon/4} \geq 1 - \frac{\epsilon}{4},$$

and plugging this in above gives the bound

$$\|f + 1\|_{p'(\nu)} > \left(1 - \frac{\epsilon}{4}\right) \left(1 + \frac{3}{4}\epsilon\right) = 1 + \frac{\epsilon}{2} - \frac{3\epsilon^2}{16} \geq 1 + \frac{\epsilon}{2},$$

which is what we needed.  $\blacksquare$

### 3. Dependent random choice

The next lemma uses a dependent random choice argument to pass the information from the previous step down to high density subsets, which allows us to iterate the argument.

**Lemma 8** (*Dependent random choice*). *Let  $G$  be a finite abelian group and let  $A$  be a subset of  $G$  with density  $\alpha$ . Let  $B_1, B_2 \subseteq G$  and  $\mu = \mu_{B_1} \circ \mu_{B_2}$ . For any function  $f : G \rightarrow \mathbf{R}_{\geq 0}$  there exist sets  $A_1 \subseteq B_1$  and  $A_2 \subseteq B_2$  with densities satisfying*

$$\min\{\mu_{B_1}(A_1), \mu_{B_2}(A_2)\} \geq \frac{1}{4}\alpha^{2p}\|\mu_A \circ \mu_A\|_{p(\mu)}^{2p}.$$

and such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, f \rangle \leq 2 \frac{\langle (\mu_A \circ \mu_A)^p, f \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p(\mu)}^p}.$$

*Proof.* For  $s = (s_1, \dots, s_p) \in G^p$  let  $A_1(s) = B_1 \cap (A + s_1) \cap \dots \cap (A + s_p)$ , and define  $A_2(s)$  analogously. First we expand

$$\begin{aligned} \langle (\mu_A \circ \mu_A)^p, f \rangle_\mu &= \mathbf{E}_{x \in G} \mu(x) (\mu_A \circ \mu_A)(x)^p f(x) \\ &= \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} \mu_{B_1}(y) \mu_{B_2}(x + y) (\mu_A \circ \mu_A)(x)^p f(x) \\ &= \frac{1}{|B_1| \cdot |B_2|} \sum_{x \in G} \sum_{y \in G} \mathbf{1}_{B_1}(y) \mathbf{1}_{B_2}(x + y) (\mu_A \circ \mu_A)(x)^p f(x). \end{aligned}$$

Renaming  $b_1 = y$  and performing the change of variable  $b_2 = x + b_1 = x + y$ , we have

$$\begin{aligned} \langle (\mu_A \circ \mu_A)^p, f \rangle_\mu &= \frac{1}{|B_1| |B_2|} \sum_{\substack{b_1 \in G \\ b_2 \in G}} \mathbf{1}_{B_1}(b_1) \mathbf{1}_{B_2}(b_2) (\mu_A \circ \mu_A)(b_2 - b_1)^p f(b_2 - b_1) \\ &= \mathbf{E}_{b_1 \in B_1, b_2 \in B_2} (\mu_A \circ \mu_A)(b_2 - b_1)^p f(b_2 - b_1) \\ &= \mathbf{E}_{b_1 \in B_1, b_2 \in B_2} \left( \alpha^{-2} \mathbf{E}_{y \in G} \mathbf{1}_A(y) \mathbf{1}_A(b_2 - b_1 + y) \right)^p f(b_2 - b_1). \end{aligned}$$

Now since  $y \in A$  if and only if  $b_1 \in A + b_1 - y$  and  $b_2 - b_1 + y \in A$  if and only if  $b_2 \in A + b_1 - y$ , so writing  $t = b_1 - y$  and changing variables, we have

$$\begin{aligned} \langle (\mu_A \circ \mu_A)^p, f \rangle_\mu &= \mathbf{E}_{b_1 \in B_1, b_2 \in B_2} \left( \alpha^{-2} \mathbf{E}_{t \in G} \mathbf{1}_{A+t}(b_1) \mathbf{1}_{A+t}(b_2) \right)^p f(b_2 - b_1) \\ &= \alpha^{-2p} \mathbf{E}_{b_1 \in B_1, b_2 \in B_2} \mathbf{E}_{s \in G^p} \mathbf{1}_{A_1(s)}(b_1) \mathbf{1}_{A_2(s)}(b_2) f(b_2 - b_1). \end{aligned}$$



Putting  $y = b_2 - b_1$  so that  $b_2 = y + b_1$ , we have

$$\begin{aligned} \langle (\mu_A \circ \mu_A)^p, f \rangle_\mu &= \alpha^{-2p} \mathbf{E}_{s \in G^p} \mathbf{E}_{b_1 \in B_1} \frac{|G|}{|B_2|} \mathbf{E}_{y \in G} \mathbf{1}_{A_1(s)}(b_1) \mathbf{1}_{A_2(s)}(y + b_1) f(y) \\ &= \frac{|G|}{\alpha^{2p} |B_2|} \mathbf{E}_{s \in G^p} \mathbf{E}_{b_1 \in B_1} \langle \mathbf{1}_{A_1(s)} \circ \mathbf{1}_{A_2(s)}, f \rangle \\ &= \frac{|G|^2}{\alpha^{2p} |B_1| \cdot |B_2|} \mathbf{E}_{s \in G^p} \langle \mathbf{1}_{A_1(s)} \circ \mathbf{1}_{A_2(s)}, f \rangle. \end{aligned}$$

Thus we let  $\beta_i = |B_i|/|G|$  and  $\alpha_i(s) = \frac{|A_i(s)|}{|G|}$  for  $i \in \{1, 2\}$ , and then apply the above in the case where  $f$  is the constant function 1 to obtain

$$\begin{aligned} \|\mu_A \circ \mu_A\|_{p(\mu)}^p &= \frac{|G|^2}{\alpha^{2p} |B_1| \cdot |B_2|} \mathbf{E}_{s \in G^p} \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} \mathbf{1}_{A_1(s)}(y) \mathbf{1}_{A_2(s)}(x + y) \\ &= \frac{1}{\alpha^{2p} \beta_1 \beta_2} \mathbf{E}_{s \in G^p} \mathbf{E}_{y \in G} \mathbf{1}_{A_1(s)}(y) \mathbf{E}_{x \in G} \mathbf{1}_{A_2(s)}(x + y) \\ &= \frac{1}{\alpha^{2p} \beta_1 \beta_2} \mathbf{E}_{s \in G^p} \alpha_1(s) \alpha_2(s) \end{aligned}$$

The constants out front do not depend on  $f$ , so we see that

$$\frac{\langle (\mu_A \circ \mu_A)^p, f \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p(\mu)}^p} = \frac{\mathbf{E}_{s \in G^p} \langle \mathbf{1}_{A_1(s)} \circ \mathbf{1}_{A_2(s)}, f \rangle}{\mathbf{E}_{s \in G^p} \alpha_1(s) \alpha_2(s)};$$

call this quotient  $\eta$  for brevity. Now we consider the quantity

$$\mathbf{E}_{s \in G^p} \mathbf{E}_{x \in G} \mathbf{1}_{A_1(s)}(x).$$

For a given  $s = (s_1, \dots, s_p) \in G^p$  and  $x \in G$ , the corresponding  $\mathbf{1}_{A_1(s)}(x)$  term is 1 if and only if  $x \in B_1$  and  $x - s_i \in A$  for all  $1 \leq i \leq p$ . Hence we have

$$\mathbf{E}_{s \in G^p} \mathbf{E}_{x \in G} \mathbf{1}_{A_1(s)}(x) = \frac{|B_1| \cdot |A|^p}{|G|^{p+1}} = \alpha^p \beta_1.$$

The analogous identity holds for  $A_2(s)$ . So, letting

$$M = \frac{1}{2} \alpha^p (\beta_1 \beta_2)^{1/2} \|\mu_A \circ \mu_A\|_{p(\mu)}^p,$$

we have

$$\begin{aligned} \mathbf{E}_{s \in G^p} \mathbf{1}_{[\alpha_1(s) \alpha_2(s) < M^2]} \alpha_1(s) \alpha_2(s) &< \mathbf{E}_{s \in G^p} M \sqrt{\alpha_1(s) \alpha_2(s)} \\ &\leq \left( \mathbf{E}_{s \in G^p} M \alpha_1(s) \right)^{1/2} \left( \mathbf{E}_{s \in G^p} M \alpha_2(s) \right)^{1/2} \\ &= M \left( \mathbf{E}_{s \in G^p} \mathbf{E}_{x \in G} \mathbf{1}_{A_1(s)}(x) \right)^{1/2} \\ &\quad \left( \mathbf{E}_{s \in G^p} \mathbf{E}_{x \in G} \mathbf{1}_{A_2(s)}(x) \right)^{1/2} \\ &= M \alpha^p \sqrt{\beta_1 \beta_2} \\ &= \frac{1}{2} \alpha^{2p} \beta_1 \beta_2 \|\mu_A \circ \mu_A\|_{p(\mu)}^p \\ &= \frac{1}{2} \mathbf{E}_{s \in G^p} \alpha_1(s) \alpha_2(s) \end{aligned}$$

and consequently

$$\mathbf{E}_{s \in G^p} \mathbf{1}_{[\alpha_1(s)\alpha_2(s) \geq M^2]} \alpha_1(s)\alpha_2(s) > \frac{1}{2} \mathbf{E}_{s \in G^p} \alpha_1(s)\alpha_2(s).$$

So we have

$$\begin{aligned} \mathbf{E}_{s \in G^p} \langle \mathbf{1}_{A_1(s)} \circ \mathbf{1}_{A_2(s)}, f \rangle &= \eta \mathbf{E}_{s \in G^p} \alpha_1(s)\alpha_2(s) \\ &< 2\eta \mathbf{E}_{s \in G^p} \alpha_1(s)\alpha_2(s) \mathbf{1}_{[\alpha_1(s)\alpha_2(s) \geq M^2]}, \end{aligned}$$

and thus there must be some  $s$  such that

$$\langle \mathbf{1}_{A_1(s)} \circ \mathbf{1}_{A_2(s)}, f \rangle < 2\eta \alpha_1(s)\alpha_2(s) \mathbf{1}_{[\alpha_1(s)\alpha_2(s) \geq M^2]}.$$

Since  $f(x) \geq 0$  for all  $x$ , the left-hand side is nonnegative, meaning that the right-hand side cannot be 0. Thus such an  $s$  must satisfy  $\alpha_1(s)\alpha_2(s) \geq M^2$ . Letting  $A_1 = A_1(s)$  and  $A_2 = A_2(s)$  for this particular  $s$ , we have

$$\frac{|A_1| \cdot |A_2|}{|G|^2} \geq \frac{1}{4} \alpha^{2p} \frac{|B_1| \cdot |B_2|}{|G|^2} \|\mu_A \circ \mu_A\|_{p(\mu)}^{2p},$$

whence

$$\mu_{B_1}(A_1)\mu_{B_2}(A_2) \geq \frac{1}{4} \alpha^{2p} \|\mu_A \circ \mu_A\|_{p(\mu)}^{2p},$$

so neither  $\mu_{B_1}(A_1)$  nor  $\mu_{B_2}(A_2)$  can be less than the right-hand side.

On the other hand, letting  $\alpha_1 = \alpha_1(s)$  and  $\alpha_2 = \alpha_2(s)$ , we also have

$$\begin{aligned} \langle \mu_{A_1} \circ \mu_{A_2}, f \rangle &= \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} \mu_{A_1}(y) \mu_{A_2}(x+y) f(x) \\ &= \alpha_1^{-1} \alpha_2^{-1} \mathbf{E}_{x \in G} \mathbf{E}_{y \in G} \mathbf{1}_{A_1}(y) \mathbf{1}_{A_2}(x+y) f(x) \\ &= \alpha_1^{-1} \alpha_2^{-1} \langle \mathbf{1}_{A_1} \circ \mathbf{1}_{A_2}, f \rangle \\ &< 2\eta \\ &= 2 \frac{\langle (\mu_A \circ \mu_A)^p, f \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p(\mu)}^p}, \end{aligned}$$

which proves the lemma.  $\blacksquare$

This lemma is slightly more general than we shall require. The version that suffices for all our applications is the following.

**Lemma 9.** *Let  $G$  be a finite abelian group, let  $p \geq 1$  be an integer, and let  $\epsilon, \delta > 0$ . Let  $B_1$  and  $B_2$  be subsets of  $G$  and let  $\mu = \mu_{B_1} \circ \mu_{B_2}$ . If  $A \subseteq G$  has density  $\alpha$  and*

$$S = \{x \in G : (\mu_A \circ \mu_A)(x) > (1 - \epsilon) \|\mu_A \circ \mu_A\|_{p(\mu)}\},$$

then there exist  $A_1 \subseteq B_1$  and  $A_2 \subseteq B_2$  with densities satisfying

$$\min\{\mu_{B_1}(A_1), \mu_{B_2}(A_2)\} = \Omega((\alpha \|\mu_A \circ \mu_A\|_{p(\mu)})^{2p+O(\epsilon^{-1} \log(\delta^{-1}))}).$$

such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_S \rangle \geq 1 - \delta.$$

*Proof.* Let  $p'$  be the smallest even integer at least  $p + \epsilon^{-1} \log(\delta^{-1})$ . By the previous lemma applied to the function  $\mathbf{1}_{G \setminus S}$ , there exist sets  $A_1 \subseteq B_1$  and  $A_2 \subseteq B_2$  with densities satisfying

$$\begin{aligned} \min\{\mu_{B_1}(A_1), \mu_{B_2}(A_2)\} &\geq \frac{1}{4} \alpha^{2p'} \|\mu_A \circ \mu_A\|_{p'(\mu)}^{2p'} \\ &\geq \frac{1}{4} (\alpha \|\mu_A \circ \mu_A\|_{p(\mu)})^{2p+2\epsilon^{-1} \log(\delta^{-1})+O(1)} \\ &= \Omega((\alpha \|\mu_A \circ \mu_A\|_{p(\mu)})^{2p+O(\epsilon^{-1} \log(\delta^{-1}))}) \end{aligned}$$

such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_{G \setminus S} \rangle \leq \frac{\langle (\mu_A \circ \mu_A)^{p'}, \mathbf{1}_{G \setminus S} \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p'(\mu)}^{p'}}.$$

Our construction of  $S$  ensures that

$$\frac{\langle (\mu_A \circ \mu_A)^{p'}, \mathbf{1}_{G \setminus S} \rangle_\mu}{\|\mu_A \circ \mu_A\|_{p'(\mu)}^{p'}} \leq (1 - \epsilon)^{p'},$$

and since  $p' \geq \epsilon^{-1} \log(\delta^{-1})$ , we have

$$(1 - \epsilon)^{p'} \leq e^{-\epsilon p'} \leq \delta.$$

Putting everything together, we have

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_{G \setminus S} \rangle \leq \delta,$$

so that

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_S \rangle = 1 - \langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_{G \setminus S} \rangle \geq 1 - \delta,$$

which completes the proof.  $\blacksquare$

#### 4. The finite-field case

We now use the methods of Kelley and Meka to give upper bounds on the size of a subset of  $\mathbf{F}_q^n$  without any three-term arithmetic progressions. First, we restate the dependent random choice lemma in the special case that applies to this finite field context.

**Corollary 10.** *Let  $p \geq 1$  be an integer and  $\epsilon \in (0, 1/2]$ . If  $A \subseteq G$  is such that  $\|\mu_A \circ \mu_A\|_p \geq 1 + \epsilon$  and  $S = \{x \in G : (\mu_A \circ \mu_A)(x) > 1 + \epsilon/2\}$ , then there are subsets  $A_1$  and  $A_2$  of  $G$ , each of density  $\Omega(\alpha^{2p+O(\epsilon^{-1} \log(\epsilon^{-1}))})$ , such that*

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_S \rangle \geq 1 - \epsilon/8.$$

*Proof.* Let  $S' = \{x \in G : (\mu_A \circ \mu_A)(x) > (1 + \epsilon)\|\mu_A \circ \mu_A\|_p\}$ , and apply Lemma 9 with the same  $p$  and  $\epsilon$ , but  $\delta$  set to  $\epsilon/8$ ,  $S$  set to  $S'$ , and  $B_1 = B_2 = G$  so that  $\mu = \mu_{B_1} = \mu_{B_2}$  is the uniform measure on  $G$ . Hence the sets  $A_1$  and  $A_2$  given by the lemma will each have density

$$\Omega((\alpha(1 + \epsilon))^{2p+O(\epsilon^{-1} \log(8/\epsilon))}) = \Omega(\alpha^{2p+O(\epsilon^{-1} \log(\epsilon^{-1}))})$$

in  $G$ , and since

$$(1 - \epsilon)\|\mu_A \circ \mu_A\|_p \geq (1 - \epsilon)(1 + \epsilon) = 1 - \epsilon^2 \geq 1 - \epsilon/2,$$

we have  $S' \subseteq S$  and thus

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_S \rangle \geq \langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_{S'} \rangle \geq 1 - \epsilon/8. \quad \blacksquare$$

There is a theorem that we need which, for now, we shall just state without proof. (This is Theorem 3.2 of [3].)

**Theorem 11** (*Schoen–Sisask*, 2016). *Let  $\epsilon \in (0, 1)$ , let  $S \subseteq \mathbf{F}_q^n$ , and let  $A_1, A_2 \subseteq \mathbf{F}_q^n$  be subsets of relative density at least  $\alpha$ . There is a subspace  $V$  of codimension  $O(\epsilon^{-2} \log(\epsilon^{-1} \alpha^{-1})^2 \log(\alpha^{-1})^2)$  such that*

$$|\langle \mu_V * \mu_{A_1} * \mu_{A_2}, \mathbf{1}_S \rangle - \langle \mu_{A_1} * \mu_{A_2}, \mathbf{1}_S \rangle| \leq \epsilon.$$

The following lemma encapsulates the density increment argument that underlies the proof of Roth's theorem.

**Lemma 12** (*Density increment*). *Let  $\epsilon \in (0, 1)$  and let  $A$  and  $C$  be subsets of  $G = \mathbf{F}_q^n$  with relative densities  $\alpha$  and  $\gamma$ , respectively. Then either*

- i)  $|\langle \mu_A * \mu_A, \mu_C \rangle - 1| \leq \epsilon$ ; or
- ii) *there is a subspace  $V$  of codimension*

$$O(\epsilon^{-2} (\log(1/\gamma) + \epsilon^{-1} \log(\epsilon^{-1}))^4 \log(1/\alpha)^4)$$

*such that  $\max_{x \in G} (\mathbf{1}_A * \mu_V)(x) \geq (1 + \epsilon/32)\alpha$ .*

*Proof.* Suppose that (i) fails. Then by Lemma 6 there is some  $p = O(\log(1/\gamma))$  such that  $\|\mu_A \circ \mu_A - 1\|_p \geq \epsilon/2$ . By the second convolution law (part (ii) of

Proposition 2),  $\mu_A \circ \mu_A$  is a nonnegative function, and note also that if  $\nu$  is the uniform measure on  $\mathbf{F}_q^n$ , then for any  $\chi : \widehat{\mathbf{F}_q^n} \rightarrow \mathbf{C}$ ,

$$\widehat{\nu}(\chi) = \mathbf{E}_{x \in G} \nu(x) \chi(-x) = \begin{cases} q^{-n}, & \text{if } \chi \text{ is the trivial character;} \\ 0, & \text{otherwise.} \end{cases}$$

This implies in particular that  $\widehat{\nu} \geq 0$ , so by Lemma 7 applied with  $f = \mu_A \circ \mu_A$  and the uniform measure for  $\nu$ , we find that  $\|\mu_A \circ \mu_A\|_{p'} \geq 1 + \epsilon/4$  for some  $p' = O((2/\epsilon) \log(2/\epsilon)p) = O(\epsilon^{-1} \log(\epsilon^{-1}) \log(1/\gamma))$ . Let  $C(\epsilon) = \epsilon^{-1} \log(\epsilon^{-1})$ . By Corollary 10, there are sets  $A_1, A_2 \subseteq G$ , each of density  $\Omega(\alpha^{2p'} + O(C(\epsilon)))$ , such that

$$\langle \mu_{A_1} \circ \mu_{A_2}, \mathbf{1}_S \rangle \geq 1 - \epsilon/32,$$

where  $S = \{x \in G : (\mu_A \circ \mu_A)(x) \geq 1 + \epsilon/8\}$ . Feeding  $-A_1$ ,  $A_2$ , and  $S$  into Theorem 11 with  $\epsilon/32$  gives us a subspace  $V$  of codimension

$$\begin{aligned} & O\left(\epsilon^{-2} \log(\epsilon^{-1} \alpha^{-2p'} - O(C(\epsilon)))^2 \log(\alpha^{-2p'} - O(C(\epsilon)))^2\right) \\ &= O\left(\epsilon^{-2} (2p' + C(\epsilon))^2 \log(1/\alpha)^2 ((2p' + C(\epsilon)) \log(1/\alpha) + \log(\epsilon^{-1}))^2\right) \\ &= O\left(\epsilon^{-2} (\log(1/\gamma) + C(\epsilon))^4 \log(1/\alpha)^4\right) \end{aligned}$$

such that

$$|\langle \mu_V * \mu_{-A_1} * \mu_{A_2}, \mathbf{1}_S \rangle - \langle \mu_{-A_1} * \mu_{A_2}, \mathbf{1}_S \rangle| \leq \epsilon/32.$$

It is easily checked that  $\mu_{-A_1} * \mu_{A_2} = \mu_{A_1} \circ \mu_{A_2}$ , so we find that

$$\langle \mu_V * (\mu_{A_1} \circ \mu_{A_2}), \mathbf{1}_S \rangle \geq 1 - \epsilon/16.$$

Now we observe that

$$\|(\mu_{A_1} \circ \mu_{A_2}) \circ \mu_{A_2}\|_1 = \mathbf{E}_{z \in G} \mu_{A_1}(z) \mathbf{E}_{y \in G} \mu_{A_2}(y + z) \mathbf{E}_{x \in G} \mu_A(x + y) = 1,$$

so that,

$$\begin{aligned} \max_{x \in G} (\mu_V * \mathbf{1}_A)(x) &= \alpha \max_{x \in G} (\mu_V * \mu_A)(x) \\ &= \alpha \|(\mu_{A_1} \circ \mu_{A_2}) \circ \mu_A\|_1 \max_{x \in G} (\mu_V * \mu_A)(x) \\ &\geq \alpha \langle \mu_V * \mu_A, (\mu_{A_1} \circ \mu_{A_2}) \circ \mu_A \rangle \\ &= \alpha \langle \mu_V * \mu_A * (\mu_{A_1} \circ \mu_{A_2}), \mu_A \rangle \\ &= \alpha \langle \mu_V * (\mu_{A_1} \circ \mu_{A_2}), \mu_A \circ \mu_A \rangle, \end{aligned}$$

where in the last two lines we have employed the adjoint property  $\langle f, g * h \rangle = \langle h \circ f, g \rangle$ , as well as the commutative properties  $f * g = g * f$  and  $\langle f, g \rangle = \langle g, f \rangle$ ,

all of which hold for functions  $f, g, h : G \rightarrow \mathbf{R}$ . But by the construction of  $S$ , we have  $\mathbf{1}_S(x)(\mu_A \circ \mu_A)(x) \geq (1 + \epsilon/8) \mathbf{1}_S$ , so

$$\begin{aligned} \langle \mu_V * (\mu_{A_1} \circ \mu_{A_2}), \mu_A \circ \mu_A \rangle &\geq \langle \mu_V * (\mu_{A_1} \circ \mu_{A_2}), \mathbf{1}_S(\mu_A \circ \mu_A) \rangle \\ &\geq (1 + \epsilon/8) \langle \mu_V * (\mu_{A_1} \circ \mu_{A_2}), \mathbf{1}_S \rangle \\ &\geq (1 + \epsilon/8)(1 - \epsilon/16) \\ &\geq 1 + \epsilon/32, \end{aligned}$$

hence we conclude that

$$\max_{x \in G} (\mathbf{1}_A * \mu_V)(x) \geq (1 + \epsilon/32) \alpha \quad \blacksquare$$

**Theorem 13** (*Finite field*). *Let  $q$  be a power of an odd prime and let  $A$  be a subset of  $G = \mathbf{F}_q^n$  of cardinality  $\alpha q^n$ . The number of (possibly trivial) three-term arithmetic progressions contained in  $A$  is at least*

$$\frac{\alpha^3}{2} q^{2n - O(\log(1/\alpha)^9)}.$$

Hence if  $A \subseteq \mathbf{F}_q^n$  contains no nontrivial three-term arithmetic progressions, then  $\alpha \leq q^{-\Omega(n^{1/9})}$ .

*Proof.* Let  $C = 2 \cdot A = \{2a : a \in A\}$ , so that  $\gamma = |C|/q^n = \alpha$ . By Lemma 12 applied to  $A$  and  $C$  with parameter  $\epsilon = 1/2$ , we find that either  $\langle \mu_A * \mu_A, \mu_C \rangle \geq 1/2$  or there is a subspace  $V$  of codimension  $O(\log(1/\alpha)^8)$  such that  $\max_{x \in G} (\mathbf{1}_A * \mu_V)(x) \geq (1 + \epsilon/64) \alpha$ .

In the second case, there exists some  $x \in G$  such that

$$\mathbf{E}_{y \in G} \mathbf{1}_A(y) \mu_V(x - y) \geq (1 + \epsilon/64) \alpha.$$

But  $x - y \in V$  if and only if  $y - x \in V$  if and only if  $y \in V + x$ , so we find that

$$|A \cap V + x| \geq (1 + \epsilon/64) \alpha |V|.$$

Now  $A$  has exactly the same number of three-term arithmetic progressions as  $A - x$ , so we can invoke Lemma 12 again with  $V$  in place of  $G$  and  $A - x$  in place of  $A$ , but note that  $\alpha$  has been replaced by  $(1 + \epsilon/64) \alpha$ , so this iteration can only happen  $\log_{1+\epsilon/64}(1/\alpha) = O(\log(1/\alpha))$  times before the second case of the lemma becomes impossible, since  $\alpha \leq 1$ . Hence we deduce that there is some subspace  $V$  of codimension  $O(\log(1/\alpha)^9)$  and some translate  $A + x'$  of  $A$  such that

$$\mathbf{E}_{x \in V} \mathbf{E}_{y \in V} \mu_{A'}(y) \mu_{A'}(x - y) \mu_{2 \cdot A'}(x) \geq \frac{1}{2},$$

where  $A' = (A + x) \cap V$  and the relative densities are taken with respect to the subspace  $V$ . Expanding further, this implies that

$$\frac{|V|^3}{|A'|^3 |V|^2} \sum_{x \in V} \sum_{y \in V} \mathbf{1}_{A'}(y) \mathbf{1}_{A'}(x - y) \mathbf{1}_{2 \cdot A'}(x) \geq \frac{1}{2};$$

that is,

$$|\{(x, y) \in (2 \cdot A) \times A : x - y \in A'\}| \geq \frac{|A'|^3 |V|^5}{2|V|^3} \geq \frac{\alpha^3}{2} q^{2n - O(\log(1/\alpha)^9)}.$$

Renaming variables, this counts the number of pairs  $(x, z) \in A' \times A'$  such that  $x + z = 2y$  for some  $y \in A'$ . Since this equation implies that  $z - y = y - x$ , the above expression counts the number of three-term arithmetic progressions in  $A'$ , including the  $|A'|$  trivial instances of  $x = y = z$ . This proves the first part of the theorem.

For the last part of the theorem statement, suppose that  $A$  does not contain any nontrivial three-term arithmetic progressions. Then

$$\frac{\alpha^3}{2} q^{2n - O(\log(1/\alpha)^9)} \leq |\{(x, y) \in (2 \cdot A) \times A : x - y \in A'\}| \leq |A'| \leq \alpha q^n,$$

whence

$$q^n \leq \frac{2q^{O(\log(1/\alpha)^9)}}{\alpha^2},$$

and taking  $q$ th logs of both sides yields

$$n \leq \log_q \left( \frac{2q^{O(\log(1/\alpha)^9)}}{\alpha^2} \right) = O(\log(1/\alpha)^9) = O(\log_q(1/\alpha)^9).$$

Letting  $C$  be the constant implicit in the last big- $O$  bound, we invert this to obtain

$$\alpha \leq q^{-n^{1/9}/C^{1/9}} = q^{-\Omega(n^{1/9})},$$

which is what we wanted.  $\blacksquare$

## 5. Bohr sets

To transfer the ideas of the finite-field proof over to the integer case, we will need the machinery of Bohr sets. These are, in some sense, the analogue in general abelian groups to subspaces in  $\mathbf{F}_q^n$ .

Let  $G$  be a finite abelian group,  $\Gamma$  be a nonempty subset of  $\widehat{G}$ , and let  $\nu : \Gamma \rightarrow [0, 2]$ . The *Bohr set*  $B = \text{Bohr}_\nu(\Gamma)$  corresponding to this data is the set

$$\text{Bohr}_\nu(\Gamma) = \{x \in G : |1 - \gamma(x)| \leq \nu(\gamma) \text{ for all } \gamma \in \Gamma\}.$$

The set  $\Gamma$  is called the *frequency set* of  $B$ , and  $\nu$  is its *width function*. We shall also say that  $B$  has rank  $d$  if  $|\Gamma| = d$ . Bohr sets contain 0, since  $\gamma(0) = 1$  for all characters  $\gamma$ , and  $x \in B$  if and only if  $-x \in B$ , since  $|1 - \gamma(x)| = |1 - \gamma(-x)|$ . Note that the set  $\text{Bohr}_\nu(\Gamma)$  does not uniquely determine the pair  $(\Gamma, \nu)$ . When we write  $B' \subseteq B$  for Bohr sets  $B'$  and  $B$ , we mean an inclusion of sets, and

do not intend to say anything about the corresponding frequency sets or width functions.

Let  $B = \text{Bohr}_\nu(\Gamma)$  be a Bohr set and let  $\rho > 0$ . Let  $\nu_\rho : \Gamma \rightarrow [0, 2]$  be the width function given by  $\nu_\rho(\gamma) = \max\{\rho \cdot \nu(\gamma), 2\}$ . We define the *dilate* of  $B$  by  $\rho$  to be the Bohr set  $\text{Bohr}_{\rho\nu}(\Gamma)$ . If  $\rho \leq 1$ ,  $\rho\nu(\gamma) \leq \nu$ , meaning that the condition  $|1 - \gamma(x)| \leq \rho\nu(\gamma)$  is now stronger, so  $B_\rho \subseteq B$ , and similarly  $B \subseteq B_\rho$  if  $\rho \geq 1$ .

We now introduce a definition that characterises when a Bohr set is approximately closed under addition. Let  $B$  be a Bohr set of rank  $d$ . We say that  $B$  is *regular* if for all  $\kappa$  with  $|\kappa| \leq 1/(100d)$ ,

$$(1 - 100d|\kappa|)|B| \leq |B_{1+\kappa}| \leq (1 + 100d|\kappa|)|B|.$$

The following proposition will be used frequently in the following sections.

**Proposition 14.** *Let  $G$  be a finite abelian group and let  $k$  be an integer with  $\gcd(k, |G|) = 1$ . If  $B = \text{Bohr}_\nu(\Gamma)$  is a regular Bohr set of rank  $d$ , then  $k \cdot B$  is also a regular Bohr set of rank  $d$ .*

*Proof.* We will construct the frequency set and width function of  $k \cdot B$ . Note first that if we have  $kx = kx'$ , then  $k(x - x') = 0$ . But since  $\gcd(k, |G|) = 1$ ,  $k(x - x') = 0$  implies that  $x - x' = 0$ , so  $x = x'$  and thus the function  $x \mapsto kx$  permutes  $G$ . We will write its inverse as  $x \mapsto x/k$ . For all  $\gamma \in \Gamma$ , let  $\gamma_k$  be given by  $\gamma_k(x) = \gamma(kx)$ , and let  $\gamma_{k^{-1}}(x) = \gamma(x/k)$ . We let

$$\Gamma' = \{\gamma_{k^{-1}} : \gamma \in \Gamma\}$$

and let  $\nu' : \Gamma' \rightarrow [0, 2]$  be given by  $\nu'(\gamma') = \gamma'_k$ . Then we have a bijection  $\gamma \mapsto \gamma'$  such that  $\nu(\gamma) = \nu'(\gamma')$ , so we see that

$$\begin{aligned} k \cdot B &= \{kx \in G : \gamma(x) \leq \nu(\gamma) \text{ for all } \gamma \in \Gamma\} \\ &= \{x \in G : \gamma_{k^{-1}}(x) \leq \nu(\gamma) \text{ for all } \gamma \in \Gamma\} \\ &= \{x \in G : \gamma'(x) \leq \nu'(\gamma') \text{ for all } \gamma' \in \Gamma'\} \\ &= \text{Bohr}_{\nu'}(\Gamma'), \end{aligned}$$

so  $k \cdot B$  is a Bohr set of rank  $d$  as well.

To check regularity, it suffices to show that  $|(k \cdot B)_{1+\kappa}| = |B_{1+\kappa}|$ . We actually have the stronger fact that

$$\begin{aligned} (k \cdot B)_{1+\kappa} &= \{x \in G : \gamma'(x) \leq (1 + \kappa)\nu'(\gamma') \text{ for all } \gamma' \in \Gamma'\} \\ &= \{x \in G : \gamma_{k^{-1}}(x) \leq (1 + \kappa)\nu(\gamma) \text{ for all } \gamma \in \Gamma\} \\ &= \{kx \in G : \gamma(x) \leq (1 + \kappa)\nu(\gamma) \text{ for all } \gamma \in \Gamma\} \\ &= k \cdot B_{1+\kappa}. \quad \blacksquare \end{aligned}$$

## References

- [1] Thomas Bloom and Olof Sisask, “The Kelley–Meka bounds for sets free of three-term arithmetic progressions,” *arXiv preprint 2302.07211* (20 pp), 2023.



- [2] Zander Kelley and Raghu Meka, “Strong bounds for 3-progressions,” *arXiv preprint 2302.05537* (2023), 78 pp.
- [3] Tomasz Schoen and Olof Sisask, “Roth’s theorem for four variables and additive structures in sums of sparse sets,” *Forum of Mathematics, Sigma* **4** (2016), article no. e5.