BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 43, Number 1, Pages 95–100
S 0273-0979(05)01101-8
Article electronically published on November 17, 2005

MR0245555 (39 #6861) 10.68

Szemerédi, E.

On sets of integers containing no four elements in arithmetic progression.

Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.

If a sequence A of integers contains no k (distinct) terms in arithmetic progression, we say that A is k-free. Let $\tau_k(n)$ denote the maximal number of elements in a k-free $A \subseteq [0, n]$; then it can be proved that $\lim_{n\to\infty} \tau_k(n)/n = \gamma_k$ exists. From elementary considerations $\gamma_k \leq 1 - 1/k$, and $\gamma_3 \leq \gamma_4 \leq \gamma_5 \leq \cdots$; and an old result of Behrend showed that either all the γ_k are 0, or $\gamma_k \to 1$ as $k \to \infty$. The question of deciding between these alternatives is notoriously difficult, and far from settled. K. F. Roth brought the subject alive when he proved, in 1953 [J. London Math. Soc. **28** (1953), 104–109; MR0051853 (14,536g); errata, MR **14**, p. 1278; ibid. **29** (1953), 20–26; MR0057894 (15,288f)], that $\gamma_3 = 0$ (he actually showed that $\tau_3(n) \ll n/\log\log n$); and the author now shows, by a combinatorial tour de force, that $\gamma_4 = 0$. The reviewer would like to do justice to this brilliant achievement by giving within a short space an outline of the method, but has to confirm that this is entirely beyond his powers. Suffice it to say that the argument is technically elementary and uses van der Waerden's famous theorem: if $u \geq u_0(l, k)$, then in any partition of $[0, u]$ into at most l classes, at least one class contains k terms in arithmetic progression. It would be interesting to eliminate dependence on the latter result (Roth did not need it in his analytic approach).

(From MathSciNet, October 2005)

H. Halberstam

MR0670131 (84b:28016) 28D05 10L20

Furstenberg, H.; Katznelson, Y.; Ornstein, D.

The ergodic theoretical proof of Szemerédi's theorem.

Bull. Amer. Math. Soc. (*N.S.*) **7** (1982), *no.* 3, 527–552.

In 1975 E. Szemerédi settled a long standing conjecture of P. Erdős and P. Turán by proving Theorem I: If Λ is a subset of the integers, of positive upper density, then Λ contains arbitrarily long arithmetic progressions [Acta Arith. **27** (1975), 199–245; MR0369312 (51 #5547)]. In 1977, Furstenberg discovered a deep and unexpected connection of this result with ergodic theory; in particular, he showed that Theorem I is equivalent to the following Theorem II: Let $(X, \mathcal{B}, \mu)$ be a probability measure space, let T be an invertible, measure-preserving transformation on $(X, \mathcal{B}, \mu)$ and let $A \in \mathcal{B}$ be a set of positive measure. Then for any positive integer k, there exist a subset $B \subset A$ with $\mu(B) > 0$ and an integer $n \geq 1$ such that $\mu(\bigcap_{j=0}^{k-1} T^{-jn} A) > 0$ [J. Analyse Math. **31** (1977), 204–256; MR0498471 (58 #16583)].

The aim of the present paper, in the authors' words, is "to give an exposition, as widely accessible as possible, of the ergodic theoretic proof of Theorem I". They have succeeded in their goal admirably. The high quality and the clarity of the exposition are perhaps best shown by the following excerpt in which the authors discuss the outline and the general philosophy of the proof:

"Theorem II is valid for all measure-preserving systems but not for the same reason. There are two distinct phenomena, mutually exclusive, which account for the existence of positive measure intersections of the form $\bigcap_{j=0}^{k-1} T^{-jn}A$. One, compactness, is seen in the case of group rotations (T being a translation by a group element on the Haar measure space of a compact group), where for appropriate values of n, T^n is 'close' to the identity so that $T^{jn}A$ differs from A by very little, for $0 < j < k$, and $\mu(\bigcap_{j=0}^{k-1} T^{-jn}A)$ is very close to $\mu(A)$. The other phenomenon is that of weak mixing, defined by the condition that for every set A, $\mu(A \cap T^{-n}A) \sim \mu(A)^2$ for most n. It can be proved (cf. Section 3) that $\mu(\bigcap_{j=0}^{k-1} T^{-jn}A) \sim \mu(A)^k$ for most values of n.

"It is not true that these two phenomena are complementary; there exist systems $(X, \mathcal{B}, \mu, T)$ which are neither weakly mixing nor group rotations. However, if $(X, \mathcal{B}, \mu, T)$ is not weakly mixing then there exists a nontrivial T-invariant sigma-algebra $\mathcal{B}_1 \subset \mathcal{B}$ such that T restricted to $\mathcal{B}_1$ acts like a group rotation, so that in any case the assertion of Theorem II is valid for all the sets A in some T-invariant sigma-algebra of $\mathcal{B}$. The strategy of the proof of Theorem II is (a) to show that there exists a T-invariant subsigma-algebra $\mathcal{B}_1 \subset \mathcal{B}$ which is maximal, with respect to inclusion, in the class of T-invariant subalgebras of $\mathcal{B}$ for which the statement of Theorem II is valid. (b) Assuming $\mathcal{B}_1 \neq \mathcal{B}$, study the behavior of sets $A \in \mathcal{B}$ under T 'relative to $\mathcal{B}_1$' and show that either we have 'relative weak mixing' or else there exist $\mathcal{B}_2 \supset \mathcal{B}_1$ for which the action of T is 'relatively compact'. In either case we show that there exists a bigger subalgebra for which the statement of Theorem II is valid, contradicting the maximality of $\mathcal{B}_1$. This implies $\mathcal{B}_1 = \mathcal{B}$ and completes the proof."

(From MathSciNet, October 2005)

V. Drobot

MR1844079 (2002k:11014) 11B25 11K38 11K45

Gowers, W. T.

A new proof of Szemerédi's theorem.

Geom. Funct. Anal. **11** (2001), *no.* 3, 465–588.

The present paper represents a major contribution to the literature surrounding E. Szemerédi's theorem on arithmetic progressions. It will be recalled that Szemerédi's theorem [Acta Arith. **27** (1975), 199–245; MR0369312 (51 #5547)] confirmed the conjecture of Erdős and Turán dating back to 1936, asserting that for a given length l and density $\delta > 0$, there is an $L(l, \delta)$ such that if $L \geq L(l, \delta)$, any subset $A \subset \{1, 2, 3, \dots, L\}$ with more than δL elements will contain an arithmetic progression of length l. The point of this conjecture was to "explain" the phenomenon in van der Waerden's classic theorem: $\exists W(l, r)$ such that if $W \geq W(l, r)$ and $\{1, 2, 3, \dots W\}$ is the union of r subsets, then one of these necessarily contains an l-term arithmetic progression. Quite likely Erdős and Turán also hoped to obtain a good estimate of $W(l, r)$ through their formulation, since $W(l, r)$ can be taken as $L(l, 1/r)$.

The Erdős-Turán conjecture was first established by K. F. Roth for 3-term arithmetic progressions [J. London Math. Soc. **28** (1953), 104–109; MR0051853 (14,536g)], then by Szemerédi for 4-term arithmetic progressions [Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104; MR0245555 (39 #6861)], and finally Szemerédi proved the conjecture in full [op. cit., 1975]. The present paper gives a new proof of this theorem, providing at the same time an upper bound to the function $L(l,\delta)$. One consequence is a significant improvement of the upper bound for $W(l,r)$; for example, for some constant c, $W(4,r)$ can be taken as $\exp(\exp(r^c))$. Generally we can take $L(l,\delta)=\exp(\delta^{-c(l)})$ where $c(l)=2^{2^{l+9}}$. These estimates fall short of providing a proof to another conjecture of Erdős: any set $A\subset\mathbb{N}$ with $\sum_{a\in A}\frac{1}{a}=\infty$ contains arbitrarily long arithmetic progressions. Nonetheless they constitute a quantum jump over previous results.

After Szemerédi's combinatorial proof appeared, this reviewer showed how Szemerédi's result would follow from an ergodic-theoretic statement in the spirit of Poincaré's "recurrence theorem". A "multiple recurrence" result was subsequently proved [H. Furstenberg, J. Analyse Math. **31** (1977), 204–256; MR0498471 (58 #16583)], and this provided a second proof for the theorem of Szemerédi. These methods could be extended to obtain results that have so far not been accessible by other means. For example, a multi-dimensional version for subsets of $\mathbb{Z}^m$ can be proved [H. Furstenberg and Y. Katznelson, J. Analyse Math. **34** (1978), 275–291 (1979); MR0531279 (82c:28032)]. V. Bergelson and A. Leibman have obtained a polynomial extension; thus a set of integers of positive density contains a k-tuple $a,a+m,a+m^2,\dots,a+m^{k-1}$ [J. Amer. Math. Soc. **9** (1996), no. 3, 725–753; MR1325795 (96j:11013)]. Also, a density version relating to a theorem of Hales and Jewett in the same way that Szemerédi's theorem relates to that of van der Waerden was established by Katznelson and the reviewer [J. Anal. Math. **57** (1991), 64–119; MR1191743 (94f:28020)]. However, the ergodic-theoretic approach depends essentially on passing to a limit whereby a set $\{1,2,3,\dots,N\}$ is replaced by a measure space, and the translations $n\to n+a$ are replaced by measure preserving transformations of this space. In passing to this limit one loses sight of the size N of the interval $\{1,2,3,\dots,N\}$. As a result this approach is incapable of giving any information regarding $L(l,\delta)$ beyond the fact that it is finite. It now remains a challenge to combine Gowers' analytic-combinatorial apparatus with ergodic-theoretic techniques to obtain precise finite versions of the various extensions of Szemerédi's result.

We will comment later on the connections between ideas in Gowers' paper and the ergodic-theoretic approach to these questions. First let us outline the strategy adopted by Gowers in his proof. Basic to the argument is the idea that a sufficiently "random" subset of given density in $\{1,2,3,\dots,N\}$ will, for large N, be expected to contain many l-term arithmetic progressions. The key to the present proof is to formulate a notion of randomness for which this heuristic argument can be made precise for the existence of arithmetic progressions of a given length. The author presents such a notion and calls it "α-uniformity of degree d". He then shows that if $A\subset\{1,\dots,N\}$ has α-uniformity of degree d for N large and α is close to 0, the number of $(d+2)$-term arithmetic progressions in the set deviates only slightly from the expected number, and will therefore be positive. What is crucial here is that if the given subset of density δ in $\{1,2,3,\dots,N\}$ is *not* α-uniform of degree d, this too can be exploited by showing that, restricted to a substantial arithmetic progression in $\{1,2,3,\dots\}$, the relative density of the set A can be increased appreciably. One then iterates this argument, and if N was chosen sufficiently large, by repeated iteration one can push up the density to exceed $1-\frac{1}{l}$, in which case the existence of l-term arithmetic progressions is obvious. In effect a similar strategy was used by Roth for the case of 3-term arithmetic progressions; in this case the appropriate notion is α-regularity of degree 1. For $d = 2$ new ideas are called for, and for $d > 2$, still further refinements are needed. Gowers treats these cases separately, thus motivating the successively more complex arguments needed in his proof, and making the final arguments as transparent as possible.

The setting for the author's definition of regularity is that of a function $f(n)$ defined for $n \in Z_N = \mathbb{Z}/N\mathbb{Z}$ with $|f(n)| \le 1$. To study a set $A \subset Z_N$ we take $f_A(n) = 1_A(n) - \delta$ where δ is the density $|A|/N$. For degree 1 the notion of uniformity has to do with $f(n)$ being uncorrelated with periodic functions. A measure of the lack of uniformity is the "average autocorrelation"

$$\frac{1}{N^3}\sum_x |\sum_y f(y)\overline{f(y-x)}|^2,$$

the sums being taken over Z_N. This can be rewritten as

$$N^{-3}\sum_{x,a,b} f(x)\overline{f(x+a)}\ \overline{f(x+b)}f(x+a+b).$$

This suggests defining higher order autocorrelation by averaging products of $f(x)$ and $\overline{f}(x)$ taken along "cubes of dimension $d+1$": $\{x + \varepsilon_1 a_1 + \varepsilon_2 a_2 + \cdots + \varepsilon_{d+1}a_{d+1} : \varepsilon_j = 0, 1\}$ (f or $\overline{f}$ being chosen according as an even or odd number of a_i appear). We then say f is *α-uniform of degree d* if the resulting autocorrelation is in absolute value $\le \alpha$. α-uniformity of degree $d+1$ implies $\sqrt{\alpha}$-uniformity of degree d, and α-uniformity for small α becomes increasingly restrictive as the degree grows. Thus if ω is a primitive Nth root of 1, the function $f(n) = \omega^{n^2}$ is $2/N$-uniform of degree 1 but not α-uniform of degree 2 for any $\alpha < 1$. An analogous statement holds for $f(n) = \omega^{p(n)}$ where $p(t)$ is a polynomial of degree d with leading coefficient relatively prime to N. The functions $f(n)$ will be α-uniform of degree $d-1$ with small α, but non-uniform of degree d. The relevance of polynomial exponential expressions to the counting of l-term arithmetic progressions is not surprising, since for any polynomial $p(t)$ of degree $< d$ one has the identity

$$\sum_{j=0}^{d}(-1)^j\binom{d}{j}p(x+jy) \equiv 0,$$

and thus the set of n for which $e^{2\pi i p(n)}$ takes values in a particular interval will contain more $(d+1)$-term arithmetic progressions than a random set of the same density. With this in mind it might be expected that it is precisely a positive correlation with such polynomial exponential expressions that is responsible for a bias in counting l-term arithmetic progressions. It can be shown however that orthogonality of $1_A(n)$ to such functions is in fact not enough to ensure the pseudorandom behavior sought, and Gowers' achievement is to have put his finger on the "right" concept, namely, α-uniformity of the appropriate degree.

The bulk of the paper is directed to drawing the implications of the *absence* of α-uniformity of degree d. The author here is guided by a well-known result of G. A. Freiman [*Foundations of a structural theory of set addition*, Translated from the Russian, Amer. Math. Soc., Providence, R. I., 1973; MR0360496 (50 #12944)] and a variant due to A. Balog and Szemerédi [Combinatorica **14** (1994), no. 3, 263–268; MR1305895 (95m:11019)]. Gowers generously acknowledges his debt to these authors as well as to I. Z. Ruzsa, who provided a simplified proof of Freiman's theorem. The Balog-Szemerédi Theorem states that if $S \subset \mathbb{Z}^r$ has n elements and the number of solutions to $a - b = c - d$ with $a, b, c, d \in S$ exceeds cn^3 for some $c > 0$, then S meets a "generalized arithmetic progression" in at least $c'n$ points, with c' depending only on c. The significance of this theorem is that it serves as a model for obtaining linear structure from information regarding the autocorrelation of a function. When $S \subset \mathbb{Z}^2$ is the graph of a function from $\mathbb{Z}$ to $\mathbb{Z}$ it can be used to show under certain conditions that the function, when properly restricted, is in fact linear. This type of result ultimately becomes a tool in showing that the lack of uniformity of degree d as attested to by higher order autocorrelation leads to a lack of uniformity of degree $d - 1$ along a subset and eventually for the function $f_A(n)$ this leads to increased density relative to some long arithmetic progression.

Gowers' notion of α-uniformity of a given degree provides, though not in a precise fashion (on account of the parameter α), a hierarchy of levels of pseudo-randomness appropriate for counting arithmetic progressions. A rather similar hierarchy appears in the ergodic-theoretic approach to Szemerédi's Theorem. This approach is based on the equivalence of the combinatorial result with what may be called the ergodic Szemerédi theorem: If $(X, \mathcal{B}, \mu)$ is a probability space and $T\colon X \to X$ is a measure preserving transformation, and if $A \in \mathcal{B}$ with $\mu(A) > 0$, then for any l there are points $x \in A$ such that for some $n \geq 1$, $T^n x \in A, T^{2n}x \in A, \dots, T^{(l-1)n}x \in A$. This involves studying the behavior of expressions of the form $f_0(x)f_1(T^n x)f_2(T^{2n}x)\cdots f_k(T^{kn}x)$, and indeed an ergodic theorem can be proved for averages of these which establishes the desired result. Here too there is the "easy" case when the $(k+1)$-tuple $x, T^n x, T^{2n}x, \dots, T^{kn}x$ behaves "randomly", and the ergodic average of the foregoing product is $\int f_0 d\mu \cdot \int f_1 d\mu \dots \int f_k d\mu$. In general there may be obstructions to this randomness, and these appear in the dynamical system in the form of special "characteristic factors" [H. Furstenberg and B. Weiss, in *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, 193–227, de Gruyter, Berlin, 1996; MR1412607 (98e:28019)]. In this context uniformity of degree d (without α) of a function $f(x)$ on X can be defined as orthogonality to all the functions that arise from the "characteristic factor of level d". One example of a characteristic factor of level 2 is the toral transformation $T\colon (\theta, \varphi) \to (\theta + \alpha, \varphi + 2\theta + \alpha)$. If $f(\theta, \varphi) = e^{2\pi i \varphi}$ then $f(T^n(\theta, \varphi)) = e^{2\pi i n^2 \alpha} e^{4\pi i n \theta} e^{2\pi i \varphi}$, which exhibits quadratic exponential behavior. That these are not, as was pointed out earlier, the only obstructions to uniformity can be predicted on the basis of the appearance of other nil-systems as characteristic factors. Specifically if N is a k-step nilpotent group and Γ a lattice in N, and a measure preserving transformation is defined on N/Γ by $Tx = ax$ for $a \in N$, then there are algebraic constraints on $x, T^n x, T^{2n}x, \dots, T^{(k+1)n}x$, and for a typical function $f(x)$ on N/Γ, the sequences $f(T^n x)$ will not be α-uniform of degree k_0 if α is small. Thus there is a rather strong parallel in the structures underlying both Gowers' proof and the ergodic-theoretic proof of Szemerédi's theorem. The parallel ends here because Gowers handles the non-uniform case by very subtle combinatorial means, whereas in the ergodic-theoretic approach the non-uniform situation can be reduced to an analysis of very special dynamical systems, the so-called characteristic systems. Nonetheless, Gowers' paper will be of significance to ergodic theorists as well as combinatorialists and number theorists.

The paper is a model of lucidity, with each new idea clearly motivated by showing why the apparatus that works in easier cases is inadequate for the general case. The author's efforts at clarity make this a profound, yet readable, paper.

(From MathSciNet, October 2005)

Hillel Furstenberg