An exposition of almost periodicity

Yuval Wigderson

1 Introduction

The goal of this note is to prove an almost periodicity result (Theorem 1.1 below), which is quoted without proof in Bloom and Sisask's exposition [1] of the recent Kelley–Meka breakthrough [7] on Roth's theorem. This exposition is meant to be entirely self-contained; thus, when combined with the Bloom–Sisask paper, one obtains a simple, self-contained proof of the Kelly–Meka breakthrough, at least in the finite field setting. As our goal is to get to the proof as quickly as possible, we will not give any background or motivation for why one should care about such a result, nor how it fits into the larger picture; interested readers are strongly encouraged to look at [1].

We first fix some notation. Throughout, $\mathbb{F}$ is a fixed finite field. For a subset $A \subseteq \mathbb{F}^n$, we denote by $\mu(A) = |A|/|\mathbb{F}^n|$ the density of A. We let 1_A be the indicator function of A, and $\mu_A = 1_A/\mu(A)$ the weighted indicator function of A, where the weighting is chosen so that μ_A is a probability measure on $\mathbb{F}^n$. In case $A = \{a\}$, we denote $1_{\{a\}}, \mu_{\{a\}}$ by $1_a, \mu_a$, respectively. Inner products, norms, and convolutions are computed relative to the uniform probability measure on $\mathbb{F}^n$, namely we have

$$\langle f, g\rangle = \mathop{\mathbb{E}}_x[f(x)\overline{g(x)}] \qquad \|f\|_p^p = \mathop{\mathbb{E}}_x[|f(x)|^p] \qquad (f * g)(y) = \mathop{\mathbb{E}}_x[f(x)g(y-x)]$$

for all functions $f, g : \mathbb{F}^n \to \mathbb{C}$ and $p \geq 1$. We note here the useful fact that $(\mu_a * f)(y) = f(y-a)$, i.e. that convolution with μ_a is simply a shift operator. As the ℓ^p norms are shift-invariant, this implies that $\|\mu_a * f\|_p = \|f\|_p$.

We denote the codimension of a subspace $V \subseteq \mathbb{F}^n$ by $\mathrm{codim}(V)$. For a real number $\alpha \in (0, 1]$, we define $\mathcal{L}(\alpha) = \log(2/\alpha)$.

The following result is the almost periodicity theorem stated without proof as [1, Theorem 11]. Versions of this result are also found in [10, Theorem 3.2] and [3, Theorem 7.4]. The technique of almost periodicity goes back to Croot and Sisask [4].

Theorem 1.1. *Fix parameters $\varepsilon, \alpha \in (0, 1]$, and let $A_1, A_2, S \subseteq \mathbb{F}^n$ satisfy $\mu(A_1), \mu(A_2) \geq \alpha$. There is a subspace $V \subseteq \mathbb{F}^n$ such that*

$$\mathrm{codim}(V) \leq C\varepsilon^{-2}\mathcal{L}(\alpha)^4$$

and

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

where $C > 0$ is an absolute constant.

We will deduce Theorem 1.1 from the following slightly more general result.

Theorem 1.2. *Fix parameters $\varepsilon, \alpha, \beta \in (0,1]$ and let $A, B \subseteq \mathbb{F}^n$ satisfy $\mu(A) \geq \alpha, \mu(B) \geq \beta$. Let $f : \mathbb{F}^n \to [0,1]$ be an arbitrary function. There is a subspace $V \subseteq \mathbb{F}^n$ such that*

$$\operatorname{codim}(V) \leq C\varepsilon^{-2}\mathcal{L}(\alpha)^3\mathcal{L}(\beta)$$

and

$$|\langle \mu_V * \mu_A * f, \mu_B \rangle - \langle \mu_A * f, \mu_B \rangle| \leq \varepsilon$$

where $C > 0$ is an absolute constant.

To deduce Theorem 1.1 from Theorem 1.2, we let $A = A_1, B = -A_2, f = 1_{-S}$. The only thing one has to observe is the adjoint property of inner products with respect to convolutions, namely

$$\langle \mu_A * f, \mu_B \rangle = \langle \mu_{A_1} * 1_{-S}, \mu_{-A_2} \rangle = \langle \mu_{A_1} * \mu_{A_2}, 1_S \rangle,$$

and similarly $\langle \mu_V * \mu_A * f, \mu_B \rangle = \langle \mu_V * \mu_{A_1} * \mu_{A_2}, 1_S \rangle$.

We break the proof of Theorem 1.2 into five steps, each of which is discussed in a separate section. We briefly go over what each of these steps is.

1. We first prove that if t is sufficiently large and $a_1, \ldots, a_t$ are chosen uniformly at random from A, then with high probability we can approximate $\mu_A * f$ by $\frac{1}{t}\sum_{i=1}^t (\mu_{a_i} * f)$, where for our purposes "approximate" means that these two functions are close in some ℓ^p norm. Ignoring the quantitative aspects (which are of course crucial for the actual applications), this is little more than the law of large numbers: we know that $\mu_A * f$ is the average of $\mu_a * f$ over all choices of $a \in A$, and the law of large numbers says that such an average is well-approximated with high probability by an empirical average over random choices $a_1, \ldots, a_t$.

2. The argument above works equally well if we replace A by a shifted set $A + x$, for any $x \in \mathbb{F}^n$. Thus, by an averaging argument, we can find a *fixed* choice of $a_1, \ldots, a_t$, such that the *fixed* function $\frac{1}{t}\sum_{i=1}^t (\mu_{a_i} * f)$ approximates $\mu_{A+x} * f$ for many shifts x. Unpacking the definitions, this implies that there is a large set X such that $\mu_x * \mu_A * f$ is close to $\mu_A * f$ for all $x \in X - X$. Here, $X - X = \{y_1 - y_2 : y_1, y_2 \in X\}$ is the difference set of X.

3. A simple application of the triangle inequality allows us to boost the previous result, so that it applies to all $x \in kX - kX$, where k is some unspecified number. Here, $kX - kX$ is the k-fold iterated sumset of $X - X$. The advantage of this boosted result is that iterated sumsets have more additive structure; informally, this means that $kX - kX$ "looks like a subspace" of $\mathbb{F}^n$.

4. So far, we have only related quantities of the form $\mu_x * \mu_A * f$ to $\mu_A * f$, but eventually we would like to understand inner products with μ_B. A simple application of Hölder's inequality allows us to transfer the results from the previous steps to a bound on such inner products.

5. Finally, we apply Fourier analysis (most importantly, a result known as Chang's lemma) to replace the iterated sumset $kX - kX$ with a subspace V. We remark that one could do this step before step 4, and in some sense, that order is more natural. However, all the previous steps work identically in any finite abelian group, whereas this step is necessarily special to $\mathbb{F}^n$ (as it concerns subspaces). By doing this step last, we are able to avoid specializing the group until it is absolutely necessary.

   We remark too that this step *can* be made to work in a general abelian group, by replacing the notion of a subspace with that of a Bohr set. However, in order to keep the technical details as simple as possible, we do not discuss Bohr sets at all.

Most of this exposition draws heavily on Lovett's expository paper [8]. Indeed, essentially all the steps discussed above can be found in Sections 4 and 5 of [8]. However, as his goal is a specific application of almost periodicity, none of the results are stated in precisely the language amenable to the application in [1]. We also mention the exposition of Pham [9], which goes into much greater detail on all these topics; however, all of the results stated there are proved in much greater generality, which introduces some additional complications we have tried to avoid in this exposition.

2 Approximating $\mu_A * f$

Let G be an arbitrary finite abelian group. In this section, we prove that if $f : G \to [0,1]$ is a bounded function and $A \subseteq G$ is any set, then we can with high probability approximate $\mu_A * f$ by $\frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f$, where $a_1, \dots, a_t$ are chosen uniformly at random from A. The following is the precise statement.

Theorem 2.1. *Let $f : G \to [0,1]$ be any function, and let $A \subseteq G$ be any non-empty set*[1]*. Let $\delta \in (0,1)$ and $p \geq 1$ be parameters, and let $t = Cp/\delta^2$, where C is an absolute constant. Then*

$$\Pr_{\substack{a_1,\dots,a_t \in A \\ \text{iid uniform}}} \left[\left\| \mu_A * f - \frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f \right\|_p \leq \delta \right] \geq \frac{1}{2}.$$

In order to prove this, we need the following quantitative version of the law of large numbers (or, more precisely, of the central limit theorem). It follows from the more general Marcinkiewicz–Zygmund inequality, but in our setting it has a simple combinatorial proof.

Lemma 2.2. *Let $X_1, \dots, X_t$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $|X_i| \leq 1$ for all i. There is an absolute constant $C > 0$ such that for any $p \geq 1$, we have*

$$\mathbb{E}\left[\left| \frac{1}{t}\sum_{i=1}^{t} X_i \right|^p \right] \leq \left(\frac{Cp}{t} \right)^{p/2}.$$

[1] Note that for this result, we make no assumption on $\mu(A)$.

Proof. By Jensen's inequality, whenever $p \leq p'$, we have that

$$\mathbb{E}[|X|^p]^{1/p} \leq \mathbb{E}[|X|^{p'}]^{1/p'},$$

for any random variable X. Therefore, up to changing the constant C, it suffices to prove the result when p is an even integer, by rounding p up to an even integer. So we henceforth assume that p is even. In that case, we have that

$$\mathbb{E}\left[\left|\sum_{i=1}^{t} X_i\right|^p\right] = \mathbb{E}\left[\left(\sum_{i=1}^{t} X_i\right)^p\right] = \sum_{i_1,\ldots,i_p \in [t]} \mathbb{E}\left[\prod_{j=1}^{p} X_{i_j}\right].$$

Recall that the X_i are independent, so $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\,\mathbb{E}[X_j] = 0$ whenever $i \neq j$. Additionally, $\mathbb{E}[X_i^m] \leq 1$ for any integer m, as $|X_i| \leq 1$ for all i. Thus, in the final sum, we know that many terms are zero—namely those terms that feature some X_i taken only to the first power. Moreover, every non-zero term is upper-bounded by 1. Thus, it suffices to upper-bound the number of sequences $(i_1, \ldots, i_p) \in [t]^p$ in which every index appears either zero times or at least twice.

To upper-bound this, we first pick the value of i_1 (there are t choices), and then pick some $j \in \{2, \ldots, p\}$ ($p-1$ choices) and assign i_j the same value as i_1. If i_2 has not yet been assigned, we have t choices for it, and then $p-3$ choices for an index to pair it to. Continuing in this way, we conclude that

$$\mathbb{E}\left[\left|\sum_{i=1}^{t} X_i\right|^p\right] \leq t^{p/2}(p-1)(p-3)\cdots 3 \cdot 1 \leq t^{p/2} p^{p/2} = (pt)^{p/2}.$$

Dividing by t^p gives the claimed bound. □

With this result in hand, we can prove Theorem 2.1.

Proof of Theorem 2.1. By Markov's inequality, we have that

$$\begin{aligned}\Pr_{\substack{a_1,\ldots,a_t \in A \\ \text{iid uniform}}}\left[\left\|\mu_A * f - \frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f\right\|_p \leq \delta\right] &= \Pr_{\substack{a_1,\ldots,a_t \in A \\ \text{iid uniform}}}\left[\left\|\mu_A * f - \frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f\right\|_p^p \leq \delta^p\right] \\ &\leq \delta^{-p} \cdot \mathop{\mathbb{E}}_{\substack{a_1,\ldots,a_t \in A \\ \text{iid uniform}}}\left[\left\|\mu_A * f - \frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f\right\|_p^p\right]. \qquad (1)\end{aligned}$$

Define $X_i = \mu_A * f - \mu_{a_i} * f$. In other words, X_i is the random function $G \to \mathbb{R}$ defined by

$$X_i(x) = (\mu_A * f)(x) - f(x - a_i).$$

In particular, we see that $\mathbb{E}[X_i(x)] = 0$ for all x, and that $|X_i(x)| \leq 1$ for all x. Therefore, by Lemma 2.2,

$$\mathop{\mathbb{E}}_{\substack{a_1,\ldots,a_t \in A \\ \text{iid uniform}}}\left[\left\|\mu_A * f - \frac{1}{t}\sum_{i=1}^{t} \mu_{a_i} * f\right\|_p^p\right] = \mathop{\mathbb{E}}_{\substack{a_1,\ldots,a_t \in A \\ \text{iid uniform}}} \mathop{\mathbb{E}}_{x}\left[\left|\frac{1}{t}\sum_{i=1}^{t} X_i(x)\right|^p\right] \leq \left(\frac{Cp}{t}\right)^{p/2}.$$

Plugging in $t = C'p/\delta^2$ for an appropriate constant C' gives the desired result, by (1). □

3 Almost periodicity with an unstructured set

Recall that for a set X, we define $X - X = \{x - x' : x, x' \in X\}$. In this section, we get closer to what we want: we find a large set X such that for every $x \in X - X$, we have that $\mu_x * \mu_A * f$ is close to $\mu_A * f$, where "close" means closeness in some (arbitrary) ℓ^p norm. This is the type of result that was originally referred to as *almost periodicity*. Indeed, it says that for every shift $x \in X - X$, the function $\mu_A * f$ is almost periodic with respect to x, in the sense that $(\mu_A * f)(y) \approx (\mu_A * f)(y - x)$ for an "average" choice of y.

Theorem 3.1. *Let $f : G \to [0, 1]$ be a function, and let $A \subseteq G$ be a set with $\mu(A) \geq \alpha$. Let $\varepsilon \in (0, 1)$ and $p \geq 1$ be parameters. There exists a set X with $\mu(X) \geq \alpha^{Cp/\varepsilon^2}$, where $C > 0$ is an absolute constant, such that*

$$\|\mu_x * \mu_A * f - \mu_A * f\|_p \leq \varepsilon$$

for every $x \in X - X$.

Proof. Let $\delta = \varepsilon/2$, and let $t = Cp/\delta^2$, where C is the constant from Theorem 2.1. Let $S(A) \subseteq A^t$ denote the set of sequences $(a_1, \ldots, a_t) \in A^t$ with the property that

$$\left\| \mu_A * f - \frac{1}{t} \sum_{i=1}^{t} \mu_{a_i} * f \right\|_p \leq \delta.$$

By Theorem 2.1, we know that

$$|S(A)| \geq \frac{1}{2}|A|^t \geq \frac{1}{2}\alpha^t |G|^t.$$

Moreover, by applying Theorem 2.1 to the shifted set $A - x$, we see that $|S(A+x)| \geq \frac{1}{2}\alpha^t |G|^t$ for all $x \in G$. We now apply an averaging argument, as follows. Consider the bipartite graph whose parts are G^t and G, and where a vertex $(a_1, \ldots, a_t) \in G^t$ is adjacent to a vertex $x \in G$ if and only if $(a_1, \ldots, a_t) \in S(A+x)$. By the argument above, every vertex on the right-hand side has degree at least $\frac{1}{2}\alpha^t |G|^t$, and thus the total number of edges is at least $\frac{1}{2}\alpha^t |G|^t \cdot |G|$. Therefore, there is a vertex $(a_1, \ldots, a_t)$ on the left-hand side whose degree is at least $\frac{1}{2}\alpha^t |G|$.

Let X be the set of neighbors of this fixed vertex. Concretely, this means that $X \subseteq G$ is a set of size at least $\frac{1}{2}\alpha^t |G|$, with the property that $(a_1, \ldots, a_t) \in S(A+x)$ for all $x \in X$. The first property is equivalent to saying that $\mu(X) \geq \frac{1}{2}\alpha^t \geq \alpha^{Cp/\varepsilon^2}$ for an appropriate constant C. By the triangle inequality, the second property implies that for all $x, x' \in X$, we have

$$\begin{aligned}
\|\mu_{x-x'} * \mu_A * f - \mu_A * f\|_p &= \|\mu_x * \mu_A * f - \mu_{x'} * \mu_A * f\|_p \\
&= \|\mu_{A+x} * f - \mu_{A+x'} * f\|_p \\
&\leq \left\| \mu_{A+x} * f - \frac{1}{t} \sum_{i=1}^{t} \mu_{a_i} * f \right\|_p + \left\| \mu_{A+x'} * f - \frac{1}{t} \sum_{i=1}^{t} \mu_{a_i} * f \right\|_p \\
&\leq 2\delta = \varepsilon,
\end{aligned}$$

where the first equality uses the fact that the ℓ^p is invariant under convolutions, i.e. that $\|\mu_{x'} * g\|_p = \|g\|_p$ for all x' and all functions g. □

4 Almost periodicity with a large sumset

We now have a powerful ℓ^p almost periodicity result: given a set A and a function f, we find a large set X such that $\mu_x * \mu_A * f \approx \mu_A * f$ for all $x \in X - X$. This is good progress, but eventually we would like to replace $X - X$ by a subspace V (and modify the almost periodicity condition appropriately).

In general, a subset of $\mathbb{F}^n$ might look nothing like a subspace. However, there is a general phenomenon in additive combinatorics, which is that iterated sum and difference sets are (much) more structured than arbitrary sets. For a positive integer k, we define

$$\begin{aligned} kX - kX &= \{x_1 + \cdots + x_k - y_1 - \cdots - y_k : x_1, \ldots, x_k, y_1, \ldots, y_k \in X\} \\ &= \{z_1 + \cdots + z_k : z_1, \ldots, z_k \in X - X\}. \end{aligned}$$

The following result, which follows immediately from Theorem 3.1 by the triangle inequality, shows that we may obtain a version of Theorem 3.1 where $X - X$ is replaced by $kX - kX$. In Section 6, we will then replace $kX - kX$ by a subspace.

Theorem 4.1. *Let $f : G \to [0,1]$ be a function, and let $A \subseteq G$ be a set with $\mu(A) \geq \alpha$. Let $\varepsilon \in (0,1)$, $p \geq 1$, and $k \in \mathbb{N}$ be parameters. There exists a set X with $\mu(X) \geq \alpha^{Ck^2p/\varepsilon^2}$, where $C > 0$ is an absolute constant, such that*

$$\|\mu_x * \mu_A * f - \mu_A * f\|_p \leq \varepsilon$$

for every $x \in kX - kX$.

Proof. Apply Theorem 3.1 with parameter $\varepsilon' = \varepsilon/k$ to obtain a set X with $\mu(X) \geq \alpha^{Cp/(\varepsilon')^2} = \alpha^{Ck^2p/\varepsilon^2}$. Let $x \in kX - kX$. By definition, we may write $x = z_1 + \cdots + z_k$, for some $z_1, \ldots, z_k \in X - X$. By telescoping, the triangle inequality, and the shift invariance of the ℓ^p norm, we find that

$$\begin{aligned} \|\mu_x * \mu_A * f - \mu_A * f\|_p &\leq \sum_{i=1}^{k} \|\mu_{z_1+\cdots+z_i} * \mu_A * f - \mu_{z_1+\cdots+z_{i-1}} * \mu_A * f\|_p \\ &= \sum_{i=1}^{k} \|\mu_{z_i} * \mu_A * f - \mu_A * f\|_p \\ &\leq k\varepsilon' = \varepsilon. \end{aligned}$$

□

5 From ℓ^p to ℓ^∞

Up to now, we have estimated norms of the form $\|\mu_x * \mu_A * f - \mu_A * f\|_p$. However, we are actually interested in estimating inner products of the form

$$\langle \mu_x * \mu_A * f, \mu_B \rangle - \langle \mu_A * f, \mu_B \rangle = \langle \mu_x * \mu_A * f - \mu_A * f, \mu_B \rangle.$$

Conveniently, it is easy to convert information about the norm into information about the inner product, using Hölder's inequality. This is the content of the next result.

Theorem 5.1. *Let $f : G \to [0,1]$ be a function, and let $A, B \subseteq G$ be sets with $\mu(A) \geq \alpha, \mu(B) \geq \beta$. Let $\varepsilon \in (0,1)$ and $k \in \mathbb{N}$ be parameters. There exists a set X with*

$$\mu(X) \geq \exp\left(-C\frac{k^2}{\varepsilon^2}\mathcal{L}(\alpha)\mathcal{L}(\beta)\right),$$

where $C > 0$ is an absolute constant, such that

$$|\langle \mu_x * \mu_A * f, \mu_B\rangle - \langle \mu_A * f, \mu_B\rangle| \leq \varepsilon \tag{2}$$

for every $x \in kX - kX$. In fact, we get the following stronger result: for every $v \in \mathbb{F}^n$ and $x \in kX - kX$,

$$|\langle \mu_v * \mu_x * \mu_A * f, \mu_B\rangle - \langle \mu_v * \mu_A * f, \mu_B\rangle| \leq \varepsilon. \tag{3}$$

Note that (3) really is stronger than (2), as (2) follows from (3) by plugging in $v = 0$.

Proof. Let $p = \mathcal{L}(\beta)$ and let q be the dual norm index of p, defined by $1/p + 1/q = 1$. Apply Theorem 4.1 with this choice of p and with parameter $\varepsilon' = \varepsilon/e$. Let X be the resulting set, and note that the lower bound on $\mu(X)$ is immediately given by plugging in the choice of p and ε' to Theorem 4.1.

Now fix $x \in kX - kX$ and $v \in \mathbb{F}^n$. By Hölder's inequality and the shift-invariance of the ℓ^p norm,

$$|\langle \mu_v * \mu_x * \mu_A * f - \mu_v * \mu_A * f, \mu_B\rangle| \leq \|\mu_v * \mu_x * \mu_A * f - \mu_v * \mu_A * f\|_p \|\mu_B\|_q \leq \frac{\varepsilon}{e}\|\mu_B\|_q.$$

Finally, we have that

$$\|\mu_B\|_q = \left(\mathop{\mathbb{E}}_y[|\mu_B(y)|^q]\right)^{1/q} \leq \left(\beta^{1-q}\right)^{1/q} = \left(\frac{1}{\beta}\right)^{1-1/q} = \left(\frac{1}{\beta}\right)^{1/\mathcal{L}(\beta)} \leq e \qquad \square$$

6 Finding a subspace using Fourier analysis

Our final step is to use Theorem 5.1 to obtain a similar result, but where the iterated sumset $kX - kX$ is replaced by a subspace V. In the literature, this step is usually referred to as *bootstrapping*. Note that, as we are interested in subspaces, we must finally specialize from an arbitrary finite abelian group G to a vector space $\mathbb{F}^n$, where $\mathbb{F}$ is some fixed finite field.

We briefly recall the basics of Fourier analysis on finite field vector spaces. A function $\chi : \mathbb{F}^n \to \mathbb{C}$ is called a *character* if $\chi(x+y) = \chi(x)\chi(y)$ for all $x, y \in \mathbb{F}^n$. The set of characters, denoted $\widehat{\mathbb{F}^n}$, forms another n-dimensional vector space over $\mathbb{F}$. If $\mathbb{F}$ has prime order[2] q, and if we fix an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{F}^n$, then we can explicitly identify $\widehat{\mathbb{F}^n}$ with $\mathbb{F}^n$: each $y \in \mathbb{F}^n$ corresponds to a character $\chi_y : \mathbb{F}^n \to \mathbb{C}$ defined by $\chi_y(x) = e^{2\pi i \langle x,y\rangle/q}$.

Given $f : \mathbb{F}^n \to \mathbb{C}$, its *Fourier transform* $\widehat{f}$ is the function $\widehat{\mathbb{F}^n} \to \mathbb{C}$ defined by

$$\widehat{f}(\chi) = \mathop{\mathbb{E}}_x[f(x)\chi(-x)].$$

[2]This assumption is not necessary, and we include it only for simplicity.

The Fourier transform satisfies the following simple properties, all of which can be checked by expanding the definitions and/or applying the orthogonality of the set of characters.

- Convolution identity: $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$.
- Parseval's identity:
$$\langle f, g \rangle = \sum_{\chi} \widehat{f}(\chi)\overline{\widehat{g}(\chi)}$$
- If f_- is defined by $f_-(x) = f(-x)$, then $\widehat{f_-}(\chi) = \overline{\widehat{f}(\chi)}$. In particular, the convolution identity implies that $\widehat{f * f_-} = |\widehat{f}|^2$.
- For any $A \subseteq \mathbb{F}^n$, we have that $\widehat{\mu_A}(0) = 1$ and $|\widehat{\mu_A}(\chi)| \leq 1$ for all $\chi \in \widehat{\mathbb{F}^n}$.
- Let $V \subseteq \mathbb{F}^n$ be a subspace. Then there exists a subspace $V^\perp \subseteq \widehat{\mathbb{F}^n}$ with $\dim V + \dim V^\perp = n$ such that
$$\widehat{\mu_V}(\chi) = \begin{cases} 1 & \text{if } \chi \in V^\perp \\ 0 & \text{otherwise.} \end{cases}$$
If we fix an inner product on $\mathbb{F}^n$ and use it to identify $\widehat{\mathbb{F}^n}$ with $\mathbb{F}^n$, then $V^\perp$ simply is the orthogonal complement of V with respect to this inner product.
Conversely, given a subspace $W \subseteq \widehat{\mathbb{F}^n}$, there exists a subspace $V \subseteq \mathbb{F}^n$ with $V^\perp = W$.

The final property we will need about the Fourier transform is Chang's lemma; in order to state it, we need the following definition. For a set $X \subseteq \mathbb{F}^n$ and a number $\delta \in (0, 1]$, we define the *δ-spectrum* of X, denoted $\mathrm{Spec}_\delta(X)$, by

$$\mathrm{Spec}_\delta(X) = \{\chi \in \widehat{\mathbb{F}^n} : |\widehat{\mu_X}(\chi)| \geq \delta\} \subseteq \widehat{\mathbb{F}^n}.$$

Using Parseval's identity, we find that $|\mathrm{Spec}_\delta(X)| \leq (\delta^2 \mu(X))^{-1} = O_\delta(1/\mu(X))$. Chang's lemma says that when $\delta = \Theta(1)$, we may place $\mathrm{Spec}_\delta(X)$ inside a subspace that is not much larger than this bound, namely in a subspace of dimension $O_\delta(\log \frac{1}{\mu(X)})$. The precise statement is the following.

Theorem 6.1 (Chang's lemma)**.** *For any $X \subseteq \mathbb{F}^n$ and any $\delta \in (0, 1]$, we have that*

$$\dim(\mathrm{span}(\mathrm{Spec}_\delta(X))) \leq C\delta^{-2} \log \frac{1}{\mu(X)},$$

where $C > 0$ is an absolute constant.

For completeness, we include a proof of Chang's lemma in Appendix A, but for now let us see how it lets us complete the proof of Theorem 1.2.

Proof of Theorem 1.2. Let $k = \log_2(3\alpha^{-1/2}\varepsilon^{-1}) \leq \mathcal{L}(\varepsilon)\mathcal{L}(\alpha)$. We apply Theorem 5.1 with parameters k and $\varepsilon' = \varepsilon/3$ to the given function f and sets A, B. We obtain a set X with

$$\mu(X) \geq \exp\left(-C\frac{\mathcal{L}(\varepsilon)^2}{\varepsilon^2}\mathcal{L}(\alpha)^3\mathcal{L}(\beta)\right),$$

for some absolute constant $C > 0$, with the property that

$$|\langle \mu_v * \mu_x * \mu_A * f, \mu_B\rangle - \langle \mu_v * \mu_A * f, \mu_B\rangle| \leq \frac{\varepsilon}{3} \tag{4}$$

for all $x \in kX - kX$ and all $v \in \mathbb{F}^n$. Let

$$W = \operatorname{span}(\operatorname{Spec}_{\frac{1}{2}}(X)) \subseteq \widehat{\mathbb{F}^n},$$

and let $V \subseteq \mathbb{F}^n$ be the subspace such that $V^\perp = W$. By Theorem 6.1, we see that

$$\operatorname{codim} V = \dim W \leq C' \log\frac{1}{\mu(X)} \leq C''\frac{\mathcal{L}(\varepsilon)^2}{\varepsilon^2}\mathcal{L}(\alpha)^3\mathcal{L}(\beta),$$

for appropriate constants $C', C'' > 0$. Our goal is to prove that this choice of V satisfies the desired condition.

Define $\tau = (\mu_X * \mu_{-X}) * \cdots * (\mu_X * \mu_{-X})$, the k-fold convolution of $\mu_X * \mu_{-X}$ with itself. τ has a very simple probabilistic interpretation: suppose we pick $x_1, \ldots, x_k, y_1, \ldots, y_k \in X$ independently and uniformly at random, and define $Z = x_1 + \cdots + x_k - y_1 - \cdots - y_k$. Then τ is simply the probability distribution of Z. In particular, as Z is an element of $kX - kX$, (4) implies that

$$|\langle \tau * \mu_A * f, \mu_B\rangle - \langle \mu_A * f, \mu_B\rangle| = |\mathbb{E}[\langle \mu_Z * \mu_A * f, \mu_B\rangle - \langle \mu_A * f, \mu_B\rangle]| \leq \frac{\varepsilon}{3}.$$

Additionally, by applying (4) to all elements $v \in V$, we see that

$$|\langle \mu_V * \tau * \mu_A * f, \mu_B\rangle - \langle \mu_V * \mu_A * f, \mu_B\rangle| = \left|\mathop{\mathbb{E}}_{v\in V}[\langle \mu_v * \mu_Z * \mu_A * f, \mu_B\rangle - \langle \mu_v * \mu_A * f, \mu_B\rangle]\right| \leq \frac{\varepsilon}{3}.$$

Therefore, to prove Theorem 1.2, it suffices to prove that

$$|\langle \tau * \mu_A * f, \mu_B\rangle - \langle \mu_V * \tau * \mu_A * f, \mu_B\rangle| \leq \frac{\varepsilon}{3},$$

or equivalently

$$|\langle (\tau - \mu_V * \tau) * \mu_A * f, \mu_B\rangle| \leq \frac{\varepsilon}{3}. \tag{5}$$

To prove (5), we compute the left-hand side in the Fourier domain. First, by the convolution identity, we have that

$$\widehat{\mu_V * \tau}(\chi) = \widehat{\mu_V}(\chi)\widehat{\tau}(\chi) = \begin{cases} \widehat{\tau}(\chi) & \text{if } \chi \in W \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Additionally, by the convolution identity and the definition of τ, we have that $\widehat{\tau}(\chi) = |\widehat{\mu_X}(\chi)|^k$. In particular, since $W \supseteq \mathrm{Spec}_{\frac{1}{2}}(X)$,

$$|\widehat{\tau}(\chi)| \leq 2^{-k} \qquad \text{if } \chi \notin W \tag{7}$$

Now applying Parseval's identity to the left-hand of (5), we find that

$$\begin{aligned}
|\langle (\tau - \mu_V * \tau) * \mu_A * f, \mu_B \rangle| &= \left| \sum_\chi (\widehat{\tau}(\chi) - \widehat{\mu_V * \tau}(\chi)) \widehat{\mu_A}(\chi) \widehat{f}(\chi) \overline{\widehat{\mu_B}(\chi)} \right| \\
&\leq \sum_\chi |\widehat{\tau}(\chi) - \widehat{\mu_V * \tau}(\chi)| |\widehat{\mu_A}(\chi)| |\widehat{f}(\chi)| && [|\widehat{\mu_B}(\chi)| \leq 1] \\
&= \sum_{\chi \notin W} |\widehat{\tau}(\chi)| |\widehat{\mu_A}(\chi)| |\widehat{f}(\chi)| && [\text{by (6)}] \\
&\leq 2^{-k} \sum_\chi |\widehat{\mu_A}(\chi)| |\widehat{f}(\chi)| && [\text{by (7)}] \\
&\leq 2^{-k} \left(\sum_\chi |\widehat{\mu_A}(\chi)|^2 \right)^{1/2} \left(\sum_\chi |\widehat{f}(\chi)|^2 \right)^{1/2} && [\text{Cauchy–Schwarz}] \\
&= 2^{-k} \|\mu_A\|_2 \|f\|_2 && [\text{Parseval}] \\
&\leq 2^{-k} \alpha^{-1/2} && [\|f\|_2 \leq \|f\|_\infty \leq 1] \\
&= \frac{\varepsilon}{3} && [\text{choice of } k].
\end{aligned}$$

This proves (5) and completes the proof. □

A A proof of Chang's lemma

Chang's lemma was first proved in [2, Lemma 3.1], and it quickly became a widely used tool in additive combinatorics. The standard proof works in any finite abelian group (once one sets up the language of Bohr sets appropriately), and is based on an analytic result known as Rudin's inequality. For an exposition of this approach, see e.g. [9, Section 4.1.1] or [11, Section 4.6].

More recently, Impagliazzo, Moore, and Russell [6] found an alternative proof using entropy (see also [5] for a slight reformulation of their technique, which is more similar to the approach we follow below). However, they only proved their result in the case of $\mathbb{F} = \mathbb{F}_2$, and to the best of my knowledge, an entropic proof of Chang's lemma for general finite field vector spaces does not exist in the literature. The proof below aims to fill that gap.

We remark that we will only deal with the case of fields of prime order, but the exact same proof works for all finite fields; one simply has to insert the trace function from $\mathbb{F}$ to its prime field in the correct places.

We begin with a brief review of entropy. For a probability distribution p on some finite set Ω, its *entropy* is defined by

$$\mathrm{H}(p) = \sum_{\omega \in \Omega} p(\omega) \log \frac{1}{p(\omega)},$$

with the convention $0 \log \frac{1}{0} = 0$. Note that if p is the uniform distribution on Ω, then $\mathrm{H}(p) = \log |\Omega|$. One key property of entropy that we will need is *subadditivity*: if p is a probability distribution on $\Omega_1 \times \Omega_2$, and p_1, p_2 denote its marginals on Ω_1, Ω_2, respectively, then $\mathrm{H}(p) \leq \mathrm{H}(p_1) + \mathrm{H}(p_2)$.

The only other property of entropy that we will need is the following fact, which is an immediate consequence of Pinsker's inequality, and which also has an elementary proof.

Lemma A.1. *If p is a probability distribution on a finite set Ω, then*

$$\mathrm{H}(p) \leq \log |\Omega| - \frac{1}{2} \sum_{\omega \in \Omega} \left(p(\omega) - \frac{1}{|\Omega|} \right)^2.$$

Proof for readers who know about Pinsker's inequality. Note that $\log |\Omega| - \mathrm{H}(p)$ is precisely the Kullback–Leibler divergence between p and the uniform distribution on Ω. Applying Pinsker's inequality, rearranging, and bringing the square into the sum yields the claimed bound. □

Proof for readers who know about Jensen's inequality. Define the function $\varphi : (0, 1] \to \mathbb{R}$ by

$$\varphi(x) = x \log \frac{1}{x} + \frac{1}{2} \left(x - \frac{1}{|\Omega|} \right)^2.$$

Note that

$$\varphi''(x) = -\frac{1}{x} + 1,$$

which is non-positive for all $x \in (0, 1]$. Thus, φ is concave on this range. Note that

$$\sum_{\omega \in \Omega} \varphi(p(\omega)) = \mathrm{H}(p) + \frac{1}{2} \sum_{\omega \in \Omega} \left(p(\omega) - \frac{1}{|\Omega|} \right)^2.$$

By Jensen's inequality, this sum is maximized when all $p(\omega)$ are equal to their average, which is $1/|\Omega|$. Thus,

$$\mathrm{H}(p) + \frac{1}{2} \sum_{\omega \in \Omega} \left(p(\omega) - \frac{1}{|\Omega|} \right)^2 = \sum_{\omega \in \Omega} \varphi(p(\omega)) \leq \sum_{\omega \in \Omega} \varphi\left(\frac{1}{|\Omega|} \right) = \sum_{\omega \in \Omega} \frac{1}{|\Omega|} \log |\Omega| = \log |\Omega|. \quad \square$$

We henceforth fix a prime q and let $\mathbb{F}$ be the finite field of order q. We fix an inner product $\langle \cdot\,, \cdot \rangle$ on $\mathbb{F}^n$. As discussed in Section 6, we can now identify an element $y \in \mathbb{F}^n$ with a character χ_y defined by $\chi_y(x) = e^{2\pi i \langle x, y \rangle / q}$. In particular, for $t \in \mathbb{F}$, we denote by $\chi_t : \mathbb{F} \to \mathbb{C}$ the character on the one-dimensional vector space $\mathbb{F}$ defined by $\chi_t(x) = e^{2\pi i t x / q}$.

We next note the following simple Fourier-analytic lemma.

Lemma A.2. *Let $\mathbb{F}$ be a finite field of order q, let $f : \mathbb{F} \to \mathbb{R}$, and let $a = \mathbb{E}[f]$. Then*

$$\sum_{x \in \mathbb{F}} (f(x) - a)^2 = q \cdot \sum_{0 \neq t \in \mathbb{F}} |\widehat{f}(\chi_t)|^2.$$

Proof. We have that

$$\sum_{x \in \mathbb{F}} (f(x) - a)^2 = q \, \|f - a\|_2^2 .$$

Note that $\widehat{f - a}(\chi_0) = 0$, and $\widehat{f - a}(\chi_t) = \widehat{f}(\chi_t)$ for all $0 \neq t \in \mathbb{F}$. Applying Parseval's identity, we see that

$$\|f - a\|_2^2 = \sum_{0 \neq t \in \mathbb{F}} |\widehat{f}(\chi_t)|^2,$$

which implies the desired result. □

We now prove the following result, which immediately implies Chang's lemma, as we will shortly see. In the study of boolean functions, this result is often called the level-1 inequality. Again, I am not aware of such a result appearing in the literature except when $\mathbb{F} = \mathbb{F}_2$.

Theorem A.3. *Fix a basis $e_1, \dots, e_n$ of $\mathbb{F}^n$. For any $X \subseteq \mathbb{F}^n$, we have*

$$\sum_{i=1}^{n} \sum_{0 \neq t \in \mathbb{F}} |\widehat{\mu_X}(\chi_{te_i})|^2 \leq 2q \log \frac{1}{\mu(X)}.$$

Proof. Note that for any $z \in \mathbb{F}^n$, the vector $(\langle z, e_1 \rangle, \dots, \langle z, e_n \rangle)$ uniquely determines z. Therefore, if we let Z be a uniformly random vector of X, the subadditivity of entropy implies that

$$\log |X| = \mathrm{H}(Z) \leq \sum_{i=1}^{n} \mathrm{H}(\langle Z, e_i \rangle).$$

Let p_i be the probability distribution on $\mathbb{F}$ defined by $\langle Z, e_i \rangle$. We now note that for any $t \in \mathbb{F}$,

$$\widehat{\mu_X}(\chi_{te_i}) = \mathop{\mathbb{E}}_{x \in \mathbb{F}^n}[\mu_X(x)\chi_{te_i}(-x)] = \mathop{\mathbb{E}}_{Z} e^{-2\pi i t Z/q} = \sum_{x \in \mathbb{F}} p_i(x) e^{-2\pi i t x/q} = q \cdot \widehat{p_i}(\chi_t),$$

where we view p_i as a function $\mathbb{F} \to \mathbb{C}$, and recall that $\chi_t : \mathbb{F} \to \mathbb{C}$ is the one-dimensional character corresponding to $t \in \mathbb{F}$. By Lemma A.1, we know that

$$\mathrm{H}(p_i) \leq \log q - \frac{1}{2} \sum_{x \in \mathbb{F}} \left(p_i(x) - \frac{1}{q} \right)^2 .$$

Note that when viewed as a function on $\mathbb{F}$, we have that $\mathbb{E}[p_i] = 1/q$ since p is a probability distribution. Therefore, by Lemma A.2,

$$\sum_{x \in \mathbb{F}} \left(p_i(x) - \frac{1}{q} \right)^2 = q \cdot \sum_{0 \neq t \in \mathbb{F}} |\widehat{p_i}(\chi_t)|^2 = \frac{1}{q} \cdot \sum_{0 \neq t \in \mathbb{F}} |\widehat{\mu_X}(\chi_{te_i})|^2.$$

Combining everything, we learn that

$$\log|X| \leq \sum_{i=1}^{n} \mathrm{H}(p_i) \leq n\log q - \frac{1}{2q}\sum_{i=1}^{n}\sum_{0\neq t\in\mathbb{F}} |\widehat{\mu_X}(\chi_{te_i})|^2$$

or equivalently

$$\sum_{i=1}^{n}\sum_{0\neq t\in\mathbb{F}} |\widehat{\mu_X}(\chi_{te_i})|^2 \leq 2q\log\frac{q^n}{|X|} = 2q\log\frac{1}{\mu(X)}. \qquad \square$$

With this result in hand, we can readily deduce Chang's lemma in the form stated in Theorem 6.1, with the constant $C = 2q$.

Proof of Theorem 6.1. Let $d = \dim(\mathrm{span}(\mathrm{Spec}_\delta(X)))$, and let $e_1, \ldots, e_d$ be linearly independent vectors in $\mathrm{Spec}_\delta(X)$. Let $e_{d+1}, \ldots, e_n$ be arbitrary elements of $\mathbb{F}^n$ which complete $e_1, \ldots, e_d$ to a basis of the whole space. By Theorem A.3, we have that

$$2q\log\frac{1}{\mu(X)} \geq \sum_{i=1}^{n}\sum_{0\neq t\in\mathbb{F}} |\widehat{\mu_X}(\chi_{te_i})|^2.$$

Every term in this sum is non-negative, so we may discard all terms with $i > d$ or $t \neq 1$ to find that

$$2q\log\frac{1}{\mu(X)} \geq \sum_{i=1}^{d} |\widehat{\mu_X}(\chi_{e_i})|^2 \geq d\cdot\delta^2,$$

since $e_i \in \mathrm{Spec}_\delta(X)$ for all $1 \leq i \leq d$. Rearranging gives the desired result. $\square$

References

[1] T. F. Bloom and O. Sisask, The Kelley–Meka bounds for sets free of three-term arithmetic progressions, 2023. Preprint available at arXiv:2302.07211.

[2] M.-C. Chang, A polynomial bound in Freiman's theorem, *Duke Math. J.* **113** (2002), 399–419.

[3] E. Croot, I. Łaba, and O. Sisask, Arithmetic progressions in sumsets and L^p-almost-periodicity, *Combin. Probab. Comput.* **22** (2013), 351–365.

[4] E. Croot and O. Sisask, A probabilistic technique for finding almost-periods of convolutions, *Geom. Funct. Anal.* **20** (2010), 1367–1396.

[5] L. Hambardzumyan and Y. Li, Chang's lemma via Pinsker's inequality, *Discrete Math.* **343** (2020), 111496, 3.

[6] R. Impagliazzo, C. Moore, and A. Russell, An entropic proof of Chang's inequality, *SIAM J. Discrete Math.* **28** (2014), 173–176.

[7] Z. Kelley and R. Meka, Strong bounds for 3-progressions, 2023. Preprint available at arXiv:2302.05537.

[8] S. Lovett, An exposition of Sanders' quasi-polynomial Freiman–Ruzsa theorem, *Theory of Computing Library, Graduate Surveys* **6** (2015), 1–14.

[9] H. T. Pham, Almost periodicity and its applications to Roth's theorem, 2019. Available online at `https://web.stanford.edu/~huypham/cam_essay_final.pdf`.

[10] T. Schoen and O. Sisask, Roth's theorem for four variables and additive structures in sums of sparse sets, *Forum Math. Sigma* **4** (2016), Paper No. e5, 28.

[11] T. Tao and V. Vu, *Additive combinatorics*, *Cambridge Studies in Advanced Mathematics*, vol. 105, Cambridge University Press, Cambridge, 2006.