Fourier Analysis and Szemerédi's Theorem

W. T. Gowers

Abstract. The famous theorem of Szemerédi asserts that for every positive integer k and every positive real number $\delta > 0$ there is a positive integer N such that every subset of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k. A second proof of the theorem was given by Furstenberg using ergodic theory, but neither this proof nor Szemerédi's gave anything other than extremely weak information about the dependence of N on k and δ. In this article we describe a new, more quantitative approach to Szemerédi's theorem which greatly improves the best known bound when $k = 4$, and which will probably do the same for general k.

1991 Mathematics Subject Classification: 11P99
Keywords and Phrases: Arithmetic progressions, Fourier analysis

§1. Introduction.

A well known result of van der Waerden [vdW], published in 1927, is the following.

Theorem 1A. *Let the natural numbers be partitioned into finitely many sets. Then one of the sets contains arbitrarily long arithmetic progressions.*

A straightforward compactness argument allows this statement to be rephrased as follows.

Theorem 1B. *For every pair of positive integers k, r there exists a positive integer M such that, whenever the set $\{1, 2, \ldots, M\}$ is partitioned into r subsets $C_1, \ldots, C_r$, at least one of the subsets contains an arithmetic progression of length k.*

This is one of the classic results of Ramsey theory: it is customary to call the cells of the partition *colours*, the partition itself an *r-colouring* and the resulting arithmetic progression *monochromatic*.

Let us define $M(k, r)$ to be the minimal M for which the conclusion of Theorem 1B holds. A compactness argument proves that $M(k, r)$ is finite but does not give any bound for it. As it happens, though, van der Waerden proved the second version of his theorem directly, and it is possible to extract from his proof an explicit estimate for $M(k, r)$. However, the estimate is enormously large, as we shall see later, and barely qualifies as a quantitative bound.

In 1936, Erdős and Turán [ET] made a conjecture which significantly strengthened van der Waerden's theorem. It soon became clear that their conjecture was very difficult, and it took almost forty years before it was solved, by Szemerédi [Sz2]. The statement is the following.

Theorem 2A. *Any subset of the natural numbers with positive upper density contains arithmetic progressions of arbitrary length.*

Again, there is a finite version.

Theorem 2B. *For every natural number k and positive real number δ there exists a natural number N such that every subset of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k.*

This certainly implies van der Waerden's theorem, as one can take $\delta = r^{-1}$ and consider the most frequently occurring colour. For this reason, the result is often called the *density version* of van der Waerden's theorem (as opposed to the *colouring version*).

It is interesting to consider why Erdős and Turán made their conjecture, and to compare it with other results in Ramsey theory. Ramsey's theorem itself states that for every k and r there exists N such that if the edges of the complete graph on N vertices are coloured with r colours, then a complete subgraph on k vertices can be found with all its edges the same colour. However, it is absolutely not true that one can do this with the most frequently occurring colour. (For example, consider a complete bipartite graph on two sets of equal size.) A theorem of Schur states that if N is sufficiently large and the set $\{1, 2, \ldots, N\}$ is coloured with r colours, then one of the colours contains a triple (x, y, z) with $x + y = z$. Again, there is no density version of the statement – just consider the set of all odd numbers less than N. The most important difference between van der Waerden's theorem and Schur's theorem in this respect is that van der Waerden's theorem is affine-invariant. This property rules out simple counterexamples such as the set of all integers satisfying some congruence.

This shows why the conjecture had a chance of being true, but the motivation for it was stronger than that. In particular, it was reasonable to think that it would not be possible to prove the conjecture using the sorts of inefficient combinatorial arguments that yielded poor bounds for van der Waerden's theorem. In that case, a proof of the conjecture would give new quantitative information even for the colouring statement. Moreover, if the bounds turned out to be good enough, one could obtain an important number-theoretic result purely combinatorially. To be precise, Erdős went on to give the following conjecture, which was possibly his favourite of all problems.

Conjecture 3. *Let A be a set of natural numbers such that $\sum_{n \in A} n^{-1} = \infty$. Then A contains arithmetic progressions of arbitrary length.*

This conjecture, if true, would imply that the primes contained arbitrarily long arithmetic progressions, and the proof would use very little about the distribution of primes – Chebyshev's theorem would suffice. However, Szemerédi's proof used van der Waerden's theorem, so, although it was a major breakthrough, it did not after all provide improved bounds, and indeed Conjecture 3 is still wide open, even for progressions of length three.

A different sort of breakthrough was made by Furstenberg [Fu] in 1977, who gave a second proof of Szemerédi's theorem, which used ergodic theory (much of which was new, fascinating and specially developed by Furstenberg for the purpose). Furstenberg's methods have since been extended, and there are now several purely combinatorial results for which the only known proofs use ergodic theory. Some of these will be discussed later in this paper. However, the ergodic theory method as it stands does not give any estimates and so in particular gives no information about Conjecture 3.

Let us now consider in more detail the best known bounds for this class of problems. In order to state them, it will be necessary to remind the reader of the *Ackermann hierarchy* of rapidly-growing functions, defined as follows. Let $A_1(n) = 2 + n$, $A_2(n) = 2n$ and $A_3(n) = 2^n$. In general one obtains $A_k(n)$ by starting with the number 2 and applying the function A_{k-1} $n-1$ times. In other words, each function iterates the previous one. A concise definition is

$$A_1(n) = 2 + n;\ A_k(1) = 2\ (k > 1);\ A_k(n) = A_{k-1}(A_k(n-1))\ (n > 1).$$

Note in particular that $A_4(n)$ is given by a tower of twos of height n, while $A_5(n)$ is given by a tower of twos of height $A_5(n-1)$.

The *Ackermann function* itself is defined as $A(n) \equiv A_n(n)$. Thus, it grows faster than any individual function A_k. In fact, it is known to grow faster than any primitive recursive function, which very roughly means any function that can be defined starting with the successor function and using a finite sequence of single inductive definitions (rather than the double induction we needed above). Nevertheless, this function does from time to time appear naturally (for a very good example see Ron Graham's account in this volume of the work of Peter Shor) and was the upper bound obtained by van der Waerden for the function $M(k, 2)$, that is, the smallest M such that every 2-colouring of $\{1, 2, \dots, M\}$ yields a monochromatic arithmetic progression of length k. (One might reasonably suppose that this was about the worst bound that could arise from *any* sensible proof of a natural combinatorial statement. If you believe this, then see [PH] or [GRS Section 6.3].) This remained the best known upper bound until 1987, by which time some people had even been tempted to wonder whether there was a comparable lower bound, although the best known lower bound was only exponential in k. Then Shelah [Sh] found a primitive recursive upper bound for $M(k, 2)$ of $A_5(k)$. To everybody's surprise, his argument was very natural, not especially difficult and in much the same spirit as that of van der Waerden. This, needless to say, did not stop it being highly ingenious.

Since Szemerédi used van der Waerden's theorem in the middle of an inductive step, one can guess that his argument, when combined with Shelah's later bound, gave an upper bound for $N(k, \delta)$ of the general form of $A_6(k)$ (for fixed δ), but this has not been checked. Despite this bound being a huge improvement on the Ackermann function, it still had the flavour of a bound that just happened to come out of a not particularly quantitative argument. Moreover, to improve it, it was clear that a substantially new proof would be necessary, one which avoided the use of van der Waerden's theorem. (Another important tool in Szemerédi's proof, his so-called uniformity lemma, also makes a big contribution. See [G1] for a proof that the function A_4 can occur in nature.)

Fortunately, there was one result in the area which was undeniably quantitative, a proof in 1953 by Roth [R1] that $N(3,\delta)$ is at most $\exp\exp(C/\delta)$ for some absolute constant C. This result was proved using Fourier analysis, and the proof will be sketched below. On the other hand, the argument did not seem to generalize to progressions of length greater than three, for reasons which will also be sketched below. Indeed, Szemerédi was able to make further progress only after he had found a different proof for progressions of length three. Even so, the existence of Roth's proof suggested that Fourier analysis (or exponential sums – they are the same in this context) ought to be used as the basis for any significant improvement to the bounds in Szemerédi's theorem. In this paper, we shall indicate how to use it for progressions of length four. More details can be found in [G2]. (It should be remarked that Roth [R2], using ideas of Szemerédi, found a proof in this case which used exponential sums, but this proof was not purely analytic. In particular it still required van der Waerden's theorem.)

§2. Roth's argument.

Let N be a prime (for convenience) and write $\mathbb{Z}_N$ for $\mathbb{Z}/N\mathbb{Z}$, the integers mod N. Let ω be the primitive N^{th} root of unity $\exp(2\pi i/N)$. Given a function $f:\mathbb{Z}_N\to\mathbb{C}$, one can define a discrete Fourier transform $\widetilde{f}$ by the formula

$$\widetilde{f}(r)=\sum_{s\in\mathbb{Z}_N}f(s)\omega^{-rs}\ .$$

One then has the inversion formula

$$f(s)=N^{-1}\sum_{r\in\mathbb{Z}_N}\widetilde{f}(r)\omega^{rs}\ ,$$

while Parseval's identity takes the form

$$\sum_{r\in\mathbb{Z}_N}|\widetilde{f}(r)|^2=N\sum_{s\in\mathbb{Z}_N}|f(s)|^2\ .$$

Since the Fourier transform is in some sense measuring periodicity, it is not surprising that it should be useful for problems to do with arithmetic progressions. Roth's argument starts with the observation (standard to analytic number theorists) that it gives a neat way of counting arithmetic progressions of length three. Consider three subsets A,B,C of $\mathbb{Z}_N$, and identify these sets with their characteristic functions. Then the number of triples (x,y,z) such that $x\in A$, $y\in B$, $z\in C$ and $x+z=2y$ (this last condition states that (x,y,z) is an arithmetic progression mod N) is

$$N^{-1}\sum_r\sum_{x,y,z}A(x)B(y)C(z)\omega^{-r(x-2y+z)}\ .$$

(Here, and from now on, all sums where the range is unspecified are over the whole of $\mathbb{Z}_N$.) To see this, notice that $\sum_r\omega^{-r(x-2y+z)}$ is zero when $x-2y+z\neq 0$, and otherwise N, while $A(x)B(y)C(z)$ is 1 if $x \in A$, $y \in B$ and $z \in C$, and otherwise zero.

Now $\omega^{-r(x-2y+z)} = \omega^{-rx}\omega^{2ry}\omega^{-rz}$, so the expression above is nothing other than

$$N^{-1}\sum_r \widetilde{A}(r)\widetilde{B}(-2r)\widetilde{C}(r)\ .$$

Notice that $\widetilde{A}(0)$ is just the cardinality of the set A, and similarly for B and C, so we can split this up as

$$N^{-1}|A||B||C| + \sum_{r\neq 0} \widetilde{A}(r)\widetilde{B}(-2r)\widetilde{C}(r)\ . \qquad (*)$$

If A, B and C have cardinalities αN, βN and γN respectively, then the first term equals $\alpha\beta\gamma N^2$. Notice that this is exactly the number of triples one would expect to have satisfying the conditions above if the sets A, B and C had been chosen randomly with their given cardinalities, since there are N^2 triples (x, y, z) in arithmetic progression mod N, and the probability that an individual one lies in $A \times B \times C$ is $\alpha\beta\gamma$. Of course, in general A, B and C are not chosen randomly, and $A \times B \times C$ may well contain no arithmetic progression mod N, but this can happen only if the first term is cancelled out by the second, and *this* can happen only if some of the non-zero Fourier coefficients of A, B and C are large.

One applies this argument as follows. Let $A \subset \{0, 1, 2, \ldots, N-1\}$ be a set of cardinality αN containing no arithmetic progression of length three. Let B and C both equal $\{x \in A : N/3 < x < 2N/3\}$. Now regard A, B and C as subsets of $\mathbb{Z}_N$ in the obvious way. Notice that if $(x, y, z) \in A\times B\times C$ and $x+z = 2y$ mod N, then either $x = y = z$ or (x, y, z) corresponds to an arithmetic progression in the original set A when it was *not* regarded as a subset of $\mathbb{Z}_N$. As explained in the previous paragraph, the fact that $A \times B \times C$ contains no arithmetic progressions mod N (apart from the degenerate ones of the form (x, x, x), but there are too few of these to be significant) implies that A, B and C have large non-zero Fourier coefficients. More precisely, it is not hard to deduce from $(*)$ that if $B = C$ has cardinality at least $\alpha N/4$, then there must exist a non-zero r such that $|\widetilde{A}(r)| \geq \alpha^2 N/20$. (If $|B| < \alpha N/4$, then A is not uniformly distributed inside $\mathbb{Z}_N$, and a similar but stronger conclusion is true.)

A good way to view the argument so far is to regard the size of the largest non-zero Fourier coefficient of A as a measure of non-randomness. Then what we have shown (or rather sketched) is that either A is random, in which case it contains plenty of arithmetic progressions of length three, just as one would expect, or it is non-random, in which case it has a non-zero Fourier coefficient which is large, where "large" means exceeding γN for some constant $\gamma > 0$ that depends only on the density α of the set A.

We must now deal with the second case, so suppose that $r \neq 0$ and $|\widetilde{A}(r)| \geq \gamma N$. Let m be a sufficiently large integer (depending on α only) and define, for $1 \leq j \leq m$, the set P_j to be $\{s \in \mathbb{Z}_N : (j-1)N/m \leq rs < jN/m\}$. The sets P_j have been chosen so that the function $s \mapsto \omega^{-rs}$ is roughly constant on each P_j.

Let s_j be an arbitrary element of P_j. Then

$$\widetilde{A}(r) = \sum_s A(s)\omega^{-rs} = \sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs}$$

is well approximated by

$$\sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs_j} = \sum_{j=1}^{m} |A\cap P_j|\omega^{-rs_j}\ .$$

Since the numbers ω^{-rs_j} are evenly spread around the unit circle, this sum cannot be large unless the sets $|A\cap P_j|$ have widely differing sizes, and because we know that $\sum_{j=1}^m |A\cap P_j| = \alpha N$, this implies that there exists j such that $|A\cap P_j| \geq (\alpha+\gamma')|P_j|$, where γ' again depends on α only.

Now P_j is nothing other than an arithmetic progression mod N with common difference r^{-1}. If the argument above is done carefully, then the size of P_j can be made proportional to N (with a constant depending on α only) and $|A\cap P_j| \geq (\alpha+c\alpha^2)|P_j|$, where c is an absolute constant. The final ingredient is a simple and standard argument, based on Dirichlet's pigeonhole principle, which shows that the set P_j can be partitioned into r sets, with r proportional to $\sqrt{N}$, which are not only arithmetic progressions mod N but are still arithmetic progressions when regarded as subsets of $\{0,1,\dots,N-1\}$. Then, by an averaging argument, we can find one of these, Q say, such that $|A\cap Q| \geq (\alpha+c\alpha^2)|Q|$ and Q has size proportional to $\sqrt{N}$.

The proof is now over, because we have managed to find a subprogression of $\{0,1,\dots,N-1\}$ inside which the density of the set A has gone up from α to $\alpha(1+c\alpha)$. We can then repeat the argument. A small calculation shows that we cannot repeat it more than C/α times, where C is another absolute constant, and another small calculation gives the bound $N(3,\delta) \leq \exp\exp(C/\delta)$, the double exponential coming from the fact that at each iteration we are taking the square root of N.

§3. Progressions of length four.

One could summarize Roth's proof as follows. If a set $A\subset\mathbb{Z}_N$ (or more accurately its characteristic function) has no large non-trivial Fourier coefficients, then it behaves randomly in a useful sense. In particular, it contains roughly the right number of arithmetic progressions of length three. On the other hand, if it has a large non-trivial Fourier coefficient, then it is not uniformly distributed inside mod-N arithmetic progressions of size proportional to N. It follows by a pigeonhole argument that there is a genuine arithmetic progression P of size proportional to $\sqrt{N}$ such that the density of $A\cap P$ inside P is significantly larger than the density of A inside $\mathbb{Z}_N$. This allows us to iterate.

It is now natural to wonder whether the "random behaviour" of the set A implies anything about the number of arithmetic progressions it contains of length four. However, it turns out that merely having small Fourier coefficients is not enough. An example to illustrate this is the set $A = \{x \in \mathbb{Z}_N : -N/1000 < x^2 < N/1000\}$ (where, for the purposes of the inequality, x^2 stands for the representative of x^2 that lies between $-N/2$ and $N/2$). It can be shown, using estimates due to Weyl [We] for exponential sums involving quadratic functions, that all the non-zero Fourier coefficients of this set are very small. We now give a very rough argument (which can easily be made rigorous) to show that A contains more mod-N arithmetic progressions of length four than one would expect. (If A has size αN and is chosen randomly, then one expects about $\alpha^4 N^2$ quadruples of the form $(a, a+d, a+2d, a+3d)$ to belong to A^4.) Suppose we know that $a-d$, a and $a+d$ all belong to A. Then $(a-d)^2$, a^2 and $(a+d)^2$ are all "small" mod N. Taking differences, this implies that $2ad - d^2$ and $2ad + d^2$ are both small, which implies that $4ad$ and $2d^2$ are both small. But then $(a+2d)^2 = a^2 + 4ad + 2.2d^2$ must be small. In other words, once we have an arithmetic progression of length three in A (and of these we have about the expected number) there is a greater chance than there should be that the next term in the progression also belongs to A. Therefore A contains more progressions of length four than it should.

With a bit more effort, one can use similar ideas to construct a set A with small Fourier coefficients and *fewer* arithmetic progressions of length four than a random set of the same cardinality. This seems to indicate that, beautiful as Roth's argument is, there is a fundamental limitation to Fourier methods which stops it generalizing. On the other hand, it is difficult to find examples to illustrate this that are fundamentally different from the set A above. That is, they all seem to involve quadratic polynomials and work for basically the same reason. It turns out that this is *necessary*, and can be proved to be necessary using Fourier methods. We now give a very brief outline of the argument.

The first step is to define a stronger notion of randomness, which we call quadratic uniformity. Let us define a set A to be *δ-uniform* if $|\widetilde{A}(r)| \leq \delta N$ for every non-zero r. Write $A + k$ for $\{x \in \mathbb{Z}_N : x - k \in A\}$. Define A to be *δ-quadratically uniform* if $A \cap (A+k)$ is δ-uniform for all but at most δN values of k. In loose terms, A is quadratically uniform if there are almost no translates of A (meaning sets of the form $A + k$) for which the intersection $A \cap (A+k)$ has a non-trivial large Fourier coefficient.

It can be shown that if A has size αN and is δ-quadratically uniform for sufficiently small δ (depending on α only) then A contains approximately the correct number of arithmetic progressions of length four, and in particular at least one such progression. (The proof is similar to the weak mixing case in Furstenberg's argument. I am grateful to Gil Kalai for pointing this out to me.) We therefore have an appropriate generalization of the first step of Roth's argument, and in fact it can be generalized further, without much difficulty, to deal with arithmetic progressions of arbitrary length.

However, it is not at all obvious what to do if A is *not* quadratically uniform. From the definition we can say that there is a set $B \subset \mathbb{Z}_N$ of size at least δN and a function $\phi : B \to \mathbb{Z}_N$ never taking the value zero such that $\left|A \cap (A+k)^{\sim}(\phi(k))\right| \geq \delta N$ for every $k \in B$, but this fact on its own does not seem particularly helpful. In order to get any further, it is useful to examine the set $A = \{x \in \mathbb{Z}_N : -N/1000 < x^2 < N/1000\}$ mentioned earlier. This is an example of a set which is uniform but not quadratically uniform. A number x belongs to $A\cap(A+k)$ only if both x^2 and $(x-k)^2$ are small, which implies that $2kx-k^2$ is small. It follows (from an easy calculation) that $A\cap(A+k)^{\sim}(2k)$ is large. Thus, the quadratic nature of the set A leads to linear behaviour of the function ϕ.

This suggests that perhaps ϕ cannot be an entirely arbitrary function, and the suggestion is correct. The rest of our proof consists in showing first that ϕ must always have a certain weakish linearity property, and then (reversing the implication from quadratic to linear above) that the linearity of ϕ implies some sort of quadratic bias to the set A. Finally, this quadratic bias implies (using Weyl's estimates for exponential sums mentioned earlier) the existence of an arithmetic progression P of size N^c such that $|A\cap P|\geq(\alpha+\gamma)|P|$ (where c and γ depend on α only).

The most interesting of the steps is finding the linearity of the function ϕ, which is itself done in two stages. The first is a somewhat algebraic argument which shows that, for a constant γ depending on α only, B^4 contains γN^3 quadruples (a,b,c,d) such that $a+b=c+d$ and $\phi(a)+\phi(b)=\phi(c)+\phi(d)$. Let us call such a quadruple *ϕ-additive*. Notice that there are only N^3 quadruples $(a,b,c,d)\in\mathbb{Z}_N^4$ such that $a+b=c+d$, so this is potentially a strong restriction on the function ϕ, and seems to put pressure on ϕ to be linear, or at least to be linear when restricted to some large subset of B.

After a little thought, however, one realizes that there are definitely non-linear examples of functions ϕ for which there are many ϕ-additive quadruples. A typical one is the following. Let m be an integer much larger than 1 and much smaller than N and let $B=\mathbb{Z}_N$. Given $0\leq x<N$, write it as $qm+r$ with $0\leq r<m$, and define $\phi(x)$ to be r. It can be checked easily that there are many ϕ-additive quadruples, and also that there is no large subset of $\mathbb{Z}_N$ on which ϕ is linear.

On the other hand, if one thinks of the numbers 1 and m as being something like a basis of $\mathbb{Z}_N$, then ϕ is something like a linear function defined on a two-dimensional set. It turns out, and this is of enormous importance for the proof, that this sort of quasi-linear behaviour is typical. That is, if there are many ϕ-additive quadruples, then there must be a large subset B' of B such that the restriction of ϕ to B' resembles a linear function defined on a space of not too high a dimension. The proof of this fact is not at all easy, because it relies on a deep theorem of Freiman [F1,2] which we now describe, and in particular a recent proof of Freiman's theorem due to Ruzsa [Ru].

Let X be a subset of $\mathbb{Z}$ of size n. The *sumset* of X, written $X+X$, is simply $\{x+y:x,y\in X\}$. Suppose that we know that the sumset of X has cardinality at most Cn (where we think of n as large and C as fixed). What does this tell us about the set X? This question is not unlike the question we have just asked about ϕ, and one can make similar remarks. The most obvious example of a set X with small sumset is an arithmetic progression. The next most obvious is a large subset of an arithmetic progression. However, these do not exhaust all possibilities. For example, if $X=\{a_1r_1+a_2r_2:0\leq a_i<s_i\}$ then it is an easy exercise to show that $|X+X|<4|X|$. Such a set is called, for obvious reasons, a *two-dimensional* arithmetic progression, and it is not hard to guess the definition of a d-dimensional arithmetic progression for arbitrary d. It is another (similar) easy exercise to show that a large subset of a low-dimensional arithmetic progression will have a small sumset. Remarkably, the converse is also true, and this is Freiman's theorem.

THEOREM 4. *Let X be a subset of $\mathbb{Z}$ such that $|X+X| \leq C|X|$. Then X is a subset of a d-dimensional arithmetic progression of size at most $D|X|$, where d and D depend on C only.*

To relate Freiman's theorem to our problem, we consider the graph of the function ϕ, which we shall call Γ. This is a subset of $\mathbb{Z}_N^2$ of size at most N which contains at least γN^3 quadruples (x, y, z, w) such that $x+y=z+w$. A theorem of Balog and Szemerédi [BS] now tells us that Γ contains a subset X of size at least ηN such that $|X+X| \leq C|X|$, with η and C constants that depend on γ (and hence α) only. It is an easy exercise to formulate an appropriate version of Freiman's theorem for subsets of $\mathbb{Z}^2$ (as we may regard X) and prove that it is equivalent to Freiman's theorem in $\mathbb{Z}$. Applying such a version of Freiman's theorem to X, we find that X is a subset of a d-dimensional arithmetic progression P of size at most $D|X|$. An easy averaging argument shows that P must contain a one-dimensional arithmetic progression Q of size proportional to $N^{1/d}$ such that $|X \cap Q| \geq D^{-1}|Q|$. Now X is the graph of the restriction of ϕ to some subset B' of B, and Q is the restriction of a linear function ψ to an arithmetic progression $R \subset \mathbb{Z}$ (of size proportional to $N^{1/d}$). The estimate for $|X \cap Q|$ tells us that $\phi(x) = \psi(x)$ for at least $D^{-1}|R|$ values of $x \in R$. We have shown that ϕ has at least some linear behaviour, and it turns out to be enough.

We shall now be even more brief. (The reader wishing for more details of the proof should consult [G2].) The linear behaviour of ϕ implies the existence of an arithmetic progression S of size proportional to $N^{1/d}$ and a quadratic function q such that, writing $f(s)$ for $A(s) - \alpha$, we have the inequality

$$\sum_{s \in S} f(s) + \Big|\sum_{s \in S} f(s)\omega^{q(s)}\Big| \geq \zeta |S|$$

with ζ depending on α only. It can be shown, using Weyl's estimates again, that S can be partitioned into arithmetic progressions $T_1, \ldots, T_m$ with $m \leq N^{1-\epsilon}$ such that the restriction of $\omega^{q(s)}$ to any T_j is approximately constant. (For Roth's theorem we needed the corresponding result for linear functions, which is much easier.) When this is done, we have that

$$\Big|\sum_{s \in S} f(s)\omega^{q(s)}\Big| \leq \sum_{j=1}^{m} \Big|\sum_{s \in T_j} f(s)\omega^{q(s)}\Big| \approx \sum_{j=1}^{m} \Big|\sum_{s \in T_j} f(s)\Big| .$$

An averaging argument then yields some j such that $|T_j| \geq N^{\epsilon}$ and $\sum_{s \in T_j} f(s) \geq \zeta'|T_j|$. The second condition is equivalent to the statement that $|A \cap T_j| \geq (\alpha + \zeta')|T_j|$. Finally, we can iterate, just as in the proof of Roth's theorem.

A small modification of the above argument (which uses Ruzsa's proof of Freiman's theorem rather than quoting the theorem directly) leads to an upper bound of $\exp\exp(\delta^{-C})$ for $N(4, \delta)$. Equivalently, if $A \subset \{1, 2, \ldots, N\}$ has cardinality at least $N(\log\log N)^{-c}$, then it must contain an arithmetic progression of length four. Here, C and c are absolute constants. Let us state this result formally.

Theorem 5. *Let $\delta > 0$ and let N be a natural number greater than or equal to* $\exp\exp(\delta^{-C})$*, where C is an absolute constant. Then every subset of the set $\{1, 2, \ldots, N\}$ of size at least δN contains an arithmetic progression of length four.*

Corollary 6. *Let r and N be natural numbers such that $N \geq \exp\exp(r^C)$, where C is an absolute constant. Then, however the set $\{1, 2, \ldots, N\}$ is coloured with r colours, there is a monochromatic arithmetic progression of length four.*

In terms of our previous notation, Corollary 6 states that $M(4, r) \leq \exp\exp(r^C)$. The bound given by Shelah's argument is more like $A_4(A_4(r))$, or in other words a tower of twos of height a tower of twos of height r. The previous best known bound for Theorem 5 was even larger, since the full strength of van der Waerden's theorem was used by Szemerédi even in this special case [Sz1]. So, as we remarked earlier, the bound was probably something like $A_6(\delta^{-1})$.

§4. Further results and questions.

The first question to deal with is whether the above argument generalizes to progressions of arbitrary length. The answer is that most of it does with no difficulty at all. However, one part involves significant extra difficulty. Let us define a set A to be *δ-cubically uniform* if the intersection

$$A \cap (A + k) \cap (A + l) \cap (A + k + l)$$

is δ-uniform for all but at most δN^2 pairs (k, l). Then if A is *not* δ-cubically uniform, one obtains a set $B \subset \mathbb{Z}_N^2$ of cardinality at least δN^2 and a function $\phi : B \to \mathbb{Z}_N$, such that, for every $(k, l) \in B$, the Fourier coefficient of the above intersection at $\phi(k, l)$ has size at least (k, l). The arguments for progressions of length four tell us a great deal about the behaviour of ϕ in each variable separately, but to prove results for longer progressions one must relate these restrictions in order to show that ϕ has some sort of *bilinear* property, and this is not easy to do. At the time of writing, I have a long preprint which deals with the general case and which is still being checked thoroughly. If it stands up to scrutiny, it will give an upper bound for $N(k, \delta)$ of $\exp\exp\big(\delta^{-\exp\exp(k+10)}\big)$.

This estimate is still far from best possible. In fact, for fixed k, the best known lower bound for $N(k, \delta)$ is $\exp\big(c(\log(1/\delta))^2\big)$ [Be]. (This bound may seem unimpressive, but it demonstrates the interesting fact that randomly chosen sets are not the worst, and thereby partly explains the difficulty of Szemerédi's theorem.) The main obstacle to further progress on bounds is that progressions of length three are not fully understood. There is now a development of Roth's argument due to Heath-Brown [H-B] and Szemerédi [Sz3], which gives an upper bound for $N(3, \delta)$ of $\exp(\delta^{-C})$, but this still greatly exceeds the lower bound just mentioned. In particular, the value of C that comes from the argument exceeds 1, which means that it does not prove the first non-trivial case of Conjecture 3. Finding the correct asymptotic behaviour of $N(3, \delta)$ is a fascinating problem, not just for its own sake, but because any methods used to solve it are almost certain to have important further applications.

As mentioned in the introduction, several generalizations of Szemerédi's theorem have been proved using ergodic theory and do not (yet) have any other proofs.

Thus, the question of obtaining *any* bounds for them, not just reasonable ones, is open. We mention three such results. The first is the density version, due to Furstenberg, of a theorem of Gallai.

Theorem 7. *Let $X \subset \mathbb{Z}^d$ and let $\delta > 0$. If N is sufficiently large then every set $A \subset \{1, 2, \ldots, N\}^d$ of size at least δN^d has a subset homothetic to X.*

It seems likely that our methods can be used to give a quantitative version of Theorem 7, but so far this has not been done.

The next result is the density version of the Hales-Jewett theorem, which itself is one of the central results of Ramsey theory. To state it, we need a small amount of notation. Let $Q(k, N)$ be the N-dimensional grid $\{1, 2, \ldots, k\}^N$. (One can think of elements of $Q(k, N)$ as words of length N in the alphabet $\{1, 2, \ldots, k\}$.) Given $x = (x_1, \ldots, x_N) \in Q(k, N)$, $r \in \{1, 2, \ldots, k\}$ and a set $W \subset \{1, 2, \ldots, N\}$, define $x \oplus rW$ to be the sequence obtained from x by replacing x_j by r whenever $j \in W$ and otherwise leaving it unchanged. A *Hales-Jewett line* in $Q(k, N)$ is a set of the form $\{x \oplus rW : 1 \leq r \leq k\}$. The density version of the Hales-Jewett theorem, proved by Furstenberg and Katznelson [FK], is the following result. (The original theorem of Hales and Jewett [HJ] is of course the colouring version.)

Theorem 8. *Let $\delta > 0$ and $k \in \mathbb{N}$. If N is sufficiently large, then every set $A \subset Q(k, N)$ of cardinality at least δk^N contains a Hales-Jewett line.*

One can easily deduce Szemerédi's theorem by projecting $Q(k, N)$ to $\mathbb{Z}$ in a sensible way. Even the case $k = 3$ of the Furstenberg-Katznelson theorem is very hard and was open for a long time. In fact, unlike with Szemerédi's theorem, the case $k = 2$ is not quite obvious either, but it follows easily from a lemma of Sperner [Sp].

Because of the difficulty of the case $k = 3$, there seems to be no immediate prospect of a quantitative version of Theorem 8. If one could somehow find a reasonably simple analytic argument when $k = 3$, then our methods might conceivably suggest a way of extending this to the general case. I would guess, however, that the problem will remain open for a long time.

Finally, we mention a beautiful generalization of Szemerédi's theorem due to Bergelson and Leibman [BL], which solved a problem that had attracted a great deal of interest for several years.

Theorem 9. *Let $\delta > 0$ and let $p_1, \ldots, p_k$ be polynomials with integer coefficients such that $p_i(0) = 0$ for every i. If N is sufficiently large, then for every set $A \subset \{1, 2, \ldots, N\}$ of size at least δN there exist integers a and d (with $d \neq 0$) such that $a + p_i(d) \in A$ for every i.*

Interestingly, the main obstacle for Bergelson and Leibman was obtaining a proof of the colouring version of Theorem 6. They could then use Furstenberg's methods to deduce the density version. Their proof of the colouring version also used ergodic theory, but it can be done purely combinatorially (see [M] or [W]).

The most elementary case of Theorem 9 that does not follow from Szemerédi's theorem is when $k = 2$, $p_1(x) = 0$ and $p_2(x) = x^2$. Then the result states that A contains a pair of the form $(a, a + d^2)$. This result was first proved by Furstenberg [Fu] and Sárközy [S]. Sárközy's argument used exponential sums and gave a sensible bound, which has subsequently been improved by Pintz, Steiger and Szemerédi [PSS] so that it is now known that a density of $C(\log N)^{-c(N)}$ will suffice, where $c(N) = \log\log\log\log N/12$. (It is still not known whether one can get away with a density of $N^{-\epsilon}$ for some $\epsilon > 0$.) It is quite possible, therefore, that some sort of mixture of our methods and other existing methods would give a quantitative version of Theorem 9. This would undoubtedly be a difficult project to carry out, not least because the methods to be mixed are all individually complicated. However, I expect it will be done by somebody in the next ten or fifteen years, if not sooner.

Let me close by saying that in this paper I have concentrated on my recent work because most of the rest is described in the proceedings of the 1994 Congress [G3], and also by Bollobás in this volume.

References

[BS] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica 14 (1994), 263-268.

[Be] F. A. Behrend, *On sets of integers which contain no three in arithmetic progression*, Proc. Nat. Acad. Sci. 23 (1946), 331-332.

[BL] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden and Szemerédi theorems*, J. Amer. Math. Soc. 9 (1996), 725-753.

[ET] P. Erdős and P. Turán, *On some sequences of integers*, J. London Math. Soc. 11 (1936), 261-264.

[F1] G. R. Freiman, Foundations of a Structural Theory of Set Addition, (in Russian), Kazan Gos. Ped. Inst., Kazan (1966).

[F2] G. R. Freiman, Foundations of a Structural Theory of Set Addition, Translations of Mathematical Monographs 37, Amer. Math. Soc., Providence, R. I., USA.

[Fu] H. Furstenberg, *Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. 31 (1977), 204-256.

[FK] H. Furstenberg and Y. Katznelson, *A density version of the Hales-Jewett theorem*, J. Anal. Math. 57 (1991), 64-119.

[G1] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, GAFA, Geom. Funct. Anal. 7 (1997), 322-337.

[G2] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA, Geom. Funct. Anal. 8 (1998), 529-551.

[G3] W. T. Gowers, *Recent results in the theory of infinite-dimensional Banach spaces*, Proc. I.C.M. Zürich 1994, Birkhäuser Verlag, Basel, Switzerland 1995.

[GRS] R. L. Graham, B. L. Rothschild and J. H. Spencer, Ramsey Theory, Wiley Interscience (2nd ed. 1990).

[HJ] A. W. Hales and R. I. Jewett, *Regularity and positional games*, Trans. Amer. Math. Soc. 106 (1963), 222-229.

[H-B] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. (2) 35 (1987), 385-394.

[M] R. McCutcheon, *Elemental methods in ergodic Ramsey theory* (to appear).

[PH] J. Paris and L. Harrington, *A mathematical incompleteness in Peano arithmetic*, in Handbook of Mathmatical Logic (J. Barwise ed.), North-Holland (1977), 1133-1142.

[PSS] J. Pintz, W. L. Steiger and E. Szemerédi, *On sets of natural numbers whose difference set contains no squares*, J. London Math. Soc. (2) 37 (1988), 219-231.

[R1] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. 28 (1953), 245-252.

[R2] K. F. Roth, *Irregularities of sequences relative to arithmetic progressions, IV*, Period. math. Hungar. 2 (1972), 301-326.

[Ru] I. Ruzsa, *Generalized arithmetic progressions and sumsets*, Acta Math. Hungar. 65 (1994), 379-388.

[S] A. Sárközy, *On difference sets of integers I*, Acta Math. Acad. Sci. Hungar. 31 (1978), 125-149.

[Sh] S. Shelah, *Primitive recursive bounds for van der Waerden numbers*, J. Amer. Math. Soc. 1 (1988), 683-697.

[Sp] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z. 27 (1928), 544-548.

[Sz1] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. 20 (1969), 89-104.

[Sz2] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 299-345.

[Sz3] E. Szemerédi, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. 56 (1990), 155-158.

[vdW] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wisk. 15 (1927), 212-216.

[W] M. J. Walters, *Combinatorial proofs of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem*, J. London Math. Soc., (to appear).

[We] H. Weyl, *Über die Gleichverteilung von Zahlen mod Eins*, Math. Annalen 77 (1913), 313-352.

W. T. Gowers
University of Cambridge
Department of Pure Mathematics
and Mathematical Statistics
16 Mill Lane
Cambridge CB2 1SB
England
wtg10@dpmms.cam.ac.uk

Editor's Remark:

Due to a failure of the printing device, Figure 1 in the article of Curtis T. McMullen on page 841 of Volume II of these Proceedings is slightly scrambled. We therefore reproduce it here in correct form:

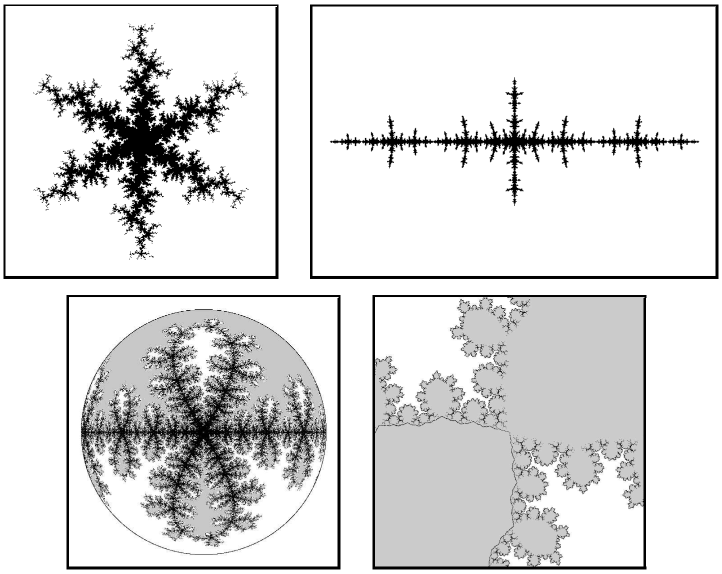

Figure 1. Dynamical systems with deep points: a totally degenerate Kleinian group, the Feigenbaum polynomial, a critical circle map and the golden mean Siegel disk.