BOURGAIN'S PROOF OF THE EXISTENCE OF LONG ARITHMETIC PROGRESSIONS IN $A + B$

OLOF SISASK

1. Introduction

The primary purpose of this note is to help me understand Bourgain's proof of the following theorem from [1].

Theorem 1.1. *Let N be a prime and suppose that A and B are subsets of $\mathbb{Z}_N$ of densities α and β. Then the sumset $A+B$ contains an arithmetic progression of length at least*

$$\exp\left(c(\alpha\beta\log N)^{1/3} - c\log\log N\right).$$

Bourgain's argument is extremely cunning; it proceeds by an in-depth analysis of sets of Fourier coefficients, introducing several substantial innovations in the process. Another purpose of this note is to highlight the following 'structural' result that can be extracted from Bourgain's paper without much effort. Theorem 1.1 then follows as an easy consequence.

Theorem 1.2 (L^p-almost-periodicity for convolutions)**.** *Let $\epsilon > 0$ and $m \in \mathbb{N}$ be two parameters, and let G be a finite abelian group. Suppose that $f, g : G \to [0,1]$ have averages $\mathbb{E}f = \alpha$ and $\mathbb{E}g = \beta$. Then there is a Bohr set $B = B(\Gamma, \rho)$ of rank $|\Gamma| \ll m^2\log(1/\epsilon)/\epsilon^2$ and radius $\rho = c\epsilon^3/m$ such that*

$$\|f * g(x+t) - f * g(x)\|_{L^{2m}(x)} \leqslant \epsilon(\alpha\beta)^{1/2}$$

for each $t \in B$.

The result says that convolutions are somewhat 'continuous' objects: one can find lots of translates that leave $f * g$ virtually unchanged in L^p; furthermore, there is a lot of structure to the set of translates. For the reader unfamiliar with Bohr sets, we give the definition and the relevant properties in the next section; the general idea to take away is that B can be considered a 'large' and 'additively structured' set.

This note is heavily based on the expository notes [4] of Ben Green, which are well worth studying, and on conversations with Tom Sanders, whose encouragement I am extremely grateful for. I should also mention that one of the motivations for putting this note together in the first place was to contrast it with a non-Fourier-analytic approach to proving Theorem 1.1 due to Ernie Croot and myself [2].

The note is structured as follows. In the next section we collect together the tools that we shall need; most of these are now considered standard in additive combinatorics. We combine these in Section 3 to establish Theorem 1.2, and in Section 4 we show how to deduce Theorem 1.1.

2. Fourier analysis, Bohr sets and dissociativity

Throughout, N will denote the order of the group G.

We shall use the basics of Fourier analysis, as can be found in [5, Chapter 4]. Thus, for a function $f : G \to \mathbb{C}$ and a character $\gamma \in \widehat{G}$—the dual of G—we write

$$\widehat{f}(\gamma) = \mathbb{E}_{x \in G} f(x)\overline{\gamma(x)}.$$

We then have Fourier inversion

$$f(x) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma(x)$$

and Parseval's identity

$$\|\widehat{f}\|_{l^2} = \|f\|_{L^2}.$$

We plainly require the notion of a Bohr set. Again the text [5] may be consulted for more detailed information, including proofs of the statements immediately following the definition.

Definition 2.1. Let $\Gamma \subseteq \widehat{G}$ be a non-empty set of characters and let $\rho > 0$ be a radius. Then we say that

$$B(\Gamma, \rho) = \{t : |\gamma(t) - 1| \leqslant \rho \text{ for all } \gamma \in \Gamma\}$$

is a *Bohr set*. We call $|\Gamma|$ the *rank* of the Bohr set and ρ the *radius*.

Bohr sets thus correspond approximately to the notion of orthogonal complements in vector spaces, but in a setting where there may not be any actual proper subgroups.

Lemma 2.2. *Let $G = \mathbb{Z}_N$ for N a prime and let B be a Bohr set of rank d and radius ρ. Then*

(i) $|B| \geqslant (\rho/2\pi)^d N$ *and*
(ii) *B contains an arithmetic progression of length at least $(\rho/2\pi)N^{1/d}$.*

We shall in fact only require the second conclusion of the lemma, which can be proved by a simple application of Dirichlet's box principle, and only then in deducing Theorem 1.1 from Theorem 1.2.

We shall also require the notion of dissociativity—making use of this was a key insight of Bourgain's.

Definition 2.3. A subset $S \in \widehat{G}$ is said to be *dissociated* if

$$\prod_{\gamma \in S} \gamma^{\sigma_\gamma} = 1 \text{ and } \sigma \in \{-1, 0, 1\}^S \text{ implies that } \sigma = 0.$$

Dissociativity is normally expressed in additive notation,

$$\sum_{\gamma\in S}\sigma_\gamma\gamma = 0 \text{ implies } \sigma = 0,$$

from which it is clearer that it represents a type of independence of the characters in S. A key fact about dissociated sets comes from the following inequality.

Theorem 2.4 (Rudin's inequality)**.** *Suppose that $S \subseteq \widehat{G}$ is a dissociated set of characters and let f be a function whose Fourier transform is supported on S. Then for each $p \geqslant 2$ we have the norm estimate*

$$\|f\|_{L^p} \ll \sqrt{p}\|\widehat{f}\|_{l^2}.$$

We shall use the following lemma to find dissociated sets of characters; see [5, Lemma 4.35].

Lemma 2.5 (Cube covering lemma)**.** *Let $S \subseteq \widehat{G}$ and let $d \geqslant 1$ be an integer. Then there is a partition*

$$S = D_1 \cup \cdots \cup D_k \cup R$$

where each D_i is dissociated and has size d, and R is contained in the $\{-1, 0, 1\}$-span of at most d elements in $\widehat{G}$.

The condition on R here means that there are d elements $\eta_1, \dots, \eta_d \in \widehat{G}$ such that each $\gamma \in R$ can be written as $\eta_1^{\sigma_1}\cdots\eta_d^{\sigma_d}$ for some $\sigma_i \in \{-1, 0, 1\}$. One can prove this lemma using the greedy algorithm, extracting dissociated subsets of S of size d one at a time until it is no longer possible to do so.

3. The main argument

We are now ready to prove Theorem 1.2. In fact, we shall prove a slightly more general version of the theorem, for it turns out that the only relevant consequence of $h = f * g$ being a convolution is that $\|\widehat{h}\|_{l^1}$ is relatively small compared to $\mathbb{E}h$. This l^1-control on $\widehat{f * g}$ follows from the fact that $\widehat{f * g} = \widehat{f}\cdot\widehat{g}$, the Cauchy-Schwarz inequality and Parseval's identity:

$$\|\widehat{f}\cdot\widehat{g}\|_{l^1} \leqslant \|\widehat{f}\|_{l^2}\|\widehat{g}\|_{l^2} \leqslant (\alpha\beta)^{1/2}.$$

Theorem 3.1 (L^p-almost-periodicity for functions with small $\|\widehat{\cdot}\|_{l^1}$)**.** *Let $\epsilon > 0$ and $m \in \mathbb{N}$ be two parameters, and let $f : G \to \mathbb{C}$ be a function on a finite abelian group. Then there is a Bohr set $B = B(\Gamma, \rho)$ of rank $|\Gamma| \ll m^2\log(1/\epsilon)/\epsilon^2$ and radius $\rho = c\epsilon^3/m$ such that*

$$\|f(x+t) - f(x)\|_{L^{2m}(x)} \leqslant \epsilon\|\widehat{f}\|_{l^1} \tag{3.1}$$

for each $t \in B$.

The rest of this section is devoted to the proof of this theorem. Let us write $\delta = \|f\|_{L^1}$ and $\kappa = \|\widehat{f}\|_{l^1}$; note that $\delta \leqslant \kappa$ since $|f(x)| \leqslant \|\widehat{f}\|_{l^1}$ for all $x \in G$ by Fourier inversion.

We begin by partitioning $\widehat{G}$ according to how each character interacts with $\widehat{f}$. Let $M \in \mathbb{N}$ be a parameter to be chosen later, and write

$$\widehat{G} = \bigcup_{k=1}^{M} \Gamma_k \cup \Gamma_{\text{small}}$$

where

$$\Gamma_k = \{\gamma : 2^{-k}\delta < |\widehat{f}(\gamma)| \leqslant 2^{-k+1}\delta\}$$

and

$$\Gamma_{\text{small}} = \{\gamma : |\widehat{f}(\gamma)| \leqslant 2^{-M}\delta\}.$$

Note that $|\widehat{f}(\gamma)| \leqslant \|f\|_{L^1} = \delta$ for each $\gamma \in \widehat{G}$, so this really does partition the set of characters. The characters in Γ_{small} should be thought of as contributing to an error term eventually, whereas we shall get a serious contribution from the characters in the Γ_ks. This contribution can be isolated further by partitioning out dissociated subsets of each Γ_k. Let d be another integer parameter to be chosen later, and apply Lemma 2.5 to Γ_k to obtain a partition

$$\Gamma_k = \bigcup_{j=1}^{v_k} \Gamma_k^{(j)} \cup \Gamma_k^{\text{struct}}$$

where each set $\Gamma_k^{(j)}$ has size d and each Γ_k^{struct} is contained in the $\{-1, 0, 1\}$-span of at most d elements in $\widehat{G}$.

Next we decompose f according to its Fourier expansion: $f(x) = \sum_\gamma \widehat{f}(\gamma)\gamma(x)$, which we write as

$$\begin{aligned} &\sum_{\gamma \in \Gamma_{\text{small}}} \widehat{f}(\gamma)\gamma(x) + \sum_{k=1}^{M}\sum_{j=1}^{v_k}\sum_{\gamma \in \Gamma_k^{(j)}} \widehat{f}(\gamma)\gamma(x) + \sum_{k=1}^{M}\sum_{\gamma \in \Gamma_k^{\text{struct}}} \widehat{f}(\gamma)\gamma(x) \\ &= g_1(x) \qquad\qquad + \qquad\qquad g_2(x) \qquad\qquad + \qquad\qquad g_3(x) \end{aligned}$$

in the indicated way. Note that $\widehat{g_1}(\gamma) = \widehat{f}(\gamma)$ for $\gamma \in \Gamma_{\text{small}}$ and is 0 otherwise, and similarly for the other g_i. From the point of view of establishing (3.1), g_1 and g_2 will constitute error terms in the sense that they contribute very little to the left-hand side, regardless of the choice of element $t \in G$. It is in dealing with the contribution of g_3 that we shall need to be careful in our choice of t, and it is here that the Bohr set of the conclusion will come into play. Thus we may write

$$\begin{aligned} \|f(x+t) - f(x)\|_{L^{2m}(x)} &\leqslant \|g_1(x+t) - g_1(x)\|_{L^{2m}(x)} + \|g_2(x+t) - g_2(x)\|_{L^{2m}(x)} \\ &\quad + \|g_3(x+t) - g_3(x)\|_{L^{2m}(x)} \\ &\leqslant 2\|g_1\|_{L^{2m}} + 2\|g_2\|_{L^{2m}} + \|g_3(x+t) - g_3(x)\|_{L^{2m}(x)} \qquad (3.2) \end{aligned}$$

We deal with each of these in turn.

3.2. **The term g_1.** Our task is to estimate $\|g_1\|_{L^{2m}}$. This turns out to be straightforward if we make use of the following (simple) special case of Young's inequality.

Lemma 3.3. *Let k be a positive integer and let $f, g : G \to \mathbb{C}$ be two functions. Then*
$$\|f * g\|_{l^k} \leqslant \|f\|_{l^1}\|g\|_{l^k}.$$

Thus
$$\|g_1\|_{L^{2m}}^{2m} = \mathbb{E}_x|g_1(x)^m|^2 = \sum_\gamma |\underbrace{\widehat{g_1} * \cdots * \widehat{g_1}}_{m}(\gamma)|^2$$
$$= \|\widehat{g_1} * \cdots * \widehat{g_1}\|_{l^2}^2 \leqslant \left(\|\widehat{g_1}\|_{l^1}^{m-1}\|\widehat{g_1}\|_{l^2}\right)^2,$$
the inequality being a repeated application of Young's inequality.

Now,
$$\|\widehat{g_1}\|_{l^1} = \sum_{\gamma\in\Gamma_{\text{small}}} |\widehat{f}(\gamma)| \leqslant \kappa$$
since $\|\widehat{f}\|_{l^1} = \kappa$, and
$$\|\widehat{g_1}\|_{l^2}^2 = \sum_{\gamma\in\Gamma_{\text{small}}} |\widehat{f}(\gamma)|^2 \leqslant 2^{-M}\delta\kappa$$
by definition of Γ_{small}. We therefore conclude the estimate
$$\|g_1\|_{L^{2m}} \leqslant 2^{-M/2m}\kappa \tag{3.3}$$
for this term, since $\delta \leqslant \kappa$.

3.4. **The term g_2.** We now estimate $\|g_2\|_{L^{2m}}$ using Rudin's inequality and the dyadic partitioning that we performed on the Γ_k. We have

$$\|g_2\|_{L^{2m}} = \|\sum_{k=1}^{M}\sum_{j=1}^{v_k}\sum_{\gamma\in\Gamma_k^{(j)}} \widehat{f}(\gamma)\gamma(x)\|_{L^{2m}} \leqslant \sum_{k=1}^{M}\sum_{j=1}^{v_k}\|\sum_{\gamma\in\Gamma_k^{(j)}} \widehat{f}(\gamma)\gamma(x)\|_{L^{2m}}$$

and, since $\Gamma_k^{(j)}$ is dissociated, this is at most an absolute constant times
$$\sqrt{m}\sum_{k=1}^{M}\sum_{j=1}^{v_k}\left(\sum_{\gamma\in\Gamma_k^{(j)}} |\widehat{f}(\gamma)|^2\right)^{1/2}.$$

But we essentially know the value of $\widehat{f}(\gamma)$ for $\gamma \in \Gamma_k$, using which we can bound this expression above by
$$2\sqrt{\frac{m}{d}}\sum_{k=1}^{M}\sum_{j=1}^{v_k}\sum_{\gamma\in\Gamma_k^{(j)}} |\widehat{f}(\gamma)| \ll \sqrt{\frac{m}{d}}\kappa.$$

Thus
$$\|g_2\|_{L^{2m}} \ll \sqrt{\frac{m}{d}}\kappa. \tag{3.4}$$

3.5. **The term g_3.** It remains to define our Bohr set $B(\Gamma,\rho)$ and estimate $\|g_3(x+t)-g_3(x)\|_{L^{2m}(x)}$ accordingly. By construction of Γ_k^{struct}, we can find elements $\eta_{k,1},\dots,\eta_{k,d}\in\widehat{G}$ for each k such that

$$\Gamma_k^{\text{struct}} \subseteq \operatorname{Span}_{\{-1,0,1\}}(\eta_{k,1},\dots,\eta_{k,d}).$$

Define

$$\Gamma = \bigcup_{k=1}^{M}\{\eta_{k,1},\dots,\eta_{k,d}\},$$

a set of at most Md characters, and let $\rho>0$ be a parameter to be chosen later. Let t be in the Bohr set $B(\Gamma,\rho)$.

Lemma 3.6. *Let $z_1,\dots,z_d\in\mathbb{C}$ be elements of absolute value 1. Then*

$$|z_1\cdots z_d-1|\leqslant|z_1-1|+\cdots+|z_d-1|.$$

Proof.

$$|x_1x_2-1|=|x_2(x_1-1)+(x_2-1)|\leqslant|x_1-1|+|x_2-1|. \qquad \square$$

Now, any character $\gamma\in\Gamma_k^{\text{struct}}$ has an expression of the form $\eta_{k,1}^{\sigma_1}\cdots\eta_{k,d}^{\sigma_d}$ for some $\sigma_j\in\{-1,0,1\}$ and so

$$|\gamma(t)-1|\leqslant d\rho$$

by the preceeding lemma and the definition of $B(\Gamma,\rho)$. Hence

$$\begin{aligned}\|g_3(x+t)-g_3(x)\|_{L^{2m}(x)} &= \|\sum_{k=1}^{M}\sum_{\gamma\in\Gamma_k^{\text{struct}}}\widehat{f}(\gamma)\gamma(x)(\gamma(t)-1)\|_{L^{2m}(x)}\\ &\leqslant \sum_{k=1}^{M}\sum_{\gamma\in\Gamma_k^{\text{struct}}}|\widehat{f}(\gamma)||\gamma(t)-1|\\ &\leqslant d\rho\kappa. \end{aligned}\tag{3.5}$$

3.7. **Combining the estimates.** From (3.2), (3.3), (3.4) and (3.5) we obtain the final estimate

$$\|f(x+t)-f(x)\|_{L^{2m}(x)}\ll\left(2^{-M/2m}+\sqrt{\frac{m}{d}}+d\rho\right)\|\widehat{f}\|_{l^1}.$$

The only remaining task is to pick parameters. We want $2^{-M/2m}+\sqrt{\frac{m}{d}}+d\rho$ to be less than some small constant times ϵ, so we pick

$$\begin{aligned}M &\approx Cm\log(1/\epsilon),\\ d &\approx Cm/\epsilon^2,\ \text{and}\\ \rho &= c\epsilon/d = c\epsilon^3/m.\end{aligned}$$

4. Finding long arithmetic progressions in $A+B$

In this section we show how to deduce Theorem 1.1 from Theorem 1.2. We use the following pigeonholing argument of Bourgain.

Lemma 4.1. *Let $f : G \to [0,1]$ and $P \subseteq G$. Suppose that*

$$\mathbb{E}_x \max_{t \in P} |f(x+t) - f(x)| < \mathbb{E} f.$$

Then $\operatorname{supp}(f) = \{x : f(x) > 0\}$ *contains a translate of P.*

Proof. By the pigeonhole principle, there is some element x for which

$$\max_{t \in P} |f(x+t) - f(x)| < f(x).$$

Hence $f(x+t) > 0$ for each $t \in P$, giving $x + P \subseteq \operatorname{supp}(f)$. □

Proof of Theorem 1.1. We are given two sets A and B of densities α and β; let us write $\delta = (\alpha\beta)^{1/2}$. Apply Theorem 1.2 to 1_A and 1_B with parameters ϵ and m to get a set Γ of characters, $|\Gamma| \leqslant Cm^2 \log(1/\epsilon)/\epsilon^2$, and a radius $\rho = c\epsilon^3/m$ such that for any $t \in B(\Gamma, \rho)$ we have

$$\|1_A * 1_B(x+t) - 1_A * 1_B(x)\|_{L^{2m}(x)} \leqslant \epsilon\delta.$$

Let P be an arithmetic progression of length k in B, where

$$k \leqslant \rho N^{1/|\Gamma|}/2\pi. \tag{4.1}$$

Then

$$\begin{aligned}
\mathbb{E}_x \max_{t \in P} |1_A * 1_B(x+t) - 1_A * 1_B(x)| &\leqslant \mathbb{E}_x \left(\sum_{t \in P} |1_A * 1_B(x+t) - 1_A * 1_B(x)|^{2m} \right)^{1/2m} \\
&\leqslant \left(\sum_{t \in P} \mathbb{E}_x |1_A * 1_B(x+t) - 1_A * 1_B(x)|^{2m} \right)^{1/2m} \\
&\leqslant k^{1/2m} \epsilon\delta.
\end{aligned}$$

We shall therefore be done if $k^{1/2m}\epsilon < \delta$. Let us pick $\epsilon = \delta/2k^{1/2m}$ and $m \approx \log k$. The only restriction on k comes from (4.1): we can find an arithmetic progression of length k in $A+B$ provided

$$k \leqslant \frac{c\delta^3}{\log k} \exp\left(\frac{c\delta^2 \log N}{\log(1/\delta)(\log k)^2} \right),$$

which is true provided

$$k \approx \exp\left(c(\delta^2 \log N)^{1/3} \right).$$

(We have assumed that δ is not too small in terms of N here; some slightly messy calculations would be needed to get a more precise allowed relationship between δ and N.) □

References

[1] J. Bourgain, *On arithmetic progressions in sums of sets of integers,* A tribute to Paul Erdős, 105–109 (CUP, 1990).

[2] E. Croot and O. Sisask, *A probabilistic technique for finding almost-periods of convolutions,* arXiv:1003.2978 (2010).

[3] B. Green, *Arithmetic progressions in sumsets,* Geom. Funct. Anal. **12** (2002), no. 3, 584–597.

[4] B. Green, expository notes at `http://www.dpmms.cam.ac.uk/~bjg23/notes.html`.

[5] T. C. Tao and V. H. Vu, *Additive Combinatorics* (CUP, 2006).