Imperfect friezes of integers

Mário Bessa and Maria Carvalho

October 1, 2009

Abstract

We show that for any positive forward density subset \( N \subset \mathbb{Z} \), there exists \( N' \in N \), such that, for all \( n \geq N \), \( N' \) contains almost perfect \( n \)-scaled reproductions of any previously chosen finite set of integers.

1 Introduction

Many problems in Number Theory are easy to state but very difficult to solve. A quintessential example is the yet unsolved famous Goldbach's conjecture which asserts that all integers greater than or equal to 4 can be written as the sum of two primes. Another renowned problem, aiming to find highly symmetric and arbitrarily long blocks of equidistant points within a given subset of the integers, is to settle whether the celebrated set of primes contains arithmetic progressions with arbitrarily large size.

We say that a set \( N \subset \mathbb{Z} \) has positive density in \( \mathbb{Z} \) if

\[
\Delta(N) := \lim_{n \to +\infty} \frac{1}{2n+1} \# \{-n \leq i \leq n : i \in N\} > 0,
\]

where \( \# \) denotes the set cardinal. The upper (resp. lower) density is defined analogously by taking the lim sup (resp. lim inf). For instance, \( \Delta(\mathbb{Z}) = 1 \), \( \Delta(\mathbb{F}) = 0 \) if \( \mathbb{F} \) is a finite set and \( \Delta(\{\text{Odd integers}\}) = \Delta(\{\text{Even integers}\}) = 1/2 \).

Szemerédi [3] proved that any positive upper density subset \( N \subset \mathbb{Z} \) contains arbitrarily long arithmetic progressions. Unfortunately, we cannot apply Szemerédi’s theorem to the set of primes because its density is zero [1]. This question was addressed recently by Ben Green and the Fields Medal winner Terence Tao, and solved positively in the remarkable work [2].

Szemerédi’s theorem guarantees that, taking \( N \subset \mathbb{Z} \) with positive upper density and an integer \( k \geq 1 \), there exist \( a, b \in \mathbb{Z} \) such that \( a + jb \in N \), for \( j = 0, ..., k-1 \). However, this result does not give any information about the common difference \( b \). In particular, we may ask if \( N \) contains a finite arithmetic progression with common difference equal to a previously fixed \( d \in \mathbb{N} \) but, in general, this is false (e.g. \( d \) an odd integer and \( N = \{\text{Even integers}\} \)). Let us see how we overcame this difficulty.

Appoint \( N \subset \mathbb{Z} \), then take \( k \in \mathbb{N} \) and consider a finite set \( Q = \{q_1, q_2, ..., q_k\} \) of \( \mathbb{Z} \) such that \( q_1 < q_2 < ... < q_k \). Fix \( n \in \mathbb{N} \), bigger or equal to \( k \), and \( \epsilon > 0 \). An \( n \)-scale of \( Q \) \( \epsilon \)-contained in \( N \) is a set \( \{r_1, r_2, ..., r_k\} \subset N \) with \( k \) elements such that

\[
|r_i - r_j| < n\epsilon,
\]

\[\star\]

Partially supported by Fundação para a Ciência e a Tecnologia (FCT), SFRH/BPD/20890/2004.

\[\dagger\]

Denote the number of primes less than \( N \) by \( \pi(N) \). Recall the asymptotic relation \( \pi(N) \sim \frac{N}{\ln(N)} \) and deduce that \( \Delta(\{\text{Primes}\}) = 0 \).
where \( \tau_1 = 0 \) and, if \( k > 1 \),

\[
\tau_2 = \frac{q_2 - q_1}{q_k - q_1} n, \quad \tau_k = \frac{q_{k-1} - q_1}{q_k - q_1} n, \quad \tau_k = n.
\]

Observe that \( r_i - r_{i-1} \), when normalized by the size of the interval \([q_1, q_i]\), is an \( n \)-homothety of \( q_i - q_{i-1} \) up to an error not exceeding \( 2n \epsilon \).

We will see that, under a sharper definition of density of \( \mathcal{N} \), \( \epsilon \)-contained \( n \)-scale sequences exist in \( \mathcal{N} \) for any \( \epsilon > 0 \) and any large enough \( n \) depending on \( \mathcal{N} \), on the fixed set \( \mathcal{Q} \) and on the required accuracy \( \epsilon \). Moreover, this result holds for any finite subset of positive integers, not necessarily within an arithmetic progression.

In the sequel, we will say that \( \mathcal{N} \subset \mathbb{Z} \) has \textit{positive forward density} if the following limit exists and is positive:

\[
\Delta^+(\mathcal{N}) := \lim_{n \to +\infty} \frac{1}{n+1} \# \{0 \leq i \leq n : i \in \mathcal{N} \}.
\]

**Theorem 1.** If \( \mathcal{N} \subset \mathbb{Z} \) has positive forward density, given \( \epsilon > 0 \), \( k \in \mathbb{N} \) and \( \mathcal{Q} \) any set of \( k \) integers, there exists \( N \in \mathbb{N} \) such that, for all \( n \geq N \), we can find an \( n \)-scale of \( \mathcal{Q} \) \( \epsilon \)-contained in \( \mathcal{N} \).

### 2 Proof of Theorem 1

Let \( X = \{0, 1\}^\mathbb{Z} \) be the space of sequences of 0’s and 1’s. We define the shift map \( \sigma : X \to X \) by

\[
\sigma(...x_{-2}x_{-1}x_0x_1x_2...) = ...x_{-1}x_0x_1x_2x_3...
\]

For example, \( \sigma(...000000... = \bar{0}) = \bar{0} \) is a fixed point of \( \sigma \); the periodic sequence \( \overline{10} \) defined by \( y_{2m} = 1 \) and \( y_{2m+1} = 0 \), for \( m \in \mathbb{Z} \), is fixed by \( \sigma \circ \sigma \).

Given \( \mathcal{N} \subset \mathbb{Z} \), we can single out in \( X \) a unique sequence \((x_m)_{m \in \mathbb{Z}}\) which \textit{detects} if an integer belongs to \( \mathcal{N} \): \( x_m = 1 \) if \( m \in \mathcal{N} \) and \( x_m = 0 \) if \( m \notin \mathcal{N} \). For example, if \( \mathcal{N} = \{ \text{Even numbers} \} \), then \((x_m)_{m \in \mathbb{Z}} = \overline{10}) \). We will call \((x_m)_{m \in \mathbb{Z}}\) the sequence that observes \( \mathcal{N} \).

Let \( \mathcal{N} \subset \mathbb{Z} \) be a positive forward density set and \((x_m)_{m \in \mathbb{Z}}\) the sequence that observes it. Consider

\[
\Gamma := \{(y_m)_{m \in \mathbb{Z}} \in X : y_0 = 1\}.
\]

**Lemma 2.1.**

\[
\lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) > 0.
\]

**Proof.** Notice that \( \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) = 1 \) if \( \sigma^i((x_m)_{m \in \mathbb{Z}}) \in \Gamma \), that is, if \( i \in \mathcal{N} \); and we have \( \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) = 0 \) otherwise. Therefore,

\[
\lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) = \lim_{n \to +\infty} \frac{1}{n} \# \{0 \leq i \leq n-1 : i \in \mathcal{N} \} = \Delta^+(\mathcal{N}) > 0.
\]
For simplicity of notation, let $\beta$ be the positive limit

$$
\beta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})).
$$

Fix now the accuracy $\epsilon > 0$ required in Theorem 1 and consider $\overline{\epsilon} = \min\{\epsilon, 1\}$ if $k = 1$, and $\overline{\epsilon} < \min\{\epsilon, \frac{1}{2(q_k - q_1)}\}$ if $k > 1$. Notice that, this way, $\overline{\epsilon} < 2$ because, if $k \geq 1$, then $q_k$ and $q_1$ belong to $\mathbb{Z}$. Then take $\delta \in [0, \beta]$ verifying

$$
\frac{\beta + \delta}{\beta - \delta} < 1 + \frac{\overline{\epsilon}}{2}
$$

(2)

and select $n_0 \in \mathbb{N}$ such that, for every $n \geq n_0$, we have

$$
\left| \beta - \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) \right| < \delta.
$$

(3)

Observe that, this way,

$$
\beta - \delta < \frac{1}{n_0} \sum_{i=0}^{n_0-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m \in \mathbb{Z}})) \leq 1.
$$

(4)

Moreover, choose an integer $N_0$ satisfying the inequality

$$
N_0 > \max\left\{ \frac{2n_0}{\overline{\epsilon} (\beta - \delta)}, \frac{4}{\overline{\epsilon}} \right\}
$$

(5)

which implies that $N_0 > n_0$ because $\beta - \delta < 1$ and $\overline{\epsilon} < 2$. Then:

**Lemma 2.2.** For all $n \geq N_0$ and all $t \in [0, 1]$ there exists $r \in \{0, 1, ..., n\}$ such that:

(i) $\sigma^r((x_m)_{m \in \mathbb{Z}})) \in \Gamma$;

(ii) $|\frac{n}{n} - t| < \overline{\epsilon}$.

**Proof.** The following argument was suggested by the proof of Lemma 3.12 of [1]. Let us assume, by contradiction, that there exist $n \geq N_0$ and $t \in [0, 1]$ such that $\sigma^r((x_m)_{m \in \mathbb{Z}}) \notin \Gamma$ for all $r \in \{0, 1, ..., n\}$; in particular, this holds for all $r \in [n(t - \overline{\epsilon}), n(t + \overline{\epsilon})]$.

Let $[s_1, s_2]$ be the maximal closed interval in $[n(t - \overline{\epsilon}), n(t + \overline{\epsilon})] \cap [0, n]$ where $s_1, s_2 \in \mathbb{Z}$.

Claim:

$$
s_2 - s_1 > \frac{n\overline{\epsilon}}{2}.
$$

(6)

In fact:

- If $n(t - \overline{\epsilon}) \geq 0$ and $n(t + \overline{\epsilon}) \leq n$, then, denoting by $\lfloor z \rfloor$ the biggest integer less or equal than $z$, we have

$$
s_2 - s_1 \geq \lfloor n(t + \overline{\epsilon}) \rfloor - \lfloor n(t - \overline{\epsilon}) \rfloor + 1 > n(t + \overline{\epsilon}) - 1 - n(t - \overline{\epsilon}) - 1 = 2n\overline{\epsilon} - 2 > \frac{n\overline{\epsilon}}{2}
$$

since $n \geq N_0$ and, by (5), $N_0\overline{\epsilon} > 4$. 


• If \( n(t - \tau) \geq 0 \) and \( n(t + \tau) > n \), then
\[
s_2 - s_1 \geq n - (\lfloor n(t - \tau) \rfloor + 1) \geq n - n(t - \tau) - 1 = n\tau + (n - 1 - nt) \geq n\tau - 1 > \frac{n\tau}{2}.
\]

• If \( n(t - \tau) < 0 \) and \( n(t + \tau) \leq n \), then
\[
s_2 - s_1 \geq \lfloor n(t + \tau) \rfloor > n(t + \tau) - 1 > n\tau - 1 > \frac{n\tau}{2}.
\]

• If \( n(t - \tau) < 0 \) and \( n(t + \tau) > n \), then \( s_2 - s_1 = n > \frac{n\tau}{2} \) since \( \tau < 2 \). \( \square \)

Let us go back to the proof of the Lemma. If \( s_1 \geq n_0 \), then, since \( s_1 \leq n \) and \( \sigma^r((x_m)_{m\in\mathbb{Z}}) \notin \Gamma \) for all \( r \in \{0,1,\ldots,n\} \), we may deduce that
\[
\beta - \delta < \frac{1}{s_2} \sum_{i=0}^{s_2-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}})) = \frac{1}{s_2} \sum_{i=0}^{s_1-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}}))
\]
\[
\leq \frac{1}{s_1 + \frac{n\tau}{2}} \sum_{i=0}^{s_1-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}})) \leq \frac{1}{s_1(1 + \frac{\tau}{2})} \sum_{i=0}^{s_1-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}}))
\]
\[
< \frac{\beta + \delta}{1 + \frac{\tau}{2}} < \beta - \delta,
\]
which is a contradiction.

On the other hand, if \( s_1 < n_0 \), then, since \( s_1 \geq 0 \) and \( n \geq N_0 \), we have
\[
s_2 \geq s_2 - s_1 > \frac{n\tau}{2} > \frac{n_0}{\beta - \delta} > n_0,
\]
(7) because \( \beta - \delta < 1 \). Therefore
\[
\beta - \delta < \frac{1}{s_2} \sum_{i=0}^{s_2-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}})) = \frac{1}{s_2} \sum_{i=0}^{s_1-1} \mathbb{I}_\Gamma(\sigma^i((x_m)_{m\in\mathbb{Z}})) < \frac{s_1}{s_1 + \frac{n\tau}{2}}
\]
\[
< \frac{n_0}{\frac{\tau}{2}} < \beta - \delta,
\]
which is again a contradiction.

\( \square \)

We may now end the proof of Theorem \( \square \). Given \( k \in \mathbb{N} \) and \( Q := \{q_1, \ldots, q_k\} \subset \mathbb{Z} \) such that \( q_1 < q_2 < \cdots < q_k \), we fix \( N_0 \) as above and \( n \geq N = \max\{k, N_0\} \). Then we apply \( k \) times the Lemma \( \square \), using the numbers \( t_1 = 0 \), and, if \( k > 1 \),
\[
t_2 = \frac{q_2 - q_1}{q_k - q_1}, \ldots, t_{k-1} = \frac{q_{k-1} - q_1}{q_k - q_1}, t_k = 1.
\]
This way we get, for \( (x_m)_{m\in\mathbb{Z}} \), a finite set \( \{r_1, \ldots, r_k\} \subset \{0,1,\ldots,n\} \) such that:

(i) \( \sigma^{r_k}((x_m)_{m\in\mathbb{Z}}) \in \Gamma \);
(ii) \(|\frac{q_i}{n} - t_i| < \epsilon\).

Item (i) means that, for all \(i = 1, \ldots, k\),

\[ x_{r_i} = 1, \text{ for all } i = 1, \ldots, k, \]

which is equivalent to say that

\[ r_i \in \mathcal{N}, \text{ for all } i = 1, \ldots, k. \]

Besides, if \(k > 1\), then \(r_i \neq r_j\) if \(i \neq j\). Indeed, by item (ii), for each \(i \in \{1, \ldots, k-1\}\), we have

\[ n \frac{q_i - q_1}{k - q_1} - n\epsilon < r_i < n \frac{q_i - q_1}{k - q_1} + n\epsilon \]

\[ n \frac{q_{i+1} - q_1}{k - q_1} - n\epsilon < r_{i+1} < n \frac{q_{i+1} - q_1}{k - q_1} + n\epsilon \]

and

Claim:

\[ n \frac{q_{i+1} - q_1}{k - q_1} - n\epsilon > n \frac{q_i - q_1}{k - q_1} + n\epsilon. \]

This means that the intervals where \(r_i\) and \(r_{i+1}\) live are disjoint, and therefore these numbers cannot be equal.

The last inequality is a consequence of the choice \(|\epsilon| < \frac{1}{2(kq_k-q_1)}\). In fact, taking into account that \(Q \subset \mathbb{Z}\), from it, we get:

\[ 2n\epsilon < n \left( \frac{1}{kq_k-q_1} \right) \leq n \left( \frac{q_{i+1} - q_i}{kq_k-q_1} \right) = n \left( \frac{q_{i+1} - q_1}{kq_k-q_1} - \frac{q_i - q_1}{kq_k-q_1} \right) \]

as wished. \(\Box\)

Finally, \(\{r_i\}_{i=1}^k\) is an \(n\)-scale of \(Q\) \(\epsilon\)-contained in \(\mathcal{N}\) due to the inequality \(\epsilon < \epsilon\), the judicious choice of the \(t_i\)'s and item (ii). \(\Box\)

References

[1] J. Bochi, Genericy of zero Lyapunov exponents. \textit{Ergodic Theory Dynam. Systems}, 22(6):1667–1696, 2002.

[2] B. Green e T. Tao, The primes contain arbitrarily long arithmetic progressions. \textit{Ann. of Math.}, 167(2):481–547, 2008.

[3] E. Szemerédi, On sets of integers containing no \(k\) elements in arithmetic progression. \textit{Acta Arith.}, 27:199–245, 1975.

Mário Bessa (bessa@fc.up.pt) FCUP, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal and ESTGOH - IPC, Rua General Santos Costa, 3400-124, Oliveira do Hospital, Portugal.

Maria Carvalho (mpcarval@fc.up.pt) CMUP, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.