Ergodic theory and arithmetic progressions

by

Marcel K. Goh

4 December 2020

1. Introduction

Let $(Z, +)$ be an abelian group. For any integer n and $z \in Z$ we let nz denote the iterated sum; for instance, $2z = z + z$ and $-3z = -z - z - z$. Letting $[a, b]$ denote the discrete interval $\{n \in \mathbf{Z} : a \le n \le b\}$ for $a, b \in \mathbf{Z}$, we define an *arithmetic progression* to be a set of the form

$$a + [0, k-1] \cdot r = \{a, a + r, a + 2r, \ldots, a + (k-1)r\}, \tag{1}$$

where $k > 1$ is the *length*, $a \in Z$ is the *base point*, and $r \in Z$ is the *common difference* or *step*. If $r \ne 0$, the progression is *nontrivial*, and this condition is assumed unless otherwise stated. In these notes, we will study the case $Z = \mathbf{Z}$.

The quest to find arithmetic progressions in subsets of $\mathbf{Z}$ began with the following celebrated theorem of B. L. van der Waerden, which answered a conjecture of P. J. H. Baudet.

Theorem V (*van der Waerden,* 1927). *For any integers r and k, there exists $N \in \mathbf{N}$ such that for any partition of $[1, N]$ into disjoint sets $C_1, C_2, \ldots, C_r$, some C_i contains a k-term arithmetic progression.* ▌

An infinitary formulation of this theorem states that if the positive integers are coloured with r colours, then some colour contains arbitrarily long arithmetic progressions. P. Erdős and P. Turán suspected that this had to do with the density of the partition. A subset A of the nonnegative integers is said to have *positive (upper) density* if

$$\limsup_{N\to\infty} \frac{\bigl|A \cap [0, N]\bigr|}{N+1} > 0. \tag{2}$$

In the more general setting where $A \subseteq \mathbf{Z}^d$ for $d \ge 1$, we get a similar definition by intersecting with $[-N, N]^d$ and dividing by $(2N + 1)^d$. (It does not really matter exactly which interval we intersect with, since we only care to know whether the density is positive or zero.) Erdős and Turán's conjecture was that van der Waerden's theorem holds because at least one of the C_i must have density at least $1/r$. This was proved for the case $k = 3$ by K. F. Roth (1953) using Fourier-analytic methods. A proof for $k = 4$ did not arrive until 1969; it was due to E. Szemerédi, who also proved the general case six years later:

Theorem S (*Szemerédi,* 1975). *Any subset of the nonnegative integers with positive upper density contains arithmetic progressions of arbitrary length. Equivalently, given any real $\delta > 0$ and $k \in \mathbf{N}$, there exists $N \in \mathbf{N}$ such that for any $A \subseteq [0, N]$ with $|A| \ge \delta(N+1)$, there exists an arithmetic progression of length k contained in A.*

The proof that Szemerédi gave was long and complicated, relying on an auxiliary "graph-regularity lemma" that has had far-reaching consequences in combinatorics. We will only supply a proof of the fact that the infinitary version of the theorem is in fact equivalent to the finitary one. (This treatment can be found in Furstenberg, Katznelson, and Ornstein (1982).)

Proof. It is clear that the finitary statement implies the infinitary one. Now suppose that the finitary statement does not hold; that is, there exists $\delta > 0$ and $k \in \mathbf{N}$ such that for all $N \in \mathbf{N}$, there exists $A_N \subseteq [0, N]$ such that $|A_N| \ge \delta(N + 1)$ but there is no arithmetic progression of length k contained in A. Note that the presence of arithmetic progressions in a set A does not change if we shift all the elements of the set by a fixed value: for any $s \in \mathbf{N}$, $s + A = \{s + a : a \in A\}$ has the same number of arithmetic progressions as A. So we can translate the sets A_N far enough apart that no arithmetic progression can belong to two of the A_N at once. Letting $A = \bigcup_{N\in\mathbf{N}} A_N$, we find that A has upper density $\ge \delta$, but does not contain arithmetic progressions of length k. ▌

Szemerédi's quantitative bounds on the size of N relative to δ and k were not as tight as Roth's bounds for the $k = 3$ case. In fact, Roth's Fourier-analytic approach was generalisable, but this was not discovered until 1998 for the case $k = 4$ and finally for the general case in 2001 (both of these results, which established much better bounds, are due to W. T. Gowers). On the other hand, if we are only interested in the infinitary statement and allow non-constructive proofs, then ergodic methods can be applied to give much shorter arguments than were employed in Szemerédi's original paper. This approach, devised by H. Furstenberg in 1977, led to many generalisations of Szemerédi's result.

2. The ergodic formulation

We will need some terminology from the realm of dynamical systems. Let $(X, \mathcal{F}, \mu)$ be a probability space and let $T : X \to X$ be measure-preserving, i.e., $\mu(T^{-n}A) = \mu(A)$ for all $A \in \mathcal{F}$. Note that $T^{-n}A$ is the set of all $x \in X$ such that $Tx \in A$; one should not be tricked into thinking that T is necessarily invertible! The quadruple $(X, \mathcal{F}, \mu, T)$ is called a *measure-preserving system.* For a bounded measurable function $f : X \to \mathbf{R}$, T acts as a shift operator by sending f to $T^n f$, given by $T^n f(x) = f(T^n x)$. We have the following lemma.

Lemma D. *Let $(X, \mathcal{F}, \mu)$ be a probability space and let f be a measurable function on X such that $\int_X f\,d\mu > 0$ and such that for all $x \in X$, $0 \le f(x) \le C$ for some constant C. Let $\delta > 0$ be such that*

$$G = \{x \in X : f(x) \ge \delta\}$$

has positive measure. For any $T : X \to X$ that is measure-preserving, there exists E_T with $\mu(E_T) \ge \mu(G)/2$ such that the set

$$R_x = \{n \in \mathbf{N} : T^n f(x) \ge \delta\}$$

has upper density $\ge \mu(G)/2$ for every $x \in E_T$.

This is a variant of Lemma 1 from a blog post by Terry Tao, dated 10 February 2008. The proof we supply follows the same outline (with some details added).

Proof. First, note that $\int_X f\,d\mu \le C$, since X is a probability space. Next, notice that the for any $x \in X$ and any n, $T^n f(x) \ge \delta$ if and only if $f(T^n x) \ge \delta$, which is equivalent to $x \in T^{-n}G$. Because μ is shift-invariant, we have

$$\mu(G) = \int_X \frac{1}{N}\sum_{n=1}^{N} \chi_G\,d\mu = \int_X \frac{1}{N}\sum_{n=1}^{N} \chi_{T^{-n}G}\,d\mu. \tag{3}$$

The integrand on the right-hand side no more than 1, so the set

$$A_N = \left\{x \in X : \frac{1}{N}\sum_{n=1}^{N} \chi_{T^{-n}G}(x) \ge \frac{\mu(G)}{2}\right\}$$

has measure at least $\mu(G)/2$. For otherwise, we would have

$$\int_X \frac{1}{N}\sum_{n=1}^{N} \chi_{T^{-n}G}\,d\mu \le \int_{A_N} 1 + \int_{{A_N}^c} \frac{1}{N}\sum_{n=1}^{N} \chi_{T^{-n}G}\,d\mu < \frac{\mu(G)}{2} + \frac{\mu({A_N}^c)\mu(G)}{2} \le \mu(G), \tag{4}$$

contradicting (3). Note that

$$\limsup_{N\to\infty} A_N = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_N$$

is the set of all $x \in X$ that are in A_N for infinitely many N. Put another way, this is the set of all $x \in X$ for which

$$\limsup_{N\to\infty} \frac{R_x \cap [1, N]}{N} \ge \frac{\mu(G)}{2}.$$

Since $\mu(A_N) \ge \mu(G)/2$ for every N, we have

$$\mu\Big(\limsup_{N\to\infty} A_N\Big) = \mu\Big(\bigcap_{n=1}^{\infty} \bigcup_{n=N}^{\infty} A_n\Big) \ge \mu(G)/2 \tag{5}$$

and $E_T = \limsup_{N\to\infty} A_N$ is a set of positive measure such that for every point $x \in E_T$, R_x has upper density $\ge \mu(G)/2$. ▮

We can now formulate Furstenberg's multiple recurrence theorem:

Theorem F (*Furstenberg,* 1977). *Let $(X, \mathcal{F}, \mu, T)$ be a measure-preserving system and let f be a bounded measurable function such that $\int_X f\,d\mu > 0$. For any $k \in \mathbf{N}$, we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{s=1}^{N} \int_X f(T^s f) \cdots (T^{(k-1)s} f)\,d\mu > 0. \quad \blacksquare \tag{6}$$

In the particular case that $k = 2$, this is a form of Poincaré's recurrence theorem. We will not be able to prove Theorem F directly in these notes, but we will prove the important correspondence between Furstenberg's multiple recurrence and arithmetic progressions in $\mathbf{N}_0$.

Lemma E. *Theorems S and F are equivalent.*

Proof. First, suppose that Theorem S holds. Let $(X, \mathcal{F}, \mu, T)$ be a measure-preserving system, f a bounded measurable function with $\int_X f\,d\mu > 0$, and let $k \in \mathbf{N}$ be arbitrary. Let $\delta > 0$ be small enough that $G = \{x \in X : f(x) \geq \delta\}$ has positive measure; by Lemma D, there exists a set E_T of measure at least $\mu(G)/2$ such that $R_x = \{n \in \mathbf{N} : T^n f(x) \geq \delta\}$ has upper density $\geq \mu(G)/2$ for every $x \in E_T$. By Theorem S, there exists an N_0 such that for any $x \in E_T$, there is a k-term arithmetic progression in $R_x \cap [0, N_0]$. But the set of k-term arithmetic progressions in $[0, N_0]$ is finite, with some cardinality K; as a crude upper bound we have $K \leq \binom{N_0}{k}$. So there exists some fixed progression $P_T = \{a, a+r, \ldots, a+(k-1)r\} \subseteq [0, N_0]$ such that the set $E'_T = \{x \in E_T : P_T \subseteq R_x\}$ has $\mu(E'_T) \geq \mu(G)/(2K)$. We have

$$\begin{aligned} \int_X f(T^r f) \cdots (T^{(k-1)r} f)\,d\mu &\geq \int_{T^{-a}E'_T} f(T^r f) \cdots (T^{(k-1)r} f)\,d\mu \\ &= \int_{E'_T} T^a f(T^{a+r} f) \cdots (T^{a+(k-1)r} f)\,d\mu \\ &\geq \frac{\delta^k \mu(G)}{2K}. \end{aligned} \tag{7}$$

The progression P_T is particular to the transformation T, but the bound depends only on f and k, so in the argument above we could have applied Lemma D to T^m for $m \in \mathbf{N}$ and obtained (7) with T^m in place of T. Since r cannot exceed N_0 no matter what T^m is, we find that

$$\frac{1}{N_0} \sum_{s'=1}^{N_0} \int_X f(T^{ms'} f) \cdots (T^{(k-1)ms'})\,d\mu \geq \frac{\delta^k \mu(G)}{2K}, \tag{8}$$

where $m \in \mathbf{N}$ is arbitrary. For simplicity's sake, suppose that $N \geq {N_0}^2$ and averaging over all $1 \leq m \leq N/N_0$, we have

$$\frac{N_0}{N} \sum_{m=1}^{N/N_0} \frac{1}{N_0} \sum_{s'=1}^{N_0} \int_X f(T^{ms'} f) \cdots (T^{(k-1)ms'})\,d\mu \geq \frac{\delta^k \mu(G)}{2K}, \tag{9}$$

and since every $1 \leq s \leq N$ has at most $\min(N_0, N/N_0) = N_0$ representations of the form ms', where $1 \leq m \leq N/N_0$ and $1 \leq s' \leq N_0$, this gives

$$\frac{1}{N} \sum_{s=1}^{N} \int_X f(T^s f) \cdots (T^{(k-1)s})\,d\mu \geq \frac{\delta^k \mu(G)}{2KN_0} \tag{10}$$

for all large enough N, yielding Theorem F[*].

For the other direction, fix a subset $A \subseteq \mathbf{N}_0$ with positive upper density and let $k \in \mathbf{N}$. We can find a sequence of natural numbers $N_1, N_2, \ldots$ such that

$$\liminf_{j\to\infty} \frac{A \cap [0, N_j]}{N_j + 1} > 0.$$

[*] Thank you to Terry Tao, whose answer to my question on his blog helped me get the correct constants in the final bound.

Consider the linear space $\ell^\infty(\mathbf{N})$ of all bounded real-valued sequences under the supremum norm. By the Hahn-Banach theorem applied to the functional $\lambda(x) = \lim_{j\to\infty} x_j$, which is defined on the subspace of convergent sequences, we can construct a linear functional $\Lambda : \ell^\infty(\mathbf{N}) \to \mathbf{R}$ such that

$$\liminf_{j\to\infty} x_j \leq \Lambda(x) \leq \limsup_{j\to\infty} x_j. \tag{11}$$

Now let $X = \{0,1\}^{\mathbf{N}_0}$ be the space of all binary sequences indexed on the nonnegative integers with the product topology and Borel σ-algebra $\mathcal{F}$. Let T be the left-shift operator given by $Tx = (x_{n+1})_{n\in\mathbf{N}_0}$ and let $a \in X$ be given by

$$a_n = \begin{cases} 1, & \text{if } n \in A; \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

For any sequence of indices $s_1, \ldots, s_m \in \mathbf{N}_0$, define a measure $\mu_{s_1,\ldots,s_m}$ on $\{0,1\}^m$ given by

$$\mu_{s_1,\ldots,s_m}(E_1 \times \ldots \times E_m) = \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j} \chi_{E_1}(T^{s_1}a_i)\cdots\chi_{E_m}(T^{s_m}a_i)\right)_{j=1}^{\infty}\right), \tag{13}$$

where $E_1, \ldots, E_m \subseteq \{0,1\}$. Permuting the indices $s_1, \ldots, s_m$ does not change the value on the right-hand side (multiplication is commutative), and for any $s_{m+1} \in \mathbf{N}_0$,

$$\begin{aligned} \mu_{s_1,\ldots,s_m,s_{m+1}}(E_1 \times \ldots \times E_m \times \{0,1\}) &= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j} \chi_{E_1}(T^{s_1}a_i)\cdots\chi_{E_m}(T^{s_m}a_i)\chi_{\{0,1\}}(T^{s_{m+1}}a_i)\right)_{j=1}^{\infty}\right) \\ &= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j} \chi_{E_1}(T^{s_1}a_i)\cdots\chi_{E_m}(T^{s_m}a_i)\right)_{j=1}^{\infty}\right) \\ &= \mu_{s_1,\ldots,s_m}(E_1 \times \ldots \times E_m). \end{aligned} \tag{14}$$

So by the Kolmogorov consistency theorem, there exists a measure μ on X such that for all $s_1, \ldots s_m \in \mathbf{N}_0$, and $E_1, \ldots, E_m \subseteq \{0,1\}$,

$$\mu\big(\{x \in X : x_{s_1} \in E_1, \ldots, x_{s_m} \in E_m\}\big) = \mu_{s_1,\ldots,s_m}(E_1 \times \ldots \times E_m).$$

This is "almost" invariant under the shift T, since

$$\begin{aligned} \mu\big(T^{-1}\{x \in X : x_{s_1} \in E_1, \ldots, x_{s_m} \in E_m\}\big) &= \mu\big(\{x \in X : x_{s_1+1} \in E_1, \ldots, x_{s_m+1} \in E_m\}\big) \\ &= \mu_{s_1+1,\ldots,s_m+1}(E_1 \times \ldots \times E_m) \\ &= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j} \chi_{E_1}(T^{s_1+1}a_i)\cdots\chi_{E_m}(T^{s_m+1}a_i)\right)_{j=1}^{\infty}\right) \\ &= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=1}^{N_j+1} \chi_{E_1}(T^{s_1}a_i)\cdots\chi_{E_m}(T^{s_m}a_i)\right)_{j=1}^{\infty}\right). \end{aligned} \tag{15}$$

But as $j \to \infty$, shifting the terms over which we take the average does not change the lim sup or the lim inf, so this is the same as $\mu\big(\{x \in X : x_{s_1} \in E_1, \ldots, x_{s_m} \in E_m\}\big)$, provided that the sequence

$$\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j} \chi_{E_1}(T^{s_1}a_i)\cdots\chi_{E_m}(T^{s_m}a_i)\right)_{j=1}^{\infty}\right)$$

is convergent. This will be enough for our purposes (though there should be some way to fix it in general). Let B be the cylinder of all $x \in X$ with $x_0 = 1$. Note that

$$\begin{aligned}\mu(B) &= \mu\big(\{x \in X : x_0 = 1\}\big)\\ &= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j}\chi_{\{1\}}(a_i)\right)_{j=1}^{\infty}\right)\\ &= \Lambda\left(\left(\frac{A\cap[0,N_j]}{N_j+1}\right)_{j=1}^{\infty}\right)\\ &= \liminf_{j\to\infty}\frac{A\cap[0,N_j]}{N_j+1},\end{aligned} \tag{16}$$

and by our choice of the sequence N_j, this is positive and equals the lim sup. By Theorem F applied to k and the function χ_B, there exists some r such that

$$\mu(B\cap T^{-r}B\cap\cdots\cap T^{-(k-1)r}B) > 0.$$

Letting $A-s$ denote the set $\{a-s : a\in A\}$, and working backwards, we discover that

$$\begin{aligned}&\limsup_{j\to\infty}\frac{A\cap(A-r)\cap\cdots\cap(A-(k-1)r)\cap[0,N_j]}{N_j+1}\\ &\quad\geq \Lambda\left(\left(\frac{A\cap(A-r)\cap\cdots\cap(A-(k-1)r)\cap[0,N_j]}{N_j+1}\right)_{j=1}^{\infty}\right)\\ &\quad= \Lambda\left(\left(\frac{1}{N_j+1}\sum_{i=0}^{N_j}\chi_{\{1\}}(a_i)\chi_{\{1\}}(T^r a_i)\cdots\chi_{\{1\}}(T^{(k-1)r}a_i)\right)_{j=1}^{\infty}\right)\\ &\quad= \mu\big(\{x\in X : x_0 = 1,\, x_r = 1,\, \ldots,\, x_{(k-1)r} = 1\}\big)\\ &\quad= \mu(B\cap T^{-r}B\cap\cdots\cap T^{-(k-1)r}B),\end{aligned} \tag{17}$$

and since this is positive, there exists an arithmetic progression of length k in A. ▮

The first part of the presented proof draws from the aforementioned blog post of Terry Tao; the second part is a fleshed-out version of the proof sketch found in Tao and Vu (2006).

3. Generalisations

The techniques that Furstenberg used to link arithmetic progressions to recurrence phenomena came to be known as the Furstenberg correspondence principle and, as mentioned above, it led to various generalisations of Szemerédi's theorem. We list a few of them here, along with their equivalent statements in the domain of dynamical systems.

Theorem M (*Furstenberg, Katznelson,* 1979). *The following are equivalent.*

i) *Let $d \geq 1$ and let $A \subseteq \mathbf{Z}^d$ have positive upper density, that is,*

$$\limsup_{N\to\infty}\frac{A\cap[-N,N]^d}{(2N+1)^d} \geq 0.$$

For any $v_1,\ldots,v_k \in \mathbf{Z}^d$, there exist infinitely many pairs $(a,r)\in\mathbf{Z}^d$ such that $\{a+rv_1,\ldots,a+rv_k\}\subseteq A$.

ii) *Let $(X,\mathcal{F},\mu)$ be a probability space and $k\in\mathbf{N}$. If $T_1,T_2,\ldots,T_k : X\to X$ are measure-preserving maps that commute wieh each other and E is a set of positive measure, then there exists $r>0$ such that ${T_1}^r E\cap {T_2}^r E\cap\cdots\cap {T_k}^r E$ is nonempty.* ▮

This multidimensional version of Szemerédi's theorem is often called the constellation theorem because, roughly speaking, it asserts that a fixed finite constellation (up to translation and scaling) can be found in any set of points with positive density. A polynomial version of the theorem is also known:

Theorem P (*Bergelson, Leibman,* 1996). *The following are equivalent.*

i) *Let $P_1, \ldots, P_k : \mathbf{Z} \to \mathbf{Z}$ be polynomials such that $P_1(0) = \cdots = P_k(0) = 0$. Let $A \subseteq \mathbf{Z}^d$ have positive upper density. Then there exist infinitely many pairs $(a, r) \in \mathbf{Z}^d \times \mathbf{N}_0$ such that $a + P_1(r), \ldots, a + P_k(r) \in A$.*

ii) *Let $(X, \mathcal{F}, \mu)$ be a probability space, let $k \in \mathbf{N}$, and let $T_1, T_2, \ldots, T_k : X \to X$ be commuting measure-preserving maps. Let $P_1, \ldots, P_k : \mathbf{Z} \to \mathbf{Z}$ be polynomials such that $P_1(0) = \cdots = P_k(0) = 0$. There exists $r > 0$ such that*

$$T^{-P_1(r)}E + T^{-P_2(r)}E \cap \cdots \cap T^{-P_k(r)}E$$

is nonempty, where $T^{(a_1, \ldots, a_k)}$ is defined to be ${T_1}^{a_1} \cdots {T_k}^{a_k}$. ▮

Lastly, we have the density Hales-Jewett theorem, which tells us that subsets of large enough density must contain a combinatorial line.

Theorem D (*Furstenberg, Katznelson,* 1991). *Let $n \geq 1$ and $0 < \delta \leq 1$. There exists an integer $d \geq 1$ such that if A is any subset of $[0, n-1]^d$ of cardinality $|A| \geq \delta n^d$, then A contains a proper arithmetic progression $\{a, a+v, \ldots, a+(k-1)v\}$ for some $a \in [0, n-1]^d$ and $v \in [0, 1]^d$.* ▮

Theorem D also has a an ergodic counterpart, but it is rather complicated. For many years, the ergodic proof of the density Hales-Jewett theorem was the only one known. In 2009, in an online initiative called the Polymath Project spearheaded by W. T. Gowers, a group of over 40 mathematicians found a purely combinatorial proof of Theorem D. The results were published under the pseudonym D. H. J. Polymath and the success of this experiment led to over a dozen other online Polymath Projects.

References

Vitaly Bergelson and Alexander Leibman, "Polynomial extensions of van der Waerden's and Szemerédi's theorems," *Journal of the American Mathematical Society* **9** (1996), 725–753.

Paul Erdős and Paul Turán, "On some sequences of integers," *Journal of the London Mathematical Society* **11** (1936), 261–264.

Hillel Furstenberg, "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions," *Journal d'Analyse Mathématique* **31** (1977), 204–256.

Hillel Furstenberg and Yitzhak Katznelson, "An ergodic Szemerédi theorem for commuting transformations," *Journal d'Analyse Mathématique* **34** (1979), 275–291.

Hillel Furstenberg and Yitzhak Katznelson, "A density version of the Hales-Jewett theorem," *Journal d'Analyse Mathématique* **57** (1991), 64–119.

William Timothy Gowers, "A new proof of Szemerédi's theorem for arithmetic progressions of length four," *Geometric and Functional Analysis* **8** (1998), 529–551.

William Timothy Gowers, "A new proof of Szemerédi's theorem," *Geometric and Functional Analysis* **11** (2001), 465–588.

D. H. J. Polymath, "A new proof of the density Hales-Jewett theorem," *Annals of Mathematics* **175** (2012), 1283–1327.

Klaus Friedrich Roth, "On certain sets of integers," *Journal of the London Mathematical Society* **28** (1953), 104–109.

Endre Szemerédi, "On sets of integers containing no k elements in arithmetic progression," *Acta Arithmetica* **27** (1975), 199–245.

Terence Tao and Van Ha Vu, *Additive Combinatorics* (Cambridge: Cambridge University Press, 2006).

Bartel Leendert van der Waerden, "Beweis einer Baudetschen Vermutung," *Nieuw Archief voor Wiskunde* **15** (1927), 212–216.