Chapter 14
Near Arithmetic Progressions in Sparse Sets

Szemeredi's theorem says that relatively dense sets contain arithmetic progressions. The purpose of this chapter is to present a result of Leth from [89] which shows that certain sparse sets contain "near" arithmetic progressions. We then detail the connection between the aforementioned theorem of Leth and the Erdős-Turán conjecture.

14.1 The Main Theorem

We begin by making precise the intuitive notion of "near" arithmetic progression mentioned in the introduction.

Definition 14.1 Fix $w \in \mathbb{N}_0$ and $t, d \in \mathbb{N}$.[1] A (t, d, w)-*progression* is a set of the form

$$\mathrm{B}(b, t, d, w) := \bigcup_{i=0}^{t-1}[b + id, b + id + w].$$

By a *block progression* we mean a (t, d, w)-progression for some t, d, w.

Note that a $(t, d, 0)$-progression is the same thing as a t-term arithmetic progression with difference d.

---

[1]In this chapter, we deviate somewhat from our conventions so as to match up with the notation from [89].

Definition 14.2 If $A \subseteq \mathbb{N}$, we say that A *nearly contains a* (t, d, w)-*progression* if there is a (t, d, w)-progression $B(b, t, d, w)$ such that $A \cap [b + id, b + id + w] \neq \emptyset$ for each $i = 1, \ldots, t - 1$.

Thus, if A nearly contains a $(t, d, 0)$-progression, then A actually contains a t-term arithmetic progression. Consequently, when A nearly contains a (t, d, w)-progression with "small" w, then this says that A is "close" to containing an arithmetic progression. The main result of this chapter allows us to conclude that even relatively sparse sets with a certain amount of density regularity nearly contain block progressions satisfying a further homogeneity assumption that we now describe.

Definition 14.3 Suppose that $A \subseteq \mathbb{N}$, I is an interval in $\mathbb{N}$, and $0 < s < 1$. We say that A *nearly contains a* (t, d, w)-*progression in* I *with homogeneity* s if there is some $B(b, t, d, w)$ contained in I such that the following two conditions hold for all $i, j = 0, 1, \ldots, t - 1$:

(i) $\delta(A, [b + id, b + id + w]) \geq (1 - s)\delta(A, I)$
(ii) $\delta(A, [b + id, b + id + w]) \geq (1 - s)\delta(A, [b + jd, b + jd + w])$.

Thus, for small s, we see that A meets each block in a density that is roughly the same throughout and that is roughly the same as on the entire interval.

The density regularity condition roughly requires that on sufficiently large subintervals of I, the density does not increase too rapidly. Here is the precise formulation:

Definition 14.4 Suppose that $I \subseteq \mathbb{N}$ is an interval, $r \in \mathbb{R}$, $r > 1$, and $m \in \mathbb{N}$. We say that $A \subseteq I$ has the (m, r)-*density property on* I if, whenever $J \subseteq I$ is an interval with $|J|/|I| \geq 1/m$, then $\delta(A, J) \leq r\delta(A, I)$.

Of course, given any $m \in \mathbb{N}$ and $A \subseteq I$, there is $r \in \mathbb{R}$ such that A has the (m, r)-density property on I. The notion becomes interesting when we think of r as fixed.

Given a hyperfinite interval $I \subseteq {}^*\mathbb{N}$, $r \in {}^*\mathbb{R}$, $r > 1$, and $M \in {}^*\mathbb{N}$, we say that an internal set $A \subseteq I$ has the *internal* (M, r)-*density property on* I if the conclusion of the definition above holds for internal subintervals J of I.

Lemma 14.5 *Suppose that* $A \subseteq [1, N]$ *is an internal set with the internal* (M, r)-*density property for some* $M > \mathbb{N}$. *Let* $f : [0, 1] \to [0, 1]$ *be the (standard) function given by*

$$f(x) := \mathrm{st}\left(\frac{|A \cap [1, xN]|}{|A \cap [1, N]|} \right).$$

Then f *is a Lipschitz function with Lipschitz constant* r.

Proof Fix $x < y$ in $[0, 1]$. Write $x := \mathrm{st}(K/N)$ and $y := \mathrm{st}(L/N)$. Since $y - x \neq 0$, we have that $\frac{L - K}{N}$ is not infinitesimal; in particular, $\frac{L - K}{N} > 1/M$. Since A has the (M, r)-density property on $[1, N]$, we have that $\delta(A, [K, L]) \leq r\delta(A, [1, N])$.

Thus, it follows that

$$f(y) - f(x) = \text{st}\left(\frac{|A \cap [K, L]|}{|A \cap [1, N]|}\right) = \text{st}\left(\delta(A, [K, L])\frac{L - K}{|A \cap [1, N]|}\right)$$

$$\leq r \,\text{st}\left(\frac{L - K}{N}\right) = r(y - x).$$

Here is the main result of this section:

Theorem 14.6 (Leth) *Fix functions $g, h : \mathbb{R}^{>0} \to \mathbb{R}^{>0}$ such that h is increasing and $g(x) \to \infty$ as $x \to \infty$. Fix also $s > 0$, $r > 1$, and $j, t \in \mathbb{N}$. Then there is $m = m(g, h, s, r, t, j) \in \mathbb{N}$ such that, for all $n > m$, whenever I is an interval of length n and $A \subseteq I$ is nonempty and has the (m, r)-density property on I, then A contains a (t, d, w)-almost progression with homogeneity s such that $w/d < h(d/n)$ and $1/g(m) < d/n < 1/j$.*

Roughly speaking, if A has sufficient density regularity, then A contains an almost-progression with "small" w (small compared to the distance of the progression).

The proof of the theorem relies on the following standard lemma; see [89, Lemma 1].

Lemma 14.7 *Suppose that $E \subseteq \mathbb{R}$ has positive Lebesgue measure and $t \in \mathbb{N}$. Then there is $v > 0$ such that, for all $0 < u < v$, there is an arithmetic progression in E of length t and difference u.*

We stress that in the previous lemma, u and v are real numbers.

Proof of Theorem 14.6 Fix g, h, s, r, j, t as in the statement of Theorem 14.6. We show that the conclusion holds for all infinite M, whence by underflow there exists $m \in \mathbb{N}$ as desired. Thus, we fix $M > \mathbb{N}$ and consider $N > M$, an interval $I \subseteq {}^*\mathbb{N}$ of length N, and a hyperfinite subset $A \subseteq I$ that has the internal (M, r)-density property on I. Without loss of generality, we may assume that $I = [1, N]$. Suppose that we can find $B, D, W \in {}^*\mathbb{N}$ and standard $c > 0$ such that $[B, B + (t - 1)D + W] \subseteq [1, N]$ and, for all $i = 0, 1, \ldots, t - 1$, we have:

$$\delta(A, [1, N])(c - \frac{s}{2}) \leq \delta(A, [B + iD, B + iD + W]) \leq \delta(A, [1, N])(c + \frac{s}{4}). \quad (\dagger)$$

We claim that A nearly contains the internal (t, D, W)-progression $\mathrm{B}(B, t, D, W)$ with homogeneity s. Indeed, item (i) of Definition 14.3 is clear. For item (ii), observe that

$$\delta(A, [B + iD, B + iD + W]) \geq \delta(A, [1, N])(c - \frac{s}{2})$$

$$\geq \delta(A, [B + jD, B + jD + W])(\frac{c - \frac{s}{2}}{c + \frac{s}{4}})$$

and note that $\frac{c-\frac{s}{2}}{c+\frac{s}{4}} > 1 - s$. Thus, it suffices to find B, D, W, c satisfying (†) and for which $W/D < h(D/N)$ and $1/g(M) < D/N < 1/j$.

Let f be defined as in the statement of Lemma 14.5. Set $b := \mathrm{st}(B/N)$, $d := \mathrm{st}(D/N)$, and $w := \mathrm{st}(W/N)$. Assume that $w \neq 0$. Then we have that

$$\mathrm{st}\left(\frac{\delta(A, [B + iD, B + iD + W])}{\delta(A, [1, N])}\right) = \frac{f(b + id + w) - f(b + id)}{w}.$$

We thus want to find B, D, W and c satisfying

$$c - \frac{s}{2} < \frac{f(b + id + w) - f(b + id)}{w} < c + \frac{s}{4}. \quad (\dagger\dagger)$$

Now the middle term in (††) looks like a difference quotient and the idea is to show that one can bound $f'(b + id)$ for $i = 0, 1, \ldots, t - 1$. Indeed, by Lemma 14.5, f is Lipschitz, whence it is absolutely continuous. In particular, by the Fundamental Theorem of Calculus, f is differentiable almost everywhere and $f(x) = \int_0^x f'(u)du$. Since $f(0) = 0$ and $f(1) = 1$, it follows that $\{x \in [0, 1] : f'(x) \geq (1 - \frac{s}{4})\}$ has positive measure. In particular, there is $c > 1$ such that

$$E := \{x \in [0, 1] : c - \frac{s}{4} \leq f'(x) \leq c\}$$

has positive measure. By Lemma 14.7, there is $b \in E$ and $0 < u < 1/j$ such that $b, b + u, b + 2u, \ldots, b + (t - 1)u \in E$. Take $B, D \in [1, N]$ such that $b = \mathrm{st}(B/N)$ and $u = \mathrm{st}(D/N)$. Note that $g(M)$ is infinite and D/N is noninfinitesimal, so $1/g(M) < D/N < 1/j$. It remains to choose W. Since f is differentiable on E, there is $w > 0$ sufficiently small so that for all $i = 0, 1, \ldots, t - 1$, we have $|f'(b + id) - \frac{f(b+id+w)-f(b+id)}{w}| < \frac{s}{4}$. For this w, (††) clearly holds; we now take W such that $w = \mathrm{st}(W/N)$. Since $h(D/N)$ is nonfinitesimal (as D/N is noninfinitesimal), if w is chosen sufficiently small, then $W/D < h(D/N)$.

Theorem 14.6 implies a very weak form of Szemeredi's theorem.

Corollary 14.8 *Suppose that* $\mathrm{BD}(A) > 0$. *Suppose that* g, h, s, t, j *are as in the hypothesis of Theorem 14.6. Then for n sufficiently large, there is an interval I of length n such that $A \cap I$ contains a (t, s, d)-almost progression in I with $w/d < h(d/n)$ and $1/g(m) < d/n < 1/j$.*

Proof Take $r \in \mathbb{R}$ with $r > 1$ satisfying $\mathrm{BD}(A) > 1/r$. Let $m := m(g, h, s, r, t, j)$ as in the conclusion of Theorem 14.6. Let $n > m$ and take an interval I of length n such that $\delta(A, I) > 1/r$. It remains to observe that $A \cap I$ has the (m, r)-density property on I.

14.2   Connection to the Erdős-Turán Conjecture

Leth's original motivation was the following conjecture of Erdős and Turán from [47]:

Conjecture 14.9 (Erdős-Turán) Suppose that $A = (a_n)$ is a subset of $\mathbb{N}$ such that $\sum 1/a_n$ diverges. Then A contains arbitrarily long arithmetic progressions.

Leth first observed the following standard fact about the densities of sequences satisfying the hypotheses of the Erdős-Turán conjecture.

Lemma 14.10 *Suppose that $A = (a_n)$ is enumerated in increasing order and is such that $\sum 1/a_n$ diverges. Then, for arbitrarily large n, one has $\delta(A, n) > 1/(\log n)^2$.*

Proof We argue by contrapositive. Suppose that $\delta(A, n) \leq 1/(\log n)^2$ for all $n \geq n_0 \geq 4$. We first show that this implies that $a_n \geq \frac{1}{2}n(\log n)^2$ for all $n > n_0$. Suppose otherwise and fix $n \geq n_0$. Then $|A \cap [1, \frac{1}{2}n(\log n)^2]| \geq n$. On the other hand, by our standing assumption, we have that

$$|A \cap [1, \frac{1}{2}n(\log n)^2]) \leq \frac{1/2n(\log n)^2}{(\log((1/2n(\log n))^2)} \leq \frac{1}{2}n,$$

yielding the desired contradiction.

Since $a_n \geq \frac{1}{2}n(\log n)^2$ eventually, we have that

$$\sum \frac{1}{a_n} \leq \sum \frac{2}{n(\log n)^2},$$

whence $\sum \frac{1}{a_n}$, converges.

The truth of the following conjecture, together with the theorem that follows it, would imply that, for sets satisfying the density condition in the previous lemma, the existence of almost arithmetic progressions implies the existence of arithmetic progressions.

Conjecture 14.11 (Leth) Fix $t \in \mathbb{N}$ and $c > 0$. Then there is $n_0 := n_0(t, c)$ such that, for all $n \geq n_0$, whenever $A \subseteq \mathbb{N}$ is such that $\delta(A, n) > 1/(c \log n)^{2 \log \log n}$, then A nearly contains a (t, d, w)-progression on $[1, n]$ with $w/d < d/n$ where d is a power of 2.

We should remark that requiring that d be a power of 2 is not much of an extra requirement. Indeed, our proof of Theorem 14.6 shows that one can take d there to be a power of 2. For any t and c, we let $L(t, c)$ be the statement that the conclusion of the previous conjecture holds for the given t and c. We let $L(t)$ be the statement that $L(t, c)$ holds for all $c > 0$.

Theorem 14.12 *Suppose that $L(t)$ is true for a given $t \in \mathbb{N}$. Further suppose that $A \subseteq \mathbb{N}$ is such that there is $c > 0$ for which, for arbitrarily large n, one has $\delta(A, n) > c/(\log n)^2$. Then A contains an arithmetic progression of length t.*

Before we prove this theorem, we state the following standard combinatorial fact, whose proof we leave as an exercise to the reader (alternatively, this is proven in [89, Proposition 1]).

Proposition 14.13 *Let $m, n \in \mathbb{N}$ be such that $m < n$, let $A \subseteq \mathbb{N}$, and let I be an interval of length n. Then there is an interval $J \subseteq I$ of length m such that $\delta(A, J) > \delta(A, I)/2$.*

Proof of Theorem 14.12 For reasons that will become apparent later in the proof, we will need to work with the set $2A$ rather than A. Note that $2A$ satisfies the hypothesis of the theorem for a different constant $c' > 0$.

By overflow, we may find $M > \mathbb{N}$ such that $\delta(^*(2A), M) > \frac{c'}{(\log M)^2}$. Take $L > \mathbb{N}$ such that $2^{2^L} \leq M < 2^{2^{L+1}}$ and set $N := 2^{2^L}$. If we apply Proposition 14.13 to any $n \leq N$ and $I = [1, N]$, we can find an interval $I_n \subseteq [1, M]$ of length n such that

$$|^*(2A) \cap I_n| > \frac{c'M}{2(\log M)^2} \geq \frac{c'M}{2(\log 2^{2^{L+1}})^2} = \frac{c'/8}{(\log N)^2}.$$

For $1 \leq k \leq L$, write $I_{2^{2^k}} = [x_k, y_k]$.

We will now construct an internal set $B \subseteq [1, N]$ such that $\delta(B, N) > \frac{1}{(c'' \log N)^{2 \log \log N}}$, where $c'' := \sqrt{8/c'}$. Since we are assuming that $L(t)$ holds, by transfer we will be able to find an internal (t, d, w)-progression nearly inside of B with $w/d < d/N$ and w and d both powers of 2. The construction of B will allow us to conclude that $^*(2A)$ contains a t-termed arithmetic progression of difference d, whence so does $2A$ by transfer, and thus so does A.

Set $B_0 := [1, N]$ and, for the sake of describing the following recursive construction, view B_0 as the union of two subintervals of length $N/2 = 2^{2^L-1} = 2^{2^L-2^0}$; we refer to these subintervals of B_0 as *blocks*. Now divide each block in B_0 into $2 = 2^{2^0}$ intervals of length $2^{2^L-2^0}/2^{2^0} = 2^{2^L-2^1}$ and, for each $0 \leq j < 2^{2^0}$, we place the jth subblock of each block in B_0 into B_1 if and only if $x_0 + j \in {}^*2A$.

Now divide each block in B_1 into 2^{2^1} intervals of length $2^{2^L-2^1}/2^{2^1} = 2^{2^L-2^2}$ and, for each $0 \leq j < 2^{2^1}$, we place the jth subblock of each block in B_1 into B_2 if and only if $x_1 + j \in {}^*2A$.

We continue recursively in this manner. Thus, having constructed the hyperfinite set B_k, which is a union of blocks of length $2^{2^L-2^k}$, we break each block of B_k into 2^{2^k} many intervals of length $2^{2^L-2^k}/2^{2^k} = 2^{2^L-2^{k+1}}$ and we place the jth subblock of each block in B_k into B_{k+1} if and only if $x_k + j \in {}^*2A$.

We set $B := B_L$. Since $|B_{k+1}|/|B_k| > \frac{c'/8}{(\log N)^2}$ for each $0 \leq k < L$, it follows that

$$|B| > \frac{(c'/8)^L N}{(\log N)^{2L}} = \frac{N}{(c'' \log N)^{2 \log \log N}}.$$

By applying transfer to $L(t)$, we have that B nearly contains an internal (t, d, w)-progression $B(b, t, d, w)$ contained in $[1, N]$ such that $w/d < d/N$ and d is a power of 2. Take k such that $2^{2^L - 2^{k+1}} \leq d < 2^{2^L - 2^k}$. Note that this implies that $2^{2^L - 2^{k+1}} \mid d$. Also, we have

$$w < (d/N) \cdot d < (2^{-2^k}) 2^{2^L - 2^k} = 2^{2^L - 2^{k+1}}.$$

We now note that $B(b, t, d, w)$ must be contained in a single block C of B_k. Indeed, since $d \mid 2^{2^L - 2^k}$ and $w \mid 2^{2^L - 2^{k+1}}$, we have $d + w < (\frac{1}{2} + \frac{1}{2^{2^k}})(2^{2^L - 2^k})$, whence the fact that $[b, b + w]$ and $[b + d, b + d + w]$ both intersect B_k would imply that $[x_{k-1}, y_{k-1}]$ contains consecutive elements of $*2A$, which is clearly a contradiction.

Now write $d = m \cdot 2^{2^L - 2^{k+1}}$. Take $0 \leq j < 2^{2^k}$ so that $[b, b + w]$ intersects B_{k+1} in the jth subblock of C so $x_k + j \in {}^*2A$. Since $[b + d, b + d + w] \cap B_{k+1} \neq \emptyset$, we have that at least one of $x_k + j + (m - 1)$, $x_k + j + m$, or $x_k + j + (m + 1)$ belong to $^*(2A)$. However, since $x_k + j$ and m are both even, it follows that we must have $x_k + j + m \in {}^*(2A)$. Continuing in this matter, we see that $x_k + j + im \in {}^*2A$ for all $i = 0, 1, \ldots, t - 1$. It follows by transfer that $2A$ contains a t-term arithmetic progression, whence so does A.

Putting everything together, we have:

Corollary 14.14 *The Erdős-Turán conjecture follows from Leth's Conjecture.*

Leth used Theorem 14.6 to prove the following theorem, which is similar in spirit to Conjecture 14.6, except that it allows sparser sequences but in turn obtains almost progressions with weaker smallness properties relating d and w.

Theorem 14.15 *Suppose that $s > 0$ and $t \in \mathbb{N}^{>2}$ are given. Further suppose that h is as in Theorem 14.6. Let $A \subseteq \mathbb{N}$ be such that, for all $\epsilon > 0$, we have $\delta(A, n) > 1/n^\epsilon$ for sufficiently large n. Then for sufficiently large n, A nearly contains an (t, d, w)-progression on $[1, n]$ of homogeneity s with $w/d < h(\log d/\log n)$, where d is a power of 2.*

Proof Suppose that the conclusion is false. Then there is N such that $^*A$ does not nearly contain any internal (t, d, w)-progression on $[1, N]$ of homogeneity s with $w/d < h(\log d/\log N)$. It suffices to show that there is $\epsilon > 0$ such that $\delta(^*A, N) < 1/N^\epsilon$. Let m be as in the conclusion of Theorem 14.6 with $r = 2$ and $g(x) = x$ (and h as given in the assumptions of the current theorem).

Claim If $I \subseteq [1, N]$ is a hyperfinite interval with $|I| > \sqrt{N}$, then $^*A$ does *not* have the $(m, 2)$-density property on I.

We will return to the proof of the claim in a moment. We first see how the claim allows us to complete the proof of the theorem. Let $K > \mathbb{N}$ be the maximal $k \in {}^*\mathbb{N}$ such that $m^{2k} \leq N$, so $m^{2K} \leq N < m^{2K+2}$. We construct, by internal induction, for $i = 0, 1, \ldots, K$, a descending chain of hyperfinite subintervals (I_i) of I of length m^{2K-i} as follows. By Proposition 14.13, we may take I_0 to be any hyperfinite subinterval of I of length m^{2K} such that $\delta(^*A, I_0) \geq \delta(^*A, N)/2$. Suppose that $i < K$ and I_i has been constructed such that $|I_i| = m^{2k-i}$. Since $^*A$ does not have the $(m, 2)$ density property on I_i, there is a subinterval I_{i+1} of length $|I_i|/m^{2k-i-1}$ with $\delta(^*A, I_{i+1}) \geq 2\delta(^*A, I_i)$. Notice now that I_K is a hyperfinite interval of length $m^K \leq \sqrt{N} < m^{K+1}$ and $\delta(^*A, I_K) \geq 2^K \delta(^*A, I_0)$. It follows that

$$\delta(^*A, N) \leq 2\delta(A, I_0) \leq 2^{-(K-1)}\delta(A, I_K) \leq 2^{-(K-1)}.$$

It follows that

$$|A \cap [1, N]| \leq 2^{-(K-1)}N \leq 2^{-(K-1)}m^{2K+2} = m^{2K+2-(K-1)\frac{\log 2}{\log m}} = (m^{2K})^{1-z}.$$

if we set $z := \frac{(K-1)\log 2}{2K \log m} - \frac{1}{K}$. If we set $\epsilon := \mathrm{st}(z/2) = \frac{\log 2}{4 \log m}$, then it follows that $|A \cap [1, N]| \leq N^{1-\epsilon}$, whence this ϵ is as desired.

We now prove the claim. Suppose, towards a contradiction, that $I \subseteq [1, N]$ is a hyperfinite interval with $|I| > \sqrt{N}$ and is such that $^*A$ does have the $(m, 2)$-density property on I. By the choice of m, $^*A$ nearly contains an internal (t, d, w)-almost progression of homogeneity s with $w/d < h(d/|I|)$ and $d > |I|/m > \sqrt{N}/m$. Notice now that $\mathrm{st}\left(\frac{\log d}{\log N}\right) \geq \mathrm{st}\left(\frac{1/2 \log N - \log m}{\log N}\right) = \frac{1}{2}$. Note that we trivially have that $d/|I| < 1/t$, whence $d/|I| < \log d/\log N$; since h is increasing, we have that $w/d < h(\log d/\log N)$, contradicting the choice of N. This proves the claim and the theorem.

In [90, Theorem 3], Leth shows that one cannot replace $(\log d)/(\log n)$ with d/n in the previous theorem.

Notes and References

There are other generalizations of arithmetic progressions appearing in the literature, e.g. the notion of *quasi-progression* appearing in [131]. It should be noted that they use the term (t, d, w)-progression in a related, but different, manner than it is used in this chapter. The Erdős-Turan conjecture, first formulated in [48], is one of the most important open problems in combinatorial number theory. A positive solution would immediately generalize both Szemeredi's Theorem and the Green-Tao theorem on the existence of arbitrarily long arithmetic progressions in the primes [64].