The Fourier-Analytic Approach to Szemerédi's Theorem

Esa V. Vesalainen

October 2009
University of Helsinki
Department of Mathematics and Statistics
Master's thesis
Advisor: Mikko Salo
Censors: Tuomas Hytönen and Mikko Salo

Contents

Introduction

In their 1936 paper [Er&T] P. Erdős and P. Turán defined for arbitrary positive integers k and N the number $r_k(N)$ as the maximum possible size of a set formed from N consecutive integers and not containing arithmetic progressions of length k, that is, not containing elements

$$x, x+d, x+2d, \ldots, x+(k-1)d$$

for some $x \in \mathbb{Z}$ and $d \in \mathbb{Z}_+$. The study of these **Erdős–Turán constants** has been an important subject in combinatorics. Erdős and Turán conjectured that $r_3(N) = o(N)$, as $N \longrightarrow \infty$, a result first proved by K. F. Roth in 1952. Roth's method of proof was Fourier-analytic and in fact gave the quantitative upper bound

$$r_3(N) \ll \frac{N}{\log\log N}. \qquad (N \longrightarrow \infty)$$

This eventually lead to the question whether also $r_k(N) = o_k(N)$, as $N \longrightarrow \infty$, for every $k \in \mathbb{Z}_+$? This turned out to be an extremely difficult question, first resolved to the positive by E. Szemerédi in 1975 [Sz2]. This result has turned out to be a quite fruitful subject of research. Besides Szemerédi's combinatorial proof, there has been other proofs, each giving rise to interesting new methods or even entire theories, such as the ergodic Ramsey theory (see p. 12 onwards). In his 1999 and 2002 articles [Go2] and [Go3] W. T. Gowers generalized Roth's Fourier-analytic approach to prove the full Szemerédi theorem, and he obtains the quantitative upper bounds

$$r_k(N) \ll \frac{N}{(\log\log N)^{c_k}}, \qquad (N \longrightarrow \infty)$$

where $k \in \mathbb{Z}_+$ is arbitrary and the numbers c_k are some absolute positive real constants.

Besides being a quite natural approach, the main advantage of the Fourier-analytic approach over the other methods is that it is the only approach that allows one to obtain reasonable upper bounds for the Erdős–Turán constants. Roth's upper bound for r_3 has been improved several times since the 1950s, each time through improvements in the Fourier-analytic approach. Gowers' upper bounds for r_k have been recently improved by B. Green and T. Tao [Gr&T5], and again the approach is a Fourier-analytic one. Our main subject will be Gower's proof of Szemerédi's theorem for four-term progressions.

Finally, we say a few words about the plan of this thesis. We begin with a chapter on historical background and some motivation. Then we turn to the theorem of Roth, which is first discussed with a historical tone by presenting

Roth's original proof of Roth's theorem. This is followed by an introduction of the important concept of discrete Fourier transform and a Fourier-analytic density increment proof of Roth's theorem which will allow us to obtain Roth's quantitative upper bound.

After discussing Roth's theorem, we dedicate a chapter for the **Gowers' uniformity norms**, a vital ingredient in Gowers' proof of Szemerédi's theorem. The last chapter presents Gowers' proof of Szemerédi's theorem in the case of four-term progressions. In other words, we will prove in the last chapter that

$$r_4(N) \ll \frac{N}{(\log\log N)^c}, \qquad (N \longrightarrow \infty)$$

for some absolute positive real constant c. We point out that our presentation of Gowers' proof is not fully self-contained as it applies some well-known results of additive combinatorics and of number theory which we do not prove.

Notation

Before proceeding, we fix some notation used throughout the text. For any positive integer N, the symbol $\mathscr{J}_N$ denotes the set $\{1, 2, \ldots, N\}$. For any $x \in \mathbb{R}$, the symbol $e(x)$ is the usual shorthand for $e^{2\pi i x}$ and $\|x\|$ denotes the distance of x from the set of integers in the real line. Also, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the standard floor and ceiling functions, and occasionally we use the notation $\{x\}$ for the fractional part of x — though we will always explicitly say so. The standard symbols log and lb are used to denote natural logarithms and logarithms in base two, respectively. The symbol $(n_1, n_2, \ldots)$ is used to denote the greatest common factor of the integers n_1, n_2, ... For any set X, we define the symbol

$$\#X \in \{0, 1, 2, 3, \ldots, \infty\}$$

to denote the number of elements in X.

For asymptotic relations we use the standard symbols O, o, Ω, $\ll$, $\gg$, Θ, etc. Let f and g be functions with equal domains of definition. Then $O(g)$ and $\Omega(g)$ denote any complex valued quantities that are bounded in modulus from above and, respectively, below by the expression $C|g|$ in the domain of definition of g for some absolute constant $C \in \mathbb{R}_+$. Such a quantity is $\Theta(g)$ if it is both $O(g)$ and $\Omega(g)$. If the constant C is allowed to depend on some parameters, those parameters are given in subindices to O, Θ and Ω.

Similarly, we write $f \ll g$ to signify that f is $O(g)$, $f \gg g$ to signify that f is $\Omega(g)$, and $f \asymp g$ to signify that f is $\Theta(g)$. If the inequalities related to O, Θ, Ω, $\ll$, $\gg$ or $\asymp$ are to hold only for the points x in the neighbourhood of some point x_0 for which f and g are defined, we append the expression $(x \longrightarrow x_0)$ to the formulas.

If g does not have zeros in some neighbourhood of x_0, where x_0 is again some element in the domain of definition of g, we write

$$f(x) = o\big(g(x)\big) \qquad (x \longrightarrow x_0)$$

to signify that

$$\frac{f(x)}{g(x)} \longrightarrow 0,$$

when x tends to x_0 in some neighbourhood of x_0 contained in the domain of definition shared by f and g.

If A is finite non-empty set and $f\colon A \longrightarrow \mathbb{C}$ it is useful to define the **expectation** of f as the normalized sum

$$\mathop{\mathscr{E}}_{x \in A} f(x) \stackrel{\text{def}}{=} \frac{1}{\#A} \sum_{x \in A} f(x).$$

We often suppress the ranges of the indices in sums and expectations if they are obvious from the context. If B is a subset of A, we say that the **density** of B (in A) is $\frac{\#B}{\#A}$. Finally, if C is any set, χ_C denotes its **characteristic function** defined by the formula

$$\chi_C(x) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } x \in C, \\ 0, & \text{if } x \notin C. \end{cases}$$

The exact domain of defition of χ_C is rarely important. However, in the case where C is a subset of, say, $\mathbb{Z}_N$ for some $N \in \mathbb{Z}_+$, we often implicitly think of χ_C as a function defined on $\mathbb{Z}_N$ in order to avoid ambiguities when using the above expectation notation.

A Brief Overview of the History of Szemerédi's Theorem

We begin by stating, motivating and describing the history of the problem of estimating the Erdős–Turán constants. In this connection, the theorems of Roth and Szemerédi are in a key position. Most of the historical information in this chapter has been assembled from the numerous books and research, survey and expository articles related to additive combinatorics in general and to the problem in particular. We mention some of these right away.

The article [Gr&T1] of B. Green and T. Tao is a non-technical overview of Szemerédi's theorem. The survey [Sh] of I. D. Shkredov also discusses the many generalizations and technical ideas related to Szemerédi's theorem and contains a wealth of references. I. Łabas' article [Ł] discusses the connections between harmonic analysis and additive combinatorics. The part VII of A. Soifer's book [Soi] contains a historical discussion of ancestors of Szemerédi's theorem. Finally, we mention that the article [Cr&L] has a section on the history of additive combinatorics.

Background

Among other things, number theory deals with different kinds of patterns in various sets of integers. Indeed, the literature is full of theorems and questions on such matters. The precursors of the theorems and questions that we focus on include the following classics, discussed for instance in [Soi] and [Sol].

Hilbert's theorem. [Hi] *If the set $\mathbb{Z}_+$ of positive integers is partitioned into finitely many disjoint sets, then one of them contains arbitrarily large* ***affine cubes****, i.e. sets of the form*

$$\left\{ x_0 + \sum_{x \in A} x \;\middle|\; A \subseteq \{x_1, x_2, \ldots, x_n\} \right\},$$

where $n \in \mathbb{Z}_+$ and $x_0, x_1, x_2, \ldots, x_n \in \mathbb{Z}_+$.

Schur's theorem. [Sc] *For any $k \in \mathbb{Z}_+$ there exists a number $N \in \mathbb{Z}_+$ such that when $\mathscr{J}_N$ is partitioned into k subsets, one of them contains a subset of the form $\{x, y, x+y\}$ with $x, y \in \mathbb{Z}_+$.*

An ancient question. *Given a number $\ell \in \mathbb{Z}_+$, are there numbers $a \in \mathbb{Z}_+$ and $d \in \mathbb{Z}_+$ for which each of the numbers*

$$a, a+d, a+2d, \ldots, a+(\ell-1)\,d$$

is a prime?

The last question appears to have been studied already by such mathematicians as J. L. Lagrange and E. Waring [Gr&T2]. It turns out to be quite relevant to our subject.

A particularly important type of patterns or configurations is that of **arithmetic progressions**, i.e. sets of the form

$$\big\{a, a+d, a+2d, \ldots, a+(\ell-1)d\big\}$$

with $a, d \in \mathbb{Z}$ and $\ell \in \mathbb{Z}_+$. In general, for any additively written Abelian group G, the term **G-arithmetic progression** means a subset of G of the above form with $a, d \in G$ and $\ell \in \mathbb{Z}_+$. In each case the number ℓ is called the **length** of the arithmetic progression and d is called the **common difference** between the terms of the progression. If $d \neq 0$, the corresponding arithmetic progression is said to be **proper** or **non-trivial**. We often call arithmetic progressions of length three **arithmetic triples** and say that a set without proper arithmetic triples is **progression-free**.

In 1927 B. L. van der Waerden published a proof [vdW1] of a conjecture which he attributed to Baudet. In the light of the detailed discussion on this subject in the book [Soi], it is perhaps appropriate to attribute the conjecture both to P. Baudet and to I. Schur. The claim was that when $\mathbb{Z}_+$ is partitioned into two disjoint subsets, one of them necessarily contains arbitrarily long proper arithmetic progressions. In fact, van der Waerden proved a slightly stronger but ultimately somewhat equivalent version of the conjecture — a one more amenable to the intricate induction arguments involved [vdW2]. The actual statement of the theorem reads as follows:

van der Waerden's theorem. *For any $k \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_+$, there exists a number $N \in \mathbb{Z}_+$ which has the property that whenever a set of N consecutive integers is partitioned into k disjoint subsets, one of them contains a non-trivial arithmetic progression of length ℓ.*

The smallest such value of N is denoted by $W(k, \ell)$. The numbers $W(k, \ell)$ are called **van der Waerden numbers**. Obtaining good bounds for these numbers is extremely difficult. The induction approach to van der Waerden's theorem can be adapted to give explicit bounds. However, due to the highly recursive nature of the arguments involved the bounds thereby obtained are phenomenally weak.

As an example we mention the bounds given by the proof of M. A. Lukomskaya which is presented in the classic book [Kh]. In that proof one first defines recursively the sequences $q_0, q_1, \ldots$ and $n_0, n_1, \ldots$ by the formulas

$$\begin{cases} q_0 = 1, & n_0 = n(k, \ell)\,, \quad \text{and} \\ q_s = 2n_{s-1}q_{s-1}, & n_s = n(k^{q_s}, \ell)\,, \quad \text{which are to hold for } s \in \mathbb{Z}_+. \end{cases}$$

Here ℓ is a positive integer and the numbers $n(1,\ell)$, $n(2,\ell)$, ... are such that

$$W(1,\ell) \leqslant n(1,\ell)\,, \quad W(2,\ell) \leqslant n(2,\ell)\,, \quad \text{etc.}$$

The existence of such numbers is of course given by the induction assumption. The number $k \in \mathbb{Z}_+$ is arbitrary and simply designates the numbers of classes into which partition is allowed to be made. The proof then goes on to show that one has the inequality

$$W(k,\ell+1) \leqslant q_k.$$

Now it is easy to obtain an explicit numerical upper bound for, say, $W(3,3)$. As for any $k \in \mathbb{Z}_+$ the true value of $W(k,2)$ is trivially $k+1$, we may take those values as our numbers $n(k,\ell)$ and the sequences $q_0, q_1, \ldots$ and $n_0, n_1, \ldots$ corresponding to the van der Waerden number $W(3,3)$ can be easily computed:

$$\begin{cases} q_0 = 1, & n_0 = n(3,2) = 4; \\ q_1 = 2n_0q_0 = 8, & n_1 = n\big(3^8,2\big) = 3^8+1; \\ q_2 = 2n_1q_1 = 16\left(3^8+1\right), & n_2 = n\Big(3^{16\left(3^8+1\right)},2\Big) = 3^{16\left(3^8+1\right)}+1; \\ q_3 = 2n_2q_2 = 32\left(3^8+1\right)\Big(3^{16\left(3^8+1\right)}+1\Big), & \ldots; \\ \ldots \end{cases}$$

Thus the upper bound obtained for $W(3,3)$ is

$$W(3,3) \leqslant 32\left(3^8+1\right)\Big(3^{16\left(3^8+1\right)}+1\Big) = 209984\left(3^{104992}+1\right).$$

However, as revealed by the tables of [L&al.], the exact value of $W(3,3)$ is 27.

The paper of Erdős and Turán

In 1936 appeared the foundational three-page paper [Er&T] by P. Erdős and P. Turán. As it is the first paper published in the subject and it is referred to everywhere in the literature, we briefly summarize its content in this section. Erdős and Turán begin by defining progression-free sets of integers and by defining the quantity $r_3(N)$ as the maximum possible size of a progression-free subset of $\mathscr{J}_N$, for any $N \in \mathbb{Z}_+$. Near the end of the paper they also explicitly define the quantities

$$r_k(N) \stackrel{\text{def}}{=} \max\big\{\#A\big|A \subseteq \mathscr{J}_N \text{ does not contain a proper arithmetic progression of length } k\big\},$$

where $N \geqslant 1$ and $k \geqslant 3$ are to be integers. These quantities are nowadays called **Erdős–Turán constants**.

There are only two theorems given in the paper. The first is the inequality

$$r_3(2N) \leqslant N,$$

proven by elementary means for integers $N \geqslant 8$. The second is the slight strengthening

$$r_3(N) < \left(\frac{4}{9}+\varepsilon\right)N,$$

which is proven for arbitrary $\varepsilon \in \mathbb{R}_+$ for sufficiently large positive integers N. Again the proof is elementary. It is mentioned that the constant $\frac{4}{9}$ can be improved to $\frac{3}{8}$ with a longer but similar argument. This leads Erdős and Turán to state the important conjecture:

"It is probable that $r_3(N) = o(N)$*."*

The rest of the paper presents two conjectures, both attributed to G. Szekeres. The first is the equality

$$r_3\left(\frac{3^k+1}{2}\right) = 2^k,$$

which is remarked to be known to hold for $k \in \{1,2,3,4\}$. The second conjecture is a generalization of the previous one:

$$r_p\left(\frac{(p-2)\,p^k+1}{p-1}\right) = (p-1)^k\,,$$

which is to hold for odd prime numbers p. As is stated in [Er&T] and proven in the next section, it is easy to see that the expressions on the right-hand side give lower bounds for the Erdős–Turán constants.

Erdős and Turán motivate the second conjecture by two impressive corollaries. The first one is that the set of prime numbers would contain arbitrarily long proper arithmetic progressions and the second one is that one would get upper bounds for the van der Waerden numbers. The precise inequality given in the paper is

$$W(k,\ell) < \ell^{c\ell \log k},$$

where k and ℓ are arbitrary positive integers and c is a positive real constant.

A remark on notation and terminology

Regarding terminology, the term "progression-free set" has superceded the older terms "$\mathscr{A}$-sequence" [Er&T] and "$\mathscr{A}$-set" [Ro1]. In the notation department, it should be noted that Erdős and Turán's original notation for r_3 was r. Instead of r_3, Roth considers the quantity $\frac{r_3(N)}{N}$ which he denotes by $A(N)$ in [Ro1] and by $a(N)$ in [Ro2]. The last of these symbols is still preferred by some. For instance, in [Sh] I. D. Shkredov considers the quantities $a_k(N) = \frac{r_k(N)}{N}$ instead of the more standard $r_k(N)$.

Lower bounds for the Erdős–Turán constants

In the following sections and chapters, the main role will be played by upper bounds for the Erdős–Turán constants. However, the first concrete results on Erdős–Turán constants obtained after the paper [Er&T] were lower bound results for r_3. The lower bound results are of course very different in spirit from the upper bound results as they always involve some rather concrete constructions. In this section we describe the famous construction of F. A. Behrend, and give some further references.

The first lower bounds for the Erdős–Turán constants were given already in the paper [Er&T], as was mentioned in the previous section. In that article it was noted that

$$r_3\Big(\frac{3^n+1}{2}\Big) \geqslant 2^n, \qquad (\alpha)$$

for all $n \in \mathbb{Z}_+$, and that more generally, for any odd prime p,

$$r_p\Big(\frac{(p-2)p^n+1}{p-1}\Big) \geqslant (p-1)^n, \qquad (\beta)$$

again for all $n \in \mathbb{Z}_+$.

The first inequality is rather easy to prove as it follows for example from the observation that the set of positive integers containing no digits other than zeroes and ones in their ternary representations furnish an explicit construction with qualities implying (α). As for (β), the set of positive integers not containing digits other than 0, 1, ..., $p-3$ and $p-2$ in their base p representation will do. Indeed, for any increasing arithmetic progression of length p consisting of such numbers, a straightforward induction argument shows that each digit of the common difference of the terms must vanish.

As remarked before, Szekeres conjectured that equalities occur in (α) and (β). This was disproven in 1942 by R. Salem and D. C. Spencer. In their article [Sal&S1] they showed that for every $\varepsilon \in \mathbb{R}_+$, one has

$$r_3(N) > N^{1-\frac{\log 2+\varepsilon}{\log\log N}}, \qquad (N \longrightarrow \infty)$$

by considering for positive integers d and n, satisfying $d > 2$ and $d \mid n$, the sets $S(n,d)$ of integers of the form

$$\sum_{\ell=0}^{n-1} a_\ell\,(2d-1)^\ell\,,$$

where $a_0, a_1, \ldots, a_{n-1} \in \{0\} \cup \mathscr{J}_{d-1}$, and for any $i \in \{0\} \cup \mathscr{J}_{d-1}$ exactly $\frac{n}{d}$ of the numbers a_0, a_1, ..., a_{n-1} are equal to i.

Behrend's construction

In 1946 F. A. Behrend refined the idea of Salem and Spencer and gave a simpler, yet stronger, construction yielding the lower bound

$$r_3(N) > N^{1-\frac{2\sqrt{2\log 2}+\varepsilon}{\sqrt{\log N}}}, \qquad (N \longrightarrow \infty)$$

for any fixed $\varepsilon \in \mathbb{R}_+$. We prove this following the original article [Beh2]. The fundamental idea behind Behrend's construction is that a line can intersect a sphere in at most two points and thus spheres do not contain any non-trivial arithmetic triples. Instead of spheres Behrend considers rather sphere-like subsets of $\mathbb{Z}_+$. Namely, he considers the sets

$$S_k(n,d) \stackrel{\text{def}}{=} \left\{ \sum_{\ell=0}^{n-1} a_\ell\,(2d-1)^\ell \;\middle|\; a_0, a_1, \ldots, a_{n-1} \in \{0\} \cup \mathscr{J}_{d-1}, \sum_{\ell=0}^{n-1} a_\ell^2 = k \right\},$$

defined for positive integers k, n and d satisfying the inequalities $d \geqslant 2$, $n \geqslant 2$, and $k \leqslant n\,(d-1)^2$.

Any such set $S_k(n,d)$ is easily seen to be free of any non-trivial arithmetic progressions of length three. To prove this, suppose that

$$a = \sum_{\ell=0}^{n-1} a_\ell\,(2d-1)^\ell\,, \quad b = \sum_{\ell=0}^{n-1} b_\ell\,(2d-1)^\ell\,, \quad \text{and} \quad c = \sum_{\ell=0}^{n-1} c_\ell\,(2d-1)^\ell$$

are three elements of $S_k(n,d)$ satisfying $a+b=2c$, where the digits a_ℓ, b_ℓ and c_ℓ are as in the definition of $S_k(n,d)$. Then, by considering a, b and c in base $2d-1$, we immediately see that we have $a_\ell + b_\ell = 2c_\ell$ for each $\ell \in \{0\} \cup \mathscr{J}_{n-1}$. Consequently,

$$\sum_{\ell=0}^{n-1} a_\ell^2 + \sum_{\ell=0}^{n-1} b_\ell^2 = 2k = 2\sum_{\ell=0}^{n-1} c_\ell^2 = 2\sum_{\ell=0}^{n-1} \left(\frac{a_\ell+b_\ell}{2}\right)^2 = \frac{1}{2}\sum_{\ell=0}^{n-1} (a_\ell+b_\ell)^2\,,$$

so that $\sum_{\ell=0}^{n-1}(a_\ell - b_\ell)^2 = 0$, and $a=b=c$.

Behrend's lower bound now follows from a straightforward application of the pigeonhole principle and judicious choices of the parameters d and n. Clearly, for any integers $n \geqslant 2$ and $d \geqslant 2$, we have

$$\begin{aligned}\#\bigcup_{k=1}^{n(d-1)^2} S_k(n,d) = \#\Big(\big\{\,\langle a_0, a_1, \ldots, a_{n-1}\rangle \,\big|\, a_0, a_1, \ldots, a_{n-1} \in \{0\} \cup \mathscr{J}_{d-1}\big\} \\ \setminus \big\{\langle 0,0,\ldots,0\rangle\big\}\Big) = d^n - 1,\end{aligned}$$

so that by the pigeonhole principle,

$$\#S_k(n,d) \geqslant \frac{d^n-1}{n(d-1)^2}$$

for some $k \in \mathscr{J}_{n(d-1)^2}$. As $S_k(n,d) \subseteq \mathscr{J}_{(2d-1)^n}$, this implies the inequalities

$$r_3\big((2d-1)^n\big) \geqslant \frac{d^n-1}{n\,(d-1)^2} > \frac{d^{n-2}}{n},$$

where the latter inequality holds since some expansion and simplifying reduce it to $2d^{n-1} > 1 + d^{n-2}$.

We finish the proof by choosing n and d appropriately for a given $N \in \mathbb{Z}_+$. Let $n = \big\lfloor \sqrt{2\,\mathrm{lb}\,N} \big\rfloor$ and let $d \in \mathbb{Z}_+$ be such that

$$(2d-1)^n \leqslant N < (2d+1)^n\,.$$

When N is sufficiently large, these choices are feasible. Then also $d > \frac{\sqrt[n]{N}-1}{2}$. We need the following simple observations: Since

$$\log \frac{n-2}{\sqrt[n]{N}} = \log(n-2) - \frac{1}{n}\log N \xrightarrow[N\longrightarrow\infty]{} -\infty,$$

we have $\frac{n-2}{\sqrt[n]{N}} \longrightarrow 0$ as $N \longrightarrow \infty$. Using the simplest asymptotic expansion for the logarithm, we see that

$$\log\Big(1 - \frac{1}{\sqrt[n]{N}}\Big)^{n-2} = (n-2)\log\Big(1-\frac{1}{\sqrt[n]{N}}\Big) = (n-2)\left(-\frac{1}{\sqrt[n]{N}} + O\Big(\frac{1}{\sqrt[n]{N^2}}\Big)\right),$$

as $N \longrightarrow \infty$, and therefore

$$\lim_{N\longrightarrow\infty}\Big(1-\frac{1}{\sqrt[n]{N}}\Big)^{n-2}=1.$$

Now we have for any arbitrarily small positive real number ε that

$$\begin{aligned} r_3(N) &\geqslant r_3\big((2d-1)^n\big) > \frac{d^{n-2}}{n} > \frac{1}{n}\left(\frac{\sqrt[n]{N}-1}{2}\right)^{n-2} \\ &= \frac{N^{\frac{n-2}{n}}}{n2^{n-2}}\left(1-\frac{1}{\sqrt[n]{N}}\right)^{n-2} > \frac{N^{\frac{n-2}{n}}}{n2^{n-1}} = N^{\frac{n-2}{n}-\frac{\log n}{\log N}-\frac{(n-1)\log 2}{\log N}} \\ &\geqslant N^{1-\frac{\sqrt{2\log 2}}{\sqrt{\log N}}-\frac{\varepsilon}{\sqrt{\log N}}-\frac{\sqrt{2\log 2}}{\sqrt{\log N}}} = N^{1-\frac{2\sqrt{2\log 2}+\varepsilon}{\sqrt{\log N}}}, \end{aligned}$$

provided N is sufficiently large. Q.E.D.

Other developments

Behrend published his construction in 1946 but surprisingly it took over sixty years before a quantitative improvement was found. In 2008 appeared the preprint [El] of M. Elkin in which he obtains the lower bound

$$r_3(N) \gg \frac{N\sqrt[4]{\log N}}{4^{\sqrt{2\,\mathrm{lb}\,N}}}. \qquad (N\longrightarrow\infty)$$

Whereas Behrend considered infinitely thin spheres with bounded radii, Elkin considers thicker layers of balls, multidimensional annuli, and proceeds to show that one may choose a large progression-free subset from one of those annuli. A much shorter proof for the same lower bound is given in [Gr&W].

Another possibility to improve Behrend's construction is given by the fact that it is not constructive as its proof invokes the pigeonhole principle and thus does not yield any explicit progression-free set. In [Mo] L. Moser gives an explicit construction which does not require the pigeonhole principle and yet produces the lower bound

$$r_3(N) > N^{1-\frac{c}{\sqrt{\log N}}}, \quad (N\longrightarrow\infty) \qquad (c \text{ a positive real constant})$$

which is essentially Behrend's lower bound. In [El] Elkin shortly discusses the possibility of avoiding the pigeonhole principle in his construction.

We leave this subject with the remark that generalizations of Behrend's construction for longer arithmetic progressions are given in [Ra] and [Ł&L].

The different approaches to Szemerédi's theorem

The conjecture $r_3(N) = o(N)$ given in [Er&T] in 1936 turned out to be quite difficult. It was proved by K. F. Roth in his 1952 article [Ro1] through a clever adaptation of the circle method. In the next year appeared the article [Ro2] in which he obtains the quantitative upper bound

$$r_3(N) \ll \frac{N}{\log\log N}. \qquad (N\longrightarrow\infty)$$

This achievement was mentioned when Roth was awarded the Fields medal in 1958. Even though the circle method had been used in the study of additive questions since 1920s, it was always applied to rather special sets. Roth was the first to apply the method in the case of a **general** set satisfying a simple arithmetic condition.

The combinatorial approach

In the late 1950s the conjecture that maybe $r_k(N) = o(N)$ for arbitrary $k \geqslant 3$ was born. A. Soifer has narrowed the date of birth of this conjecture between the years 1957–1959. This estimate is based on the two observations that Erdős' 1957 open problem paper [Er1] does not contain the conjecture whereas his 1961 problem paper [Er2] (which was submitted in 1960) does include it [Soi, p. 352]. This conjecture turned out to be substantially more difficult than its easiest special case $k = 3$ solved by Roth.

The special case $k = 4$ was settled by E. Szemerédi in the 1969 paper [Sz1]. In 1973 Szemeredi managed to adapt his approach to the case $k = 5$ which he did not publish [Soi, p. 349], and two years later, in 1975, appeared the article [Sz2] in which he settled the conjecture for all values of k by an elementary combinatorial argument of legendary difficulty. This earned Szemerédi a $1000 Erdős prize and the reputation of being a wizard of combinatorics [Soi, pp. 349–350].

Besides the original article [Sz2], the proof is also presented in T. Tao's exposition [T2]. His article [T8] also briefly discusses Szemerédi's proof.

Szemerédi's proof introduced as an auxiliary lemma the Szemerédi regularity lemma. It has become ubiquitous in modern graph theory. For a general reference we refer to the surveys [Ko&S, Ko&al.].

The ergodic theoretic approach

Two years later, in 1977, H. Furstenberg published an article [Fu] in which he proves Szemerédi's theorem by ergodic theoretical means. Furstenberg's proof applied the axiom of choice and did not yield any quantitative upper bounds for the Erdős–Turán constants. On the other hand, the proof introduced to the world a way of proving combinatorial theorems by translating them into the language of ergodic theory. This basically opened up a new field of research — the ergodic Ramsey theory.

Even if the ergodic theoretical approach does not allow good quantitative bounds, it is so flexible that it does allow striking generalizations of many important theorems of additive combinatorics. This includes the so-called density version of the Hales–Jewett theorem, which is a deep generalization of van der Waerden's theorem [Fu&K2], a multi-dimensional version of Szemerédi's theorem [Fu&K1], and a generalization of Szemerédi's theorem, due to V. Bergelson and A. Leibman [Ber&L], which guarantees that a set of integers of positive upper density contains arbitrary polynomial patterns. It is notable that many of these generalizations have turned out to be difficult to be reproduced by other means.

Regarding the literature, we mention that the article [Kr]of B. Kra discusses ergodic methods in additive combinatorics, that the presentations [Ber1, Ber2, Ber3, Ber4] of Bergelson discuss ergodic Ramsey theory in general, and that the

article [T8] of T. Tao discusses both combinatorial and ergodic approaches to Szemerédi's theorem.

Even though Furstenberg's proof of Szemerédi's theorem did not give any quantitative upper bounds, Tao has recently given an ergodic theoretic proof that in principle gives quantitative upper bounds, though rather weak ones. Besides the original article [T6], this proof is also discussed in the Tao's expository note [T3].

The Fourier-analytic approach

Despite their depth and impact, both Szemerédi's and Furstenberg's proofs leave something to be desired regarding the original motivations. One is that by initiating the study of large sets without arithmetic progressions of prescribed length, Erdős and Turán hoped for stronger upper bounds for van der Waerden numbers and may have been hoping that a proof of such bounds would also give a more natural proof of van der Waerden's theorem. The proofs of Szemerédi and Furstenberg fail to give reasonable upper bounds for Erdős–Turán constants and van der Waerden numbers, for Szemerédi's proof applies van der Waerden's theorem and Furstenberg's proof applies the axiom of choice. Another remaining question is whether Roth's approach could be adapted to give a proof for the entire Szemerédi theorem. Roth's original approach, which gave such a good upper bound for r_3, did not generalize to a proof for the entire Szemerédi's theorem.

All this lead W. T. Gowers to seek a generalization of Roth's argument for longer progressions. He made progress in understanding where the difficulties lie and in 1999 he published the article [Go2] in which the case $k = 4$ is settled by Fourier-analytic means and in fact he obtains the quantitative bound $r_4(N) \ll \frac{N}{(\log\log\log N)^c}$, which is to hold for $N \longrightarrow \infty$, and where $c \in \mathbb{R}_+$ is a constant. The remaining cases $k > 4$ provided further complications but three years later, in 2002, the over 120 pages long paper [Go3] appeared, furnishing a proof of the full Szemerédi theorem and giving the upper upper bounds

$$r_k(N) \ll_k \frac{N}{(\log\log N)^{c_k}},$$

as $N \longrightarrow \infty$, for all integers $k \geqslant 3$, where c_3, c_4, ... are positive real constants. The last chapter of this text presents some of the ideas used in the proof.

Gowers' proof yields first reasonable bounds for van der Waerden's constants. The corollary 18.7 of the paper [Go3] says that

$$W(2,k) \leqslant 2^{2^{2^{2^{2^{k+9}}}}}$$

for any positive integer k.

One of the fundamental ideas behind Gowers' proof is that the kind of linear Fourier-analysis, used in Roth's proof of Roth's theorem, simply is not enough to detect longer progressions. The proof deals with higher degree phase factors instead. This has recently inspired work in developing "quadratic Fourier-analysis". Green's lecture notes [Gre7] discuss these matters.

The article [Go3] introduced the **Gowers uniformity norms** that have shown to be very useful. Besides their use in additive combinatorics, they

have also found computer science applications to ZOR lemmas for correlation with low degree $GF(2)$ polynomials and to PCPs and query verifiers. These applications are discussed in chapters 8 and 9 of the lecture notes [Bar&al.].

The hypergraph approach

There is one more approach to Szemerédi's theorem, one based on the 1978 observation of I. Z. Ruzsa and Szemerédi (published in [Ru&Sz]) that Roth's theorem follows quite effortlessly from Szemerédi's regularity lemma for ordinary graphs. This inspired a lot of research into finding and proving appropriate generalizations of the regularity lemma for hypergraphs, so that Szemerédi's theorem would then easily follow from these. This work was finished by B. Nagle, V. Rődl, M. Schacht and J. Skokan in the series of articles [Nag&al., Rö&Sc1, Rö&Sc2, Rö&Sk1, Rö&Sk2], published in the years 2006, 2007, 2007, 2004 and 2006, respectively, and in a different way by Gowers in [Go6] in 2007.

Primes in arithmetic progression

The ancient problem of studying the possible existence of long arithmetic progressions consisting of distinct prime numbers is a particularly striking application of the wide circle of ideas connected with Szemerédi's theorem. Before 2005, the only result on this question was J. G. van der Corput's 1939 theorem that there are infinitely many proper three-term arithmetic progressions in the set of primes (see [vdC]; also [Cho] and [Chuda] are likely to be relevant).

In 2005 B. Green proved the following quantitative version of van der Corput's theorem. Write ϖ for the set of prime numbers. Green's theorem says that when $N \in \mathbb{Z}_+$ is sufficiently large, any subset A of $\varpi \cap \mathscr{J}_N$ of cardinality

$$\#A \gg \frac{N\sqrt{\log\log\log\log\log N}}{\log N \cdot \sqrt{\log\log\log\log N}}$$

contains a proper arithmetic triple. The paper [Gre6] was fittingly titled "Roth's Theorem in the Primes". K. O. Chipeniuk has written an exposition [Chi] on Green's proof.

The true breakthrough came in 2004, when Green and T. Tao announced a proof for the existence of arbitrarily long arithmetic progressions in ϖ [Gr&T2]. The theorem is ergodic theoretical in nature, and relies on Szemerédi's theorem, Gowers norms, and some recent number theoretical results of D. Goldston and C. Y. Yıldırım.

The Green–Tao theorem has been generalized to polynomial patterns [T&Z] and linear equations [Gr&T6] in primes. Tao has recently proved that the set of Gaussian primes contains arbitrarily chosen constellations [T7].

The Green–Tao theorem and the ideas behind its proof are discussed in many different expository articles. We only mention [Gre5], [Ho], [T4] and [T5].

Erdős' conjecture

No discussion of the history of Szemerédi's theorem could be complete without mentioning the outstanding conjecture of Erdős.

Erdős' conjecture. *Let A be a set of positive integers such that $\sum_{x\in A}\frac{1}{x} = \infty$. Then the set A must contain arbitrarily long proper arithmetic progressions.*

It is observed in [Soi, sect. 35.4] that it is difficult to tell when Erdős first formulated this conjecture. Erdős originally offered $3000 for its solutions but shortly before his death raised this to $5000.

The conjecture seems to be very difficult. Not even the simplest special case of concluding the existence of arithmetic triples in such a set has been proved. We remark that in view of L. Euler's observation $\sum_{p\in\varpi}\frac{1}{p}=\infty$, the conjecture is strong enough to imply the existence of arbitrarily long arithmetic progressions in the primes on density grounds only.

Roth's Theorem

In this chapter we will discuss Roth's theorem in detail. We begin by giving some equivalent forms of the result. After briefly introducing Dirichlet's lemma on rational approximation, we present Roth's original proof of his theorem and then give Varnavides' improvement of the conclusion.

Then we turn to the modern point of view. We recall the elements of the discrete Fourier transform on residue class groups $\mathbb{Z}_N$. In the next section we give a density increment proof for Roth's theorem and obtain Roth's upper bound $r_3(N) \ll \frac{N}{\log\log N}$. These two sections are the most important ones regarding the following two chapters.

We end the chapter by mentioning some quantitative improvements obtained for r_3 after Roth.

Different forms of the theorems of Roth and Szemerédi

Important mathematical results often come in many forms and Roth's and Szemerédi's theorems are no exceptions. In this section we concisely introduce and discuss some of the different forms in which these results appear in the literature.

The upper density form

The **upper (asymptotic) (Banach) density** of a set $A \subseteq \mathbb{Z}_+$ is defined to be the quantity

$$d^*(A) \stackrel{\text{def}}{=} \limsup_{N \longrightarrow \infty} \frac{\#(A \cap \mathscr{J}_N)}{N}.$$

The form of Szemerédi's theorem that is perhaps the most relevant one regarding the intuition behind van der Waerden's theorem is the following:

> *Any set $A \subseteq \mathbb{Z}_+$ of positive upper density contains arbitrarily long proper arithmetic progressions.*

The equivalence between this and the usual formulation

> *For each $k \in \mathbb{Z}_+$, we have $r_k(N) = o(N)$ as $N \longrightarrow \infty$.*

is easily seen. In fact, they are equivalent for each fixed length k separately. For if it is known that $r_k(N) = o(N)$ for some $k \in \mathbb{Z}_+$, and if $A \subseteq \mathbb{Z}_+$ is a set of positive upper density δ, then one can find a lower bound $N_0 \in \mathbb{Z}_+$ so that $r_k(N) < \frac{\delta N}{2}$ for any integer $N > N_0$. As A has positive upper density δ, we

must have an integer $N \in \mathbb{Z}_+$ for which both $N > N_0$ and $\#(A \cap \mathscr{J}_N) > \frac{\delta N}{2}$, so that A contains a proper arithmetic progression of length k.

For the other direction, suppose that for some $k \in \mathbb{Z}_+$, we do not have $r_k(N) = o(N)$. Then there must exist a number $\varepsilon \in \mathbb{R}_+$ and an increasing sequence

$$N_1 < N_2 < N_3 < \dots$$

of positive integers and a sequence of sets

$$A_1 \subseteq \mathscr{J}_{N_1}, \quad A_2 \subseteq \mathscr{J}_{N_2}, \quad A_3 \subseteq \mathscr{J}_{N_3}, \quad \dots,$$

for which $\#A_\ell \geqslant \varepsilon N_\ell$ for each $\ell \in \mathbb{Z}_+$, and yet none of these sets contains a proper arithmetic progression of length k. Through restriction into a subsequence of $\langle N_\ell \rangle_{\ell=1}^{\infty}$, if necessary, we may assume that $N_{\ell+1} \geqslant 3N_\ell$ for each $\ell \in \mathbb{Z}_+$. Then the set

$$A = \bigcup_{\ell=1}^{\infty} (2N_\ell + A_\ell) \subseteq \mathbb{Z}_+$$

has a positive upper density at least $\frac{\varepsilon}{3}$ but can not contain any proper arithmetic progressions of length k.

To see this, suppose that x, $x+d$, $x+2d$, …, $x+(k-1)d$ (here $x, d \in \mathbb{Z}_+$) is a proper arithmetic progression of length k contained in A. Not all of the terms may be contained in a single set $2N_\ell + A_\ell$ for some $\ell \in \mathbb{Z}_+$, for the sets A_1, A_2, … were assumed to be free of any proper arithmetic progressions of length k. Let $i \in \mathbb{Z}_+$ and $j \in \mathbb{Z}_+$ be the unique indices for which $x \in 2N_i + A_i$ and $x+(k-1)\,d \in 2N_j + A_j$, respectively. Now $x+d$ can not belong to $2N_j + A_j$ for then d would have to be larger than the distance between $2N_i + A_i$ and $2N_j + A_j$. In particular d would have to be larger than N_j. This would mean that the progression under consideration could only contain one term in $\mathscr{J}_{N_j}$ and only one term in $2N_j + A_j$ and would therefore be of length at most two. Thus, $x + d \in \mathscr{J}_{N_j}$. But this means that $d < N_j$, and the progression must contain an element in $N_j + \mathscr{J}_{N_j}$, which is impossible. We have reached a contradiction.

The limit form

The following simple lemma is proven in [Polla] and [Vau] and is taken for granted in papers such as [Ro1, Ro2].

Lemma. *Suppose that $f\colon \mathbb{Z}_+ \longrightarrow \mathbb{Z}_+$ is increasing and subadditive. That is, suppose that for any $m \in \mathbb{Z}_+$ and $n \in \mathbb{Z}_+$ with $m \leqslant n$, f satisfies the inequalities $f(m) \leqslant f(n)$ and $f(m+n) \leqslant f(m) + f(n)$. Then*

$$\lim_{N \longrightarrow \infty} \frac{f(N)}{N} = \inf_{N \in \mathbb{Z}_+} \frac{f(N)}{N}.$$

Proof. For simplicity, we write $C = \inf_{N \in \mathbb{Z}_+} \frac{f(N)}{N}$ and $f(0) = 0$. Let $\varepsilon \in \mathbb{R}_+$ be arbitrary. Then there exists a number $N_0 \in \mathbb{Z}_+ \setminus \{1\}$ such that $\frac{f(N_0)}{N_0} < C + \varepsilon$. Let N be an integer larger than N_0. The division equation gives a positive integer s and a nonnegative integer r such that $N = sN_0 + r$ and $r < N_0$.

A direct application of the subadditivity condition satisfied by f yields

$$\begin{aligned} C \leqslant \frac{f(N)}{N} =& \frac{f(sN_0+r)}{N} \leqslant \frac{sf(N_0)+f(r)}{N} \\ &\leqslant \frac{f(N_0)}{N_0} + \frac{\max\{f(0), f(1), \ldots, f(N_0-1)\}}{N} \\ &< C + \varepsilon + \frac{\max\{f(0), f(1), \ldots, f(N_0-1)\}}{N} \xrightarrow[N\longrightarrow\infty]{} C+\varepsilon. \end{aligned}$$

Hence $\lim\limits_{N\longrightarrow\infty} \dfrac{f(N)}{N} = C$. q.e.d.

This lemma immediately guarantees the existence of the limit

$$C_k \stackrel{\text{def}}{=} \lim_{N\longrightarrow\infty} \frac{r_k(N)}{N}$$

for every $k \in \mathbb{Z}_+$, since each of the functions r_k is easily seen to be subadditive. It is clear that $0 \leqslant C_1 \leqslant C_2 \leqslant C_3 \leqslant \ldots$ and that $C_k \leqslant 1$ for each $k \in \mathbb{Z}_+$, so that the limit $C = \lim_{k\longrightarrow\infty} C_k \in [0,1]$ exists. In this setting, Roth's theorem takes the form $C_3 = 0$, and Szemerédi's theorem reduces to the equality $C = 0$.

Perhaps the first published result on the Erdős–Turán constants after the paper [Er&T] is the following theorem of Behrend mentioned in [Polla, p. 239]. In his 1938 article [Beh1] Behrend proved that $C \in \{0,1\}$. That is, that either $C = C_1 = C_2 = \ldots = 0$, or $C_1 \leqslant C_2 \leqslant \ldots$ and $C = 1$.

Other forms

Fix some $k \in \mathbb{Z}_+$. For any finite Abelian group G we may define the analog $r_k(G)$ of the kth Erdős–Turán constant as

$$r_k(G) \stackrel{\text{def}}{=} \max\big\{\#A \big| A \subseteq G \text{ does not contain a proper arithmetic progression of length } k\big\}.$$

For cyclic groups $G = \mathbb{Z}_N$ ($N \in \mathbb{Z}_+$) we have the obvious inequalities

$$r_k\Big(\Big\lfloor \frac{N}{k} \Big\rfloor\Big) \leqslant r_k(\mathbb{Z}_N) \leqslant r_k(N),$$

showing that the case $r_k(N) = o(N)$ of Szemerédi's theorem could as well be given in the form $r_k(\mathbb{Z}_N) = o(N)$.

Many approaches to Roth's theorem handle it in a much more general form inspired by a theorem of Varnavides which we will discuss later. The statement is as follows:

> *Let $N \in \mathbb{Z}_+$ be odd and suppose that $f\colon \mathbb{Z}_N \longrightarrow [0,\infty[$ does not vanish identically. Then*
>
> $$\mathop{\mathscr{E}}_{x}\mathop{\mathscr{E}}_{d} f(x)\, f(x+d)\, f(x+2d) = \Omega_{\mathscr{E}f}(1).$$

For the characteristic functions $f = \chi_A$ ($A \subseteq \mathbb{Z}_N$) this implies the original theorem of Varnavides. The functional on the left-hand side and its generalizations will appear rather frequently later on.

To conclude this section, we mention that the theorems of Roth and Szemerédi, together with their various quantitative forms, also generalize to groups more general than the finite cyclic groups. The book [T&V] of Tao and V. Vu discusses these questions systematically. The finite group and finite field analogs of these theorems are not only intrinsically interesting, but for suitably chosen groups or finite vector spaces, the strategies used in proofs of Roth's or Szemerédi's theorems take cleaner and more straightforward forms, perhaps thereby revealing more of what is really essential in the related arguments. This point of view is discussed in detail in [Gre2].

Dirichlet's lemma on rational approximation

Now we pause to introduce Dirichlet's classic lemma on rational approximation. One of the standard references is the book [Chand].

Dirichlet's lemma. *For any $\delta \in]0,1[$ and any $\vartheta \in \mathbb{R}$ there exists a number $d \in \mathscr{J}_{\lfloor \frac{1}{\delta} \rfloor}$ for which $\|d\vartheta\| \leqslant \delta$.*

The proof is easy: Write N for $\lfloor \frac{1}{\delta} \rfloor$ and consider the N fractional parts

$$\{1\vartheta\}, \{2\vartheta\}, \dots, \{N\vartheta\}, \tag{α}$$

and the disjoint intervals

$$\left[0, \frac{1}{N+1}\right[, \left[\frac{1}{N+1}, \frac{2}{N+1}\right[, \dots, \left[\frac{N}{N+1}, 1\right[.$$

If one of the numbers in the list (α), say $\{i\vartheta\}$, where $i \in \mathscr{J}_N$, lies in the first or the last of these intervals, then we may simply choose $d = i$. Otherwise all the N numbers lie in the remaining $N-1$ intervals and so Dirichlet's pigeonhole principle gives us two distinct numbers $i \in \mathscr{J}_N$ and $j \in \mathscr{J}_N$ for which the distance between the fractional parts of $i\vartheta$ and $j\vartheta$ is at most $\frac{1}{N+1}$, so that we can choose $d = |i-j|$. q.e.d.

Applying this result with $\delta = \frac{1}{\sqrt{N}}$ for some $N \in \mathbb{Z}_+$ we immediately see that the lemma may also be given in the form:

Dirichlet's lemma. *Suppose that $\vartheta \in \mathbb{R}$ and $N \in \mathbb{Z}_+$. Then there exist numbers $h \in \mathbb{Z}$, $q \in \mathbb{Z}_+$ and $\varrho \in \mathbb{R}$ such that*

$$(h,q) = 1, \quad q \leqslant \sqrt{N}, \quad q|\varrho| \leqslant \frac{1}{\sqrt{N}}, \quad \textit{and} \quad \vartheta = \frac{h}{q} + \varrho.$$

The original proof of Roth's theorem

The most important conjecture in the paper [Er&T] was without any doubt the statement

"It is probable that $r_3(N) = o(N)$."

In his 1952 note [Ro1] Roth showed that this is true, and in his 1953 paper [Ro2] he finished the proof with some more accuracy to get the result

$$r_3(N) \ll \frac{N}{\log\log N}. \qquad (N \longrightarrow \infty)$$

His proof was an adaptation of the circle method and basically considers some integrals and asymptotic relations. This is in contrast with the modern way of presenting this Fourier-analytic approach in the form of a density increment argument. Gowers' proof of Szemerédi's theorem generalizes the Fourier analytic approach and it is thus perhaps appropriate to present the original proof. We closely follow the presentation of [Ro1], occasionally taking advantage of the less concise presentation of [Ro2].

Our goal is to prove that

$$C_3 = \lim_{N\longrightarrow\infty} \frac{r_3(N)}{N} = 0.$$

For this purpose, we will consider maximal progression-free subsets A of $\mathscr{J}_N$ and, in particular, exponential sums of the form

$$f(\alpha) = \sum_{x\in A} e(\alpha x)\,. \qquad (\alpha \in \mathbb{R})$$

As we will see later, we are basically manipulating the discrete Fourier transform of the characteristic function of A. The core idea of the proof is that a maximal progression-free set is rather uniformly distributed over $\mathscr{J}_N$ and thus the functions similar to f, which appear in the integral which counts arithmetic triples in a progression-free set, may be replaced by uniform exponential sums such as

$$F(\alpha) = \frac{r_3(N)}{N} \sum_{\ell=1}^{N} e(\alpha\ell)\,. \qquad (\alpha \in \mathbb{R})$$

One then proceeds to show that such uniform sets in general contain a lot of arithmetic triples, and therefore the maximal progression-free sets have to be rather small. The theorem of Varnavides, which we will discuss in the next section, perhaps substantiates this ideology further.

Act one

Suppose that $N \in \mathbb{Z}_+$, let $A \subseteq \mathscr{J}_N$ be a progression-free set, and let $\alpha \in \mathbb{R}$ be arbitrary. The theorem of Dirichlet now tells us that there exist numbers $h \in \mathbb{Z}$, $q \in \mathbb{Z}_+$ and $\beta \in \mathbb{R}$ such that $(h,q) = 1$, $q \leqslant \sqrt{N}$, $q\big|\beta\big| \leqslant \frac{1}{\sqrt{N}}$, and

$$\alpha = \frac{h}{q} + \beta.$$

Roth's original proof consists of two parts. For now, we focus on proving that for arbitrary $m \in \mathscr{J}_{N-1}$, the sums

$$S = \sum_{x\in A} e(\alpha x)\,, \quad \text{and} \quad S' = \frac{r_3(m)}{mq} \sum_{\nu=1}^{q} e\Big(\frac{h\nu}{q}\Big) \sum_{n=1}^{N} e(\beta n)$$

satisfy the inequality

$$\big|S - S'\big| < \frac{Nr_3(m)}{m} - \#A + O\Big(m\sqrt{N}\Big)\,. \qquad (\beta)$$

Each element $x \in A$ belongs to exactly mq intervals of the form $[n, n+mq[$ ($n \in \mathbb{Z}_+$), except for the at most mq elements x of A for which $x < mq$. Thus,

$$S = \sum_{x\in A} e(\alpha x) = \frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N}\sum_{\substack{x\in A\cap[n,n+mq[\\ x\equiv\nu\,(\mathrm{mod}\,q)}} e(\alpha x) + O(mq)\,.$$

Write the number of terms in the innermost sum in the form $r_3(m) - D_{nmq\nu}$. The set $A\cap[n, n+mq[$ is progression-free and therefore there are at most $r_3(mq)$ terms in the innermost sums, so that $D_{nmq\nu} \geqslant 0$. Using the rational approximation $\alpha = \frac{h}{q} + \beta$ each term $e(\alpha x)$ of the innermost sum can be written as

$$\begin{aligned} e(\alpha x) = e\Big(\frac{hx}{q}\Big)\,e(\beta x) &= e\Big(\frac{h\nu}{q}\Big)\,e(\beta n) + e\Big(\frac{h\nu}{q}\Big)\,\big(e(\beta x) - e(\beta n)\big) \\ &= e\Big(\frac{h\nu}{q}\Big)\,e(\beta n) + O\big(mq\big|\beta\big|\big)\,. \end{aligned}$$

We get

$$\begin{aligned} S &= \frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N}\sum_{\substack{x\in A\cap[n,n+mq[,\\ x\equiv\nu\,(\mathrm{mod}\,q)}} e(\alpha x) + O(mq) \\ &= \frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N}\sum_{\substack{x\in A\cap[n,n+mq[\\ x\equiv\nu\,(\mathrm{mod}\,q)}} \Big(e\Big(\frac{h\nu}{q}\Big)\,e(\beta n) + O\big(mq\big|\beta\big|\big)\Big) + O(mq) \\ &= \frac{r_3(m)}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N} e\Big(\frac{h\nu}{q}\Big)\,e(\beta n) \\ &\qquad - \frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N} e\Big(\frac{h\nu}{q}\Big)\,e(\beta n)\,D_{nmq\nu} + O\big(Nmq\big|\beta\big|\big) + O(mq)\,. \end{aligned}$$

Since α was arbitrary and we have not made any use of the coprimality of h and q so far, this last relation must also hold for $\alpha = \frac{0}{q} + 0$, i.e. for $h = \beta = 0$. Thus

$$\frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N} D_{nmq\nu} = \frac{Nr_3(m)}{m} - \#A + O(mq)\,,$$

and we get

$$\begin{aligned}\big|S-S'\big| &= \left|-\frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N} e\Big(\frac{h\nu}{q}\Big)\, e(\beta n)\, D_{nmq\nu} + O\big(Nmq\big|\beta\big|\big) + O(mq)\right| \\ &\leqslant \frac{1}{mq}\sum_{\nu=1}^{q}\sum_{n=1}^{N} D_{nmq\nu} + O\big(Nmq\big|\beta\big|\big) + O(mq) \\ &= \frac{Nr_3(m)}{m} - \#A + O(mq) + O\big(Nmq\big|\beta\big|\big) + O(mq) \\ &= \frac{Nr_3(m)}{m} - \#A + O\Big(m\sqrt{N}\Big)\,,\end{aligned}$$

which is precisely (β).

Act two

Now that we have armed ourselves with the inequality (β), we may proceed to the second part of the proof.

Let $\varepsilon \in \left]0,1\right[$ be arbitrary. As one might expect, we will let $\varepsilon \longrightarrow 0+$ in the very end. Next, let m be a positive integer so large that

$$0 \leqslant \frac{r_3(n)}{n} - C_3 < \varepsilon$$

for every integer $n \geqslant m$. The lemma which was proven in p. 18 shows that such a number m exists, and furthermore, we may let m tend to infinity when the time is ripe. We also let N denote some positive integer larger than m. Again, we will let $N \longrightarrow \infty$. To finish the setup, we suppose that A is a progression-free subset of $\mathscr{J}_{2N}$ having cardinality $r_3(2N)$.

Next we introduce some notation. We let B stand for the set of integers x for which $2x \in A$. The set B is obviously a progression-free subset of $\mathscr{J}_N$ and thus has cardinality at most $r_3(N)$. On the other hand, $A \setminus 2B$ is a progression-free subset of the arithmetic progression $2\mathscr{J}_N - 1$ which has length N, and so $\#B \geqslant r_3(2N) - r_3(N)$. Using the definition of m we see that

$$\begin{aligned}\left|\frac{2Nr_3(m)}{m} - r_3(2N)\right| &\leqslant 2N\left(\left|\frac{r_3(m)}{m} - C_3\right| + \left|C_3 - \frac{r_3(2N)}{2N}\right|\right) \\ &< 2N\,(\varepsilon+\varepsilon) \ll \varepsilon N,\end{aligned}$$

and that similarly,

$$\left|\frac{Nr_3(m)}{m} - \#B\right| \ll \varepsilon N,$$

since when $\dfrac{Nr_3(m)}{m} \leqslant \#B$, we have

$$\begin{aligned}\left|\frac{Nr_3(m)}{m} - \#B\right| &= \#B - \frac{Nr_3(m)}{m} \leqslant r_3(N) - \frac{Nr_3(m)}{m} \\ &\leqslant N\left(\left|\frac{r_3(N)}{N} - C_3\right| + \left|C_3 - \frac{r_3(m)}{m}\right|\right) < N\,(\varepsilon+\varepsilon) \ll \varepsilon N,\end{aligned}$$

and when $\dfrac{Nr_3(m)}{m} \geqslant \#B$, we have

$$\begin{aligned}\left|\frac{Nr_3(m)}{m} - \#B\right| &= \frac{Nr_3(m)}{m} - \#B \leqslant \frac{Nr_3(m)}{m} - r_3(2N) + r_3(N)\\ &= \frac{2Nr_3(m)}{m} - r_3(2N) - \frac{Nr_3(m)}{m} + r_3(N)\\ &\leqslant 2N\left(\left|\frac{r_3(m)}{m} - C_3\right| + \left|C_3 - \frac{r_3(2N)}{2N}\right|\right)\\ &\qquad + N\left(\left|\frac{r_3(m)}{m} - C_3\right| + \left|C_3 - \frac{r_3(N)}{N}\right|\right)\\ &< 2N\left(\varepsilon+\varepsilon\right) + N\left(\varepsilon+\varepsilon\right) \ll \varepsilon N.\end{aligned}$$

We now introduce the generating functions of A and B:

$$\begin{cases} f = \alpha \longmapsto \displaystyle\sum_{x\in A} e(\alpha x)\colon \mathbb{R} \longrightarrow \mathbb{C},\\ g = \alpha \longmapsto \displaystyle\sum_{x\in B} e(\alpha x)\colon \mathbb{R} \longrightarrow \mathbb{C},\end{cases}$$

and their uniform substitutes

$$\begin{cases} F = \alpha \longmapsto \dfrac{r_3(m)}{m}\displaystyle\sum_{n=1}^{2N} e(\alpha n)\colon \mathbb{R} \longrightarrow \mathbb{C}, \quad \text{and}\\ G = \alpha \longmapsto \dfrac{r_3(m)}{m}\displaystyle\sum_{n=1}^{N} e(\alpha n)\colon \mathbb{R} \longrightarrow \mathbb{C}.\end{cases}$$

The rest of the proof will revolve around estimating the integral

$$\mathscr{I} = \int_{-\eta}^{1-\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha,$$

where η is shorthand for $\frac{1}{\sqrt{\varepsilon}N}$. We must have $0 < \eta < \frac{1}{2}$ for sufficiently large values of N, and we assume that this is the case. The meaning of $\mathscr{I}$ is easily seen:

$$\begin{aligned}\int_{-\eta}^{1-\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha &= \int_{-\eta}^{1-\eta} \sum_{y\in A} e(\alpha y) \sum_{x\in B} e(-\alpha x) \sum_{z\in B} e(-\alpha z)\,\mathrm{d}\alpha\\ &= \sum_{x\in B}\sum_{y\in A}\sum_{z\in B} \int_{-\eta}^{1-\eta} e\big(-(x-y+z)\alpha\big)\,\mathrm{d}\alpha = \sum_{\substack{x\in B\\ x-y+z=0}}\sum_{y\in A}\sum_{z\in B} 1,\end{aligned}$$

In other words, the integral $\mathscr{I}$ counts the arithmetic triples $\langle 2x, y, 2z\rangle$ of A with $x \in B$ and $z \in B$. Since A is progression-free, there are exactly $\#B$ such triples, namely the trivial arithmetic triples of $2B$. We conclude that

$$\mathscr{I} = \#B \leqslant N.$$

The time has come to estimate the error caused by the substitution of f by F and g by G in the integral $\mathscr{I}$. In the spirit of the circle method, we split the interval of interest $]-\eta,1-\eta[$ into two cases $]-\eta,\eta[$ and $]\eta,1-\eta[$ and consider them separately.

First we suppose that $\alpha\in\,]-\eta,\eta[$. Then $\alpha=\frac{0}{1}+\alpha$ with 0 and 1 coprime, $1\leqslant\sqrt{N}\leqslant\sqrt{2N}$, and $\big|\alpha\big|<\eta=\frac{1}{\sqrt{\varepsilon}N}$. In addition, for sufficiently large values of N, we have $\big|\alpha\big|<\frac{1}{\sqrt{N}}$. Therefore the sum S' corresponding to the current set $A\subseteq\mathscr{J}_{2N}$ and the current value of α reduces to

$$S'=\frac{r_3(m)}{m}\sum_{n=1}^{2N}e(\alpha n)=F(\alpha)\,,$$

and the inequality (β) says that

$$\big|f(\alpha)-F(\alpha)\big|<\frac{2Nr_3(m)}{m}-r_3(2N)+O\Big(m\sqrt{2N}\Big)\ll\varepsilon N+m\sqrt{N}.$$

Similarly, the sums S and S' corresponding to the set $B\subseteq\mathscr{J}_N$ are

$$S=\sum_{x\in B}e(\alpha x)=g(\alpha)\,,\quad\text{and}\quad S'=\frac{r_3(m)}{m}\sum_{n=1}^{N}e(\alpha n)=G(\alpha)\,,$$

so that we infer from (β) that

$$\big|g(\alpha)-G(\alpha)\big|<\frac{Nr_3(m)}{m}-\#B+O\Big(m\sqrt{N}\Big)\ll\varepsilon N+m\sqrt{N}.$$

Before we can proceed to the estimation of the integral of $f(\cdot)\,g^2(-\cdot)$ over the interval $]-\eta,\eta[$, we remark that

$$\begin{aligned}\int_{-\frac{1}{2}}^{\frac{1}{2}}F(\alpha)\,G^2(-\alpha)\,\mathrm{d}\alpha&=\Big(\frac{r_3(m)}{m}\Big)^3\int_{-\frac{1}{2}}^{\frac{1}{2}}\sum_{\varrho=1}^{2N}e(\alpha\varrho)\sum_{\sigma=1}^{N}e(-\alpha\sigma)\sum_{\tau=1}^{N}e(-\alpha\tau)\,\mathrm{d}\alpha\\&=\Big(\frac{r_3(m)}{m}\Big)^3\sum_{\varrho=1}^{2N}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\int_{-\frac{1}{2}}^{\frac{1}{2}}e\big((\varrho-\sigma-\tau)\alpha\big)\,\mathrm{d}\alpha\\&=\Big(\frac{r_3(m)}{m}\Big)^3\sum_{\substack{\varrho=1\\ \varrho-\sigma-\tau=0}}^{2N}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}1=\Big(\frac{r_3(m)}{m}\Big)^3\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}1=\Big(\frac{r_3(m)}{m}\Big)^3N^2,\end{aligned}$$

and we mention the following simple lemma:

Lemma. *For any $L\in\mathbb{Z}_+$ and arbitrary $\vartheta\in\mathbb{R}\setminus\mathbb{Z}$, we have*

$$\left|\sum_{\ell=1}^{L}e(\vartheta\ell)\right|\leqslant\frac{1}{\|\vartheta\|}.$$

The proof is straightforward:

$$\left|\sum_{\ell=1}^{L} e(\vartheta\ell)\right| = \left|\frac{e(L\vartheta)-1}{e(\vartheta)-1}\right| \leqslant \frac{2}{\bigl|e(\vartheta)-1\bigr|} = \frac{2}{\sqrt{(\cos 2\pi\vartheta - 1)^2 + \sin^2 2\pi\vartheta}}$$
$$= \frac{2}{\sqrt{2-2\cos 2\pi\vartheta}} = \frac{1}{\bigl|\sin\pi\vartheta\bigr|} \leqslant \frac{1}{\bigl\|\vartheta\bigr\|}. \quad \text{q.e.d.}$$

We are now in a position to estimate the integral

$$\int_{-\eta}^{\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha = \int_{-\eta}^{\eta} F(\alpha)\, G^2(-\alpha)\,\mathrm{d}\alpha + \int_{-\eta}^{\eta} \bigl(f(\alpha)-F(\alpha)\bigr)\, G^2(-\alpha)\,\mathrm{d}\alpha$$
$$+ \int_{-\eta}^{\eta} f(\alpha)\,\bigl(g(-\alpha)+G(-\alpha)\bigr)\,\bigl(g(-\alpha)-G(-\alpha)\bigr)\,\mathrm{d}\alpha.$$

We extend the domain of integration of the first integral to $\left]-\frac{1}{2},\frac{1}{2}\right[$ and estimate the resulting error by applying the above lemma to the functions F and G in the intervals $\left]-\frac{1}{2},-\eta\right[$ and $\left]\eta,\frac{1}{2}\right[$. In the second integral, we apply the bound we obtained for the difference $\bigl|f(\cdot)-F(\cdot)\bigr|$ in the interval $]-\eta,\eta[$ and the trivial estimate

$$\bigl|G(-\alpha)\bigr| \ll N.$$

In the third integral we use the bound we obtained for the difference of g and G in the interval $]-\eta,\eta[$ and apply trivial estimates to the rest of the integral. We get

$$\int_{-\eta}^{\eta} f(\alpha)\, g^2(\alpha)\,\mathrm{d}\alpha = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(\alpha)\, G^2(-\alpha)\,\mathrm{d}\alpha - \left(\int_{-\frac{1}{2}}^{-\eta} + \int_{\eta}^{\frac{1}{2}}\right) F(\alpha)\, G^2(-\alpha)\,\mathrm{d}\alpha$$
$$+ O\Bigl(\eta\bigl(\varepsilon N + m\sqrt{N}\bigr)N^2\Bigr) + O\Bigl(\eta\cdot N\cdot N\cdot\bigl(\varepsilon N + m\sqrt{N}\bigr)\Bigr)$$
$$= \left(\frac{r_3(m)}{m}\right)^3 N^2 + O\left(\int_{\eta}^{\frac{1}{2}} \frac{\mathrm{d}\alpha}{\alpha^3}\right) + O\Bigl(\eta N^2\bigl(\varepsilon N + m\sqrt{N}\bigr)\Bigr)$$
$$= \left(\frac{r_3(m)}{m}\right)^3 N^2 + O\Bigl(\frac{1}{\eta^2} + \eta N^2\bigl(\varepsilon N + m\sqrt{N}\bigr)\Bigr).$$

Next we suppose that $\alpha \in]\eta, 1-\eta[$. Approximate

$$\alpha = \frac{h}{q} + \beta,$$

with $h\in\mathbb{Z}$, $q\in\mathbb{Z}_+$, $\beta\in\mathbb{R}$, $(h,q)=1$, $q\leqslant\sqrt{2N}$, and $q\bigl|\beta\bigr| \leqslant \frac{1}{\sqrt{2N}}$. If $q=1$, we have $\bigl|\beta\bigr| = \alpha$, so that $\bigl\|\beta\bigr\| = \bigl\|\alpha\bigr\| > \eta$ and

$$\sum_{n=1}^{2N} e(\beta n) \ll \frac{1}{\bigl\|\beta\bigr\|} < \frac{1}{\eta} = \sqrt{\varepsilon}N,$$

and we get

$$\frac{r_3(m)}{mq}\sum_{\nu=1}^{q} e\Big(\frac{h\nu}{q}\Big)\sum_{n=1}^{2N} e(\beta n) \ll \frac{r_3(m)}{mq}\cdot q\cdot\sqrt{\varepsilon}N \ll \sqrt{\varepsilon}N.$$

If $q>1$ instead, then the sum $\displaystyle\sum_{\nu=1}^{q} e\Big(\frac{h\nu}{q}\Big)$ simply vanishes, so that

$$\frac{r_3(m)}{mq}\sum_{\nu=1}^{q} e\Big(\frac{h\nu}{q}\Big)\sum_{n=1}^{2N} e(\beta n) = 0 \ll \sqrt{\varepsilon}N.$$

Therefore (β) gives us the relation

$$\begin{aligned} f(\alpha)=\sum_{x\in A} e(\alpha x) &\ll \left|\frac{r_3(m)}{mq}\sum_{\nu=1}^{q} e\Big(\frac{h\nu}{q}\Big)\sum_{n=1}^{2N} e(\beta n)\right| \\ &\quad + \left|\frac{2Nr_3(m)}{m} - r_3(2N)\right| + m\sqrt{N} \\ &\ll \sqrt{\varepsilon}N + \varepsilon N + m\sqrt{N} \ll \sqrt{\varepsilon}N + m\sqrt{N}. \end{aligned}$$

With this relation and the trivial estimate

$$\int_0^1 \big|g(\alpha)\big|^2 \mathrm{d}\alpha \leqslant N,$$

we get the estimate

$$\int_{-\eta}^{1-\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha \ll \Big(\sqrt{\varepsilon}N + m\sqrt{N}\Big)\int_0^1 \big|g\,(\alpha)\big|^2 \mathrm{d}\alpha \ll \sqrt{\varepsilon}N^2 + mN\sqrt{N}.$$

Combining all the above observations gives

$$\begin{aligned} N \geqslant \#B &= \int_{-\eta}^{1-\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha = \int_{-\eta}^{\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha + \int_{\eta}^{1-\eta} f(\alpha)\, g^2(-\alpha)\,\mathrm{d}\alpha \\ &= \left(\frac{r_3(m)}{m}\right)^3 N^2 + O\Big(\frac{1}{\eta^2} + \eta N^2\big(\varepsilon N + m\sqrt{N}\big)\Big) + O\Big(\sqrt{\varepsilon}N^2 + mN\sqrt{N}\Big) \\ &= \left(\frac{r_3(m)}{m}\right)^3 N^2 + O\Big(\varepsilon N^2 + \sqrt{\varepsilon}N^2 + \frac{1}{\sqrt{\varepsilon}}mN\sqrt{N} + \sqrt{\varepsilon}N^2 + mN\sqrt{N}\Big) \\ &= \left(\frac{r_3(m)}{m}\right)^3 N^2 + O\left(\sqrt{\varepsilon}N^2 + \frac{mN\sqrt{N}}{\sqrt{\varepsilon}}\right), \end{aligned}$$

and we have

$$\left(\frac{r_3(m)}{m}\right)^3 \ll \sqrt{\varepsilon} + \frac{m}{\sqrt{\varepsilon N}}.$$

Letting $N \longrightarrow \infty$ gives

$$\left(\frac{r_3(m)}{m}\right)^3 \ll \sqrt{\varepsilon},$$

and letting $m \longrightarrow \infty$ gives

$$C_3^3 \ll \sqrt{\varepsilon}.$$

Finally, letting $\varepsilon \longrightarrow 0+$, we see that $C_3 = 0$. Q.E.D.

In [Ro2] Roth uses the exact same approach but this time the goal is to obtain a functional inequality for r_3, from which he then derives the classical quantitative bound $r_3(N) \ll \frac{N}{\log\log N}$ after one page of tedious calculations. We omit the details here as we will prove the same upper bound later in this chapter in a more transparent way.

The presentation in [Vau] is similar in spirit, but much simpler as instead of exchanging each of the factors in the integrand $f(\alpha)\, g^2(-\alpha)$ into a more evenly distributed one, there one changes only one of the factors. Consequently the quantitative bound is obtained in just a few lines of reverse engineering.

In fact, Roth's original method of proof easily generalises to sets of integers not containing elements x_1, x_2, …, x_n ($n \in \mathbb{Z}_+$) satisfying relations

$$\sum_{\ell=1}^{n} a_{\mu\ell} x_\ell = 0, \qquad \big(\mu \in \{1, 2, \ldots, \ell\}, \quad \ell \in \mathbb{Z}_+\big)$$

where $\big[a_{\mu\ell}\big]$ is a suitable $\ell \times n$ matrix with integral elements. Roth describes these generalizations in his 1954 paper [Ro3]. It is curious that this method just barely fails to prove Szemerédi's theorem for four-term progressions. Only after Szemerédi's paper [Sz1] Roth managed to adapt his Fourier-analytic approach to four-term progressions by using some combinatorial ideas of Szemerédi [Ro4, Ro5, Ro6]. However, Roth does not consider the possible upper bounds the method could obtain. It is speculated in the survey article [Ko&S] that the method might yield an upper bound of the form $r_4(N) \ll \frac{N}{\log\log\cdots\log N}$, for some fixed number of iterated logarithms.

Varnavides' theorem

We now know for sure that for any density $\delta \in]0, 1[$, there exists an integer $N_\delta \in \mathbb{Z}_+$ with the property that for any integer N larger than N_δ, each subset A of $\mathscr{J}_N$, having cardinality at least δN, necessarily contains **at least one** proper arithmetic progression of length three. However, it is intuitively clear that when N is very large, the set A is likely to contain many proper arithmetic triples. Thus it is natural to ask how much can one improve the lower bound "at least one"?

When the set A is constructed randomly by including any $x \in \mathscr{J}_N$ in A with uniform probability δ, then any triplet a, $a+d$, $a+2d$ of elements in $\mathscr{J}_N$ with $a, d \in \mathscr{J}_N$ is contained in A with a probability of magnitude δ^3, as the events $a \in A$, $a + d \in A$ and $a + 2d \in A$ are virtually independent. Thus it would be natural to expect A to have $\Omega_\delta\big(N^2\big)$ arithmetic triples. In the same vein, the different technical notions of quasirandomness of A imply the inclusion of a similar quantity of arithmetic triples. We remark that $\Omega_\delta\big(N^2\big)$ arithmetic

triples immediately yield $\Theta_\delta\big(N^2\big)$ arithmetic triples as there are trivially less than N^2 arithmetic triples in the whole of $\mathscr{J}_N$.

Soon after Roth's articles [Ro1] and [Ro2] the paper [Var1] by P. Varnavides appeared. In it Varnavides shows that one can assert the existence of $\Omega_\delta(N\log N)$ arithmetic triples. In 1959 he published another article [Var2] in which he improves the bound to $\Omega_\delta\big(N^2\big)$ arithmetic triples. That is, he proved the following theorem, which we also prove here by following the original presentation of the article [Var2].

Varnavides' theorem. *For any $\delta \in \,]0,1[$ there exists a constant $C \in \mathbb{R}_+$ such that for sufficiently large positive integers N, each subset A of $\mathscr{J}_N$, having cardinality at least δN, necessarily contains at least CN^2 proper arithmetic triples.*

Proof of Varnavides' theorem

Let $\delta \in \,]0,1[$ be arbitrary. Then Roth's theorem guarantees the existence of a positive integer $\ell > 8$ for which any subset of cardinality at least $\frac{\delta\ell}{2}$ of any arithmetic progression of length ℓ contains a proper arithmetic triple. For the present time, we fix $N \in \mathbb{Z}_+$ and $A \subseteq \mathscr{J}_N$ with $\#A \geqslant \delta N$.

For any $u \in \mathbb{Z}_+$ and $d \in \mathbb{Z}_+$ we define the arithmetic progression of length ℓ

$$P_{u,d} \stackrel{\text{def}}{=} \big\{u, u+d, u+2d, \ldots, u+(\ell-1)\,d\big\},$$

and the number

$$f(u,d) \stackrel{\text{def}}{=} \#(A \cap P_{u,d})\,.$$

We call the progression $P_{u,d}$ **admissible**, if

$$d < \frac{\delta N}{\ell^2}, \quad P_{u,d} \subseteq \mathscr{J}_N, \quad \text{and} \quad f(u,d) \geqslant \frac{\delta\ell}{2}.$$

The point of all of this is that there will turn out to be $\Omega_\delta\big(N^2\big)$ admissible progressions, each giving rise to a proper arithmetic triple in A.

Fix temporarily the positive integer $d < \frac{\delta N}{\ell^2}$ and define

$$G_d \stackrel{\text{def}}{=} \#\big\{u \in \mathbb{Z}_+ \,\big|\, P_{u,d} \text{ is admissible}\big\}.$$

The core of this proof is in the observation that

$$G_d > \frac{\delta N}{4}.$$

This is seen by estimating the sum $\sum_u f(u,d)$ both from above and below.

Since every $a \in A$, for which $a \geqslant \ell d$ and $a \leqslant N - \ell d$, is contained in exactly ℓ progressions $P_{u,d}$ (d is fixed), and since there are at least $\delta N - 2\ell d$ such elements a, we have

$$\sum_u f(u,d) \geqslant \ell\,(\delta N - 2\ell d) > \ell\left(\delta N - \frac{2\delta N}{\ell}\right) = \left(1 - \frac{2}{\ell}\right)\ell\delta N > \frac{3\ell\delta N}{4}.$$

On the other hand, since there are G_d progressions $P_{u,d}$ containing at least $\frac{\delta\ell}{2}$ elements of A, and since there are at most N such progressions with less than $\frac{\delta\ell}{2}$ elements of A, we trivially have the inequality

$$\sum_u f(u,d) < N \cdot \frac{\delta\ell}{2} + G_d \cdot \ell.$$

Combining these two inequalities gives us

$$G_d > \frac{1}{\ell}\left(\frac{3\ell\delta N}{4} - \frac{\delta\ell N}{2}\right) = \frac{\delta N}{4}.$$

Now we see that for $N > \frac{2\ell^2}{\delta}$ there are at least

$$\sum_{d<\frac{\delta N}{\ell^2}} G_d > \sum_{d<\frac{\delta N}{\ell^2}} \frac{\delta N}{4} = \left(\left\lceil \frac{\delta N}{\ell^2} \right\rceil - 1\right)\frac{\delta N}{4} \geqslant \frac{\delta^2 N^2}{4\ell^2} - \frac{\delta N}{4}$$

$$= \frac{\delta^2 N^2}{8\ell^2} + N\left(\frac{\delta^2 N}{8\ell^2} - \frac{\delta}{4}\right) \geqslant \frac{\delta^2 N^2}{8\ell^2} \gg_\delta N^2$$

admissible progressions, and by the definition of ℓ, each of them contains at least one proper arithmetic triple also fully contained in A. We are not yet quite done, as not all of these arithmetic triples are distinct. However, this is not a significant obstacle as arbitrarily chosen one of these, one having a common difference $d < \frac{\delta N}{\ell^2}$, say, is contained in at most $\ell - 2$ admissible progressions with the same common difference d and in at most $\frac{\ell}{2}(\ell-2)$ admissible progressions with common difference at most d. This is so because only for common differences $\partial \in \mathbb{Z}_+$ of the form $\frac{d}{t}$, with positive integers $t < \frac{\ell}{2}$, can $P_{u,\partial}$ contain the original arithmetic triple.

As ℓ depends only on δ, we conclude that for any $N > \frac{2\ell^2}{\delta}$, the set A contains at least

$$\frac{2}{\ell\,(\ell-2)} \cdot \frac{\delta^2 N^2}{8\ell^2} \gg_\delta N^2$$

proper arithmetic triples. Q.E.D.

The discrete Fourier transform

We need to deal with a multitude of complex valued functions defined on $\mathbb{Z}_N$ for suitably chosen values of $N \in \mathbb{Z}_+$. In this section we introduce the useful concept of discrete Fourier transforms of such functions. The chapter 4 of the book [T&V] contains an extended discussion of Fourier analysis from the point of view of additive combinatorics. In order to simplify our notation, we fix a number $N \in \mathbb{Z}_+$ and two functions $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ and $g\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$. We recall that

$$\mathop{\mathscr{E}}_{x\in\mathbb{Z}_N} f(x) = \frac{1}{N}\sum_{x\in\mathbb{Z}_N} f(x),$$

and that we often leave out the ranges of the indices.

The **discrete Fourier transform** of f is defined to be the function

$$\widehat{f} \stackrel{\text{def}}{=} \xi \longmapsto \mathop{\mathscr{E}}_{x} f(x)\, e\Big(\frac{-x\xi}{N}\Big)\colon \mathbb{Z}_N \longrightarrow \mathbb{C},$$

where the value of

$$e\Big(\frac{-x\xi}{N}\Big) = e^{-\frac{2\pi i x\xi}{N}}$$

is defined in the natural way. Sometimes it is necessary to use the inferior notation $f^\wedge$ instead of $\widehat{f}$.

We also further simplify the notation by defining the **convolution** of f and g as the function

$$f * g \stackrel{\text{def}}{=} x \longmapsto \mathop{\mathscr{E}}_{y} f(y)\, g(x-y)\colon \mathbb{Z}_N \longrightarrow \mathbb{C},$$

and by defining for any $p \in [1, \infty[$ the $\mathscr{L}^p$**-** and ℓ^p**-norms** of f by the formulas

$$\big\|f\big\|_{\mathscr{L}^p} \stackrel{\text{def}}{=} \sqrt[p]{\mathop{\mathscr{E}}_{x} \big|f(x)\big|^p}, \quad \text{and} \quad \big\|f\big\|_{\ell^p} \stackrel{\text{def}}{=} \sqrt[p]{\sum_{\xi} \big|f(\xi)\big|^p}, \quad \text{respectively.}$$

We also define

$$\big\|f\big\|_{\mathscr{L}^\infty} \stackrel{\text{def}}{=} \big\|f\big\|_{\ell^\infty} \stackrel{\text{def}}{=} \max_{x\in\mathbb{Z}_N} \big|f(x)\big|.$$

From the simple observation that for any $x \in \mathbb{Z}_N$ we have

$$\sum_{\xi} e\Big(\frac{-x\xi}{N}\Big) = \begin{cases} N, & \text{if } x = 0, \\ 0, & \text{otherwise,} \end{cases}$$

we easily deduce the discrete Fourier inversion formula

$$\begin{aligned} \sum_{\xi} \widehat{f}(\xi)\, e\Big(\frac{x\xi}{N}\Big) &= \sum_{\xi} \mathop{\mathscr{E}}_{y} f(y)\, e\Big(\frac{-y\xi}{N}\Big)\, e\Big(\frac{x\xi}{N}\Big) \\ &= \mathop{\mathscr{E}}_{y} f(y) \sum_{\xi} e\Big(\frac{(x-y)\,\xi}{N}\Big) = f(x)\,, \end{aligned}$$

which holds for all $x \in \mathbb{Z}_N$, and the following analog of Parseval's formula:

$$\begin{aligned} \sum_{\xi} \widehat{f}(\xi)\, \overline{\widehat{g}(\xi)} &= \sum_{\xi} \mathop{\mathscr{E}}_{x} f(x)\, e\Big(\frac{-x\xi}{N}\Big)\, \overline{\mathop{\mathscr{E}}_{y} g(y)\, e\Big(\frac{-y\xi}{N}\Big)} \\ &= \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{y} f(x)\, \overline{g(y)} \sum_{\xi} e\Big(\frac{-\,(x-y)\,\xi}{N}\Big) = \mathop{\mathscr{E}}_{x} f(x)\, \overline{g(x)}. \end{aligned}$$

As a corollary we obtain an analog for Plancherel's identity as well:

$$\big\|f\big\|_{\mathscr{L}^2} = \big\|\widehat{f}\,\big\|_{\ell^2}.$$

We close this section with the remark that for any $\xi \in \mathbb{Z}_N$ we have

$$\begin{aligned} \widehat{f * g}(\xi) &= \mathop{\mathscr{E}}_{x} \big(f * g\big)(x)\, e\Big(\frac{-x\xi}{N}\Big) \\ &= \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{y} f(y)\, g(x-y)\, e\Big(\frac{-y\xi}{N}\Big)\, e\Big(\frac{-(x-y)\xi}{N}\Big) \\ &= \mathop{\mathscr{E}}_{z} \mathop{\mathscr{E}}_{y} f(y)\, e\Big(\frac{-y\xi}{N}\Big)\, g(z)\, e\Big(\frac{-z\xi}{N}\Big) = \widehat{f}(\xi)\, \widehat{g}(\xi)\,. \end{aligned}$$

A modern density increment proof of Roth's theorem

In this section we prove Roth's theorem with a density increment argument. Despite it being much shorter and easier, the structure of the proof has many similarities to Gowers' proof of Szemerédi's theorem, the discussion of which will occupy us in the following chapters. We primarily follow the presentations of [Go4] and [Gre8]. The precise conclusion is the classic

$$r_3(N) \ll \frac{N}{\log\log N}, \qquad (N \longrightarrow \infty)$$

or in other words:

Roth's theorem. *For sufficiently large integers N, any subset of $\mathscr{J}_N$ with cardinality at least $\frac{CN}{\log\log N}$ contains a proper arithmetic triple. Here $C \in \mathbb{R}_+$ is an absolute constant independent of N.*

The proof is based on

The density increment strategy. *Let $\alpha \in]0,1[$. Suppose that $P \subseteq \mathbb{Z}$ is an arithmetic progression with cardinality greater than $C\alpha^{-C}$, and that A is a subset of P with cardinality $\alpha\#P$. If A does not contain a proper arithmetic triple, then there is a subprogression of P of length at least $\sqrt[3]{\#P}$ in which A has density at least $\alpha + c\alpha^2$. Here $c \in]0,1[$ and $C \in]1,\infty[$ are absolute constants independent of α.*

We begin by deducing Roth's theorem from this result. Let c and C be as in the above statement, and let $\alpha \in]0,1[$ and let $N \in \mathbb{Z}_+$ be such that $N > C\alpha^{-C}$. Suppose that A is a progression-free subset of $\mathscr{J}_N$ with cardinality at least αN. Then we may iterate the density increment argument to get a decreasing sequence of arithmetic progressions

$$\mathscr{J}_N \supseteq Q_1 \supseteq Q_2 \supseteq \dots$$

such that

$$\sqrt[3]{N} \leqslant \#Q_1, \quad \sqrt[3]{\#Q_1} \leqslant \#Q_2, \quad \dots,$$

and that the density of A increases by at least $c\alpha^2$ in each iteration.

We continue this iteration until some of the conditions for the density increment fail. Since the density of A doubles in at most $\frac{1}{c\alpha}$ iterations, it must reach one in at most

$$\frac{1}{c\alpha} + \frac{1}{2c\alpha} + \frac{1}{4c\alpha} + \dots = \frac{2}{c\alpha}$$

iterations. Hence the sequence $Q_1, Q_2, \dots$ must be finite, consisting of, say, $m \in \mathbb{Z}_+$ elements. The only assumption that can fail after m iterations is the lower bound for $\#Q_m$. Hence $\#Q_m \leqslant C(\alpha + mc\alpha^2)^{-C}$. Now

$$N^{\left(\frac{1}{3}\right)^{\left(\frac{2}{c\alpha}\right)}} \leqslant N^{\left(\frac{1}{3}\right)^m} \leqslant \#Q_m \leqslant C\left(\alpha + mc\alpha^2\right)^{-C} \leqslant C\alpha^{-C},$$

so that

$$-\frac{2}{c\alpha}\log 3+\log\log N\leqslant\log\log\left(C\alpha^{-C}\right)\leqslant\log\left(C\log\frac{\sqrt[c]{C}}{\alpha}\right)\leqslant\frac{C\sqrt[c]{C}}{\alpha},$$

and finally:
$$\alpha\leqslant\left(\frac{2\log 3}{c}+C\sqrt[c]{C}\right)\frac{1}{\log\log N}.$$
Q.E.D.

Now we turn to the implementation of the density increment strategy. Suppose that α, P and A are as in the statement of the claim. Without loss of generality, we may suppose that $P=\mathscr{J}_N$, where $N\in\mathbb{Z}_+$ is such that $N>C\alpha^{-C}$. The actual values of c and C will be discovered later.

It is convenient to assume that the set

$$B\stackrel{\text{def}}{=}A\cap\left]\frac{N}{3},\frac{2N}{3}\right]$$

has cardinality at least $\frac{\alpha N}{5}$. If this is not the case, then one of the sets $A\cap\left]0,\frac{N}{3}\right]$ and $A\cap\left]\frac{2N}{3},N\right]$ will have at least $\frac{1}{2}\left(\alpha N-\frac{\alpha N}{5}\right)=\frac{4\alpha N}{10}$ elements, thus yielding the sought-for density increment in a subprogression of length $\left\lfloor\frac{N}{3}\right\rfloor$ or $\left\lceil\frac{N}{3}\right\rceil$.

We now embed P and A into $\mathbb{Z}_N$ via the canonical surjection

$$n\longmapsto n+N\mathbb{Z}\colon\mathbb{Z}\longrightarrow\mathbb{Z}_N,$$

and loosely consider them as subsets of $\mathbb{Z}_N$. We will later revert back to $\mathbb{Z}$. As we switch back and forth between $\mathbb{Z}$ and $\mathbb{Z}_N$ it is useful to say that a $\mathbb{Z}_N$-arithmetic progression is a **$\mathbb{Z}$-arithmetic progression** when its image under the mapping

$$x\longmapsto x\cap\mathscr{J}_N\colon\mathbb{Z}_N\longrightarrow\mathbb{Z}$$

is a true $\mathbb{Z}$-arithmetic progression.

As $A\subseteq\mathbb{Z}_N$, we may define the function $f_A=\chi_A-\alpha\chi_{\mathbb{Z}_N}$, called the **balanced function** of A. This is useful for two reasons. For technical reasons it is nice that $\mathscr{E}_x f_A(x)=0$. The other reason is that we are interested in large values of $\widehat{\chi_A}(\xi)$, but only for $\xi\neq 0$ as the value $\widehat{\chi_A}(0)=\alpha$ is uninteresting and quite large. Thus it is practical that we have $\widehat{f_A}(\xi)=\widehat{\chi_A}(\xi)$ for $\xi\in\mathbb{Z}_N\setminus\{0\}$ and yet $\widehat{f_A}(0)=0$.

The argument now splits into two cases depending on whether we have $\left\|\widehat{f_A}\right\|_{\ell^\infty}\leqslant\frac{\alpha^2}{20}$ or $\left\|\widehat{f_A}\right\|_{\ell^\infty}\geqslant\frac{\alpha^2}{20}$. This is a prime example of the phenomenon often called the "dichotomy between structure and randomness" (see e.g. [T5]). The idea is that given a fixed density, subsets of $\mathscr{J}_N$ of that density will contain non-trivial arithmetic triples, but for different sets this may have different causes. We have pointed out in our discussion of Varnavides' theorem (p. 28 onwards) that when a set is constructed by allowing arbitrary element of $\mathscr{J}_N$ in it with a given fixed positive density, it will in general contain many arithmetic triples. It is intuitively clear that such a random set is likely to have very small Fourier coefficients since in the exponential sums that define the Fourier transform of its balanced function the terms are all unimodular and their arguments are uniformly distributed to all directions, leading to a lot of cancellation.

This is precisely the fundamental idea behind the proof. If $\left\|\widehat{f_A}\right\|_{\ell^\infty}\leqslant\frac{\alpha^2}{20}$, then our set resembles a random set and some simple inequalities force it to have

so many arithmetic triples that the trivial ones can not account for all of them. Therefore we much have $\|\widehat{f_A}\|_{\ell^\infty} \geqslant \frac{\alpha^2}{20}$ and we may use this new structural information to obtain the density increment.

The quasirandom case

We first suppose that $\|\widehat{f_A}\|_{\ell^\infty} \leqslant \frac{\alpha^2}{20}$. For simplicity, we write $\#B = \beta N$, where $\beta \in]0,1]$. The number of arithmetic progressions x, y, z in A for which $y, z \in B$ is

$$\begin{aligned}
&\sum_{\substack{x\\ x-2y+z=0}} \sum_{y} \sum_{z} \chi_A(x)\,\chi_B(y)\,\chi_B(z) \\
&= \frac{1}{N} \sum_{x} \sum_{y} \sum_{z} \chi_A(x)\,\chi_B(y)\,\chi_B(z) \sum_{\xi} e\Big(\frac{-(x-2y+z)\xi}{N}\Big) \\
&= N^2 \sum_{\xi} \mathop{\mathscr{E}}_{x} \chi_A(x)\, e\Big(\frac{-x\xi}{N}\Big) \mathop{\mathscr{E}}_{y} \chi_B(y)\, e\Big(\frac{2y\xi}{N}\Big) \mathop{\mathscr{E}}_{z} \chi_B(z)\, e\Big(\frac{-z\xi}{N}\Big) \\
&= N^2 \sum_{\xi} \widehat{\chi_A}(\xi)\,\widehat{\chi_B}(-2\xi)\,\widehat{\chi_B}(\xi) \\
&= N^2 \widehat{\chi_A}(0)\,\widehat{\chi_B}^2(0) + N^2 \sum_{\xi\neq 0} \widehat{\chi_A}(\xi)\,\widehat{\chi_B}(-2\xi)\,\widehat{\chi_B}(\xi) \\
&= \alpha\beta^2 N^2 + N^2 \sum_{\xi\neq 0} \widehat{\chi_A}(\xi)\,\widehat{\chi_B}(-2\xi)\,\widehat{\chi_B}(\xi)\,.
\end{aligned}$$

Here the first term is at least $\frac{\alpha^2\beta N^2}{5}$, while the Cauchy-Schwarz-Bunyakovsky inequality and some elementary observations imply that for the second term

$$\begin{aligned}
&\Big| N^2 \sum_{\xi\neq 0} \widehat{\chi_A}(\xi)\,\widehat{\chi_B}(-2\xi)\,\widehat{\chi_B}(\xi) \Big| \\
&\leqslant N^2 \Big(\max_{\xi\neq 0} |\widehat{f_A}(\xi)| \Big) \sqrt{\sum_{\xi\neq 0} |\widehat{\chi_B}(-2\xi)|^2} \sqrt{\sum_{\xi\neq 0} |\widehat{\chi_B}(\xi)|^2} \\
&\leqslant N^2 \|\widehat{f_A}\|_{\ell^\infty} \cdot \sqrt{2} \left(\sum_{\xi} |\widehat{\chi_B}(\xi)|^2 \right) \leqslant 2N^2 \|\widehat{f_A}\|_{\ell^\infty} \|\widehat{\chi_B}\|_{\ell^2}^2 \\
&= 2N^2 \|\widehat{f_A}\|_{\ell^\infty} \|\chi_B\|_{\mathscr{L}^2}^2 = 2N^2 \|\widehat{f_A}\|_{\ell^\infty} \big(\sqrt{\beta}\big)^2 \leqslant 2N^2 \cdot \frac{\alpha^2}{20}\beta = \frac{\alpha^2\beta N^2}{10}.
\end{aligned}$$

The triangle inequality now gives the lower bound

$$\Big| \alpha\beta^2 N^2 + N^2 \sum_{\xi\neq 0} \widehat{\chi_A}(\xi)\,\widehat{\chi_B}(-2\xi)\,\widehat{\chi_B}(\xi) \Big| \geqslant \frac{\alpha^2\beta N^2}{5} - \frac{\alpha^2\beta N^2}{10} \geqslant \frac{\alpha^3 N^2}{50}.$$

Thus, for values of N greater than $\frac{50}{\alpha^3}$ there are more than N arithmetic progressions of length three in A with two last terms in B. Thus at least one of them must be a non-constant progression, but obviously it must be also a proper $\mathbb{Z}$-arithmetic progression of length three. This is against the assumption that A is progression-free.

The non-quasirandom case

After the considerations made above we may safely suppose that $\left\|\widehat{f_A}\right\|_{\ell^\infty} \geqslant \frac{\alpha^2}{20}$, provided that $N > \frac{50}{\alpha^3}$. We do not need $\mathbb{Z}_N$ anymore. The assumption is used in a different form: we know that for some $\vartheta \in \mathbb{R}$ we have

$$\left|\sum_{x\in\mathscr{J}_N} f_A(x)\, e(x\vartheta)\right| \geqslant \frac{\alpha^2 N}{20},$$

and from this assumption we want to deduce that for some arithmetic progression $Q \subseteq \mathscr{J}_N$ with length at least $\sqrt[3]{N}$, we have

$$\sum_{x\in Q} f_A(x) \geqslant c\alpha^2 \#Q,$$

where $c \in]0,1[$ will be an absolute constant, for then

$$\#(A\cap Q) = \sum_{x\in Q}\chi_A(x) = \sum_{x\in Q} f_A(x) + \alpha\#Q \geqslant \left(\alpha + c\alpha^2\right)\#Q,$$

and this is exactly the density increment we aim for.

We want to partition $\mathscr{J}_N$ into disjoint arithmetic progressions

$$Q_1, Q_2, \ldots, Q_k \quad \left(k \in \mathbb{Z}_+\right),$$

each having a cardinality at least $\sqrt[3]{N}$, and for which

$$\left|e(x\vartheta) - e(y\vartheta)\right| \leqslant \frac{\alpha^2}{40}$$

for any x and y taken from one of the progressions. Suppose that we have such a partition. Choose an element x_ℓ from Q_ℓ for each $\ell \in \mathscr{J}_k$. Then

$$\begin{aligned}
&\sum_{\ell=1}^{k}\frac{\alpha^2\#Q_\ell}{40} = \frac{\alpha^2 N}{20} - \frac{\alpha^2 N}{40} \leqslant \left|\sum_{x\in\mathscr{J}_N} f_A(x)\, e(x\vartheta)\right| - \frac{\alpha^2 N}{40}\\
&\leqslant \sum_{\ell=1}^{k}\left|\sum_{x\in Q_\ell} f_A(x)\, e(x_\ell\vartheta)\right| + \sum_{\ell=1}^{k}\left|\sum_{x\in Q_\ell} f_A(x)\left(e(x\vartheta) - e(x_\ell\vartheta)\right)\right| - \frac{\alpha^2 N}{40}\\
&\leqslant \sum_{\ell=1}^{k}\left|\sum_{x\in Q_\ell} f_A(x)\right| + \sum_{\ell=1}^{k}\frac{\alpha^2}{40}\#Q_\ell - \frac{\alpha^2 N}{40}\\
&= \sum_{\ell=1}^{k}\left(\left|\sum_{x\in Q_\ell} f_A(x)\right| + \sum_{x\in Q_\ell} f_A(x)\right),
\end{aligned}$$

and thus

$$\sum_{x\in Q_\ell} f_A(x) \geqslant \frac{\alpha^2\#Q_\ell}{80}$$

for some $\ell \in \mathscr{J}_k$.

It suffices to show that the described partition is feasible. For that we need the lemma of Dirichlet which we have already proved (see p. 20). Applying Dirichlet's lemma to the numbers $\delta = \frac{\alpha^2}{800\sqrt[3]{N}}$ and ϑ yields a number $d \in \mathscr{J}_{\left\lfloor \frac{800\sqrt[3]{N}}{\alpha^2} \right\rfloor}$ such that $\left\|d\vartheta\right\| \leqslant \frac{\alpha^2}{800\sqrt[3]{N}}$. Let P be any arithmetic progression of integers with common difference d and length at most $2\sqrt[3]{N}$. Then for any two elements x and y of P we have

$$\begin{aligned}\left|e(x\vartheta) - e(y\vartheta)\right| &\leqslant 2\sqrt[3]{N}\left|e(d\vartheta) - 1\right| = 2\sqrt[3]{N}\cdot 2\left|\sin \pi d\vartheta\right| \leqslant 4\pi\sqrt[3]{N}\left\|d\vartheta\right\| \\ &\leqslant 4\pi\sqrt[3]{N}\frac{\alpha^2}{800\sqrt[3]{N}} = \frac{\pi\alpha^2}{200} \leqslant \frac{5\alpha^2}{200} = \frac{\alpha^2}{40}.\end{aligned}$$

Clearly $d \leqslant \sqrt{N}$ when $N > 800^6\alpha^{-12}$. We shall assume that so is the case for the rest of the section.

We now show that $\mathscr{J}_N$ can be partitioned into arithmetic progressions of common difference d and lengths between $\sqrt[3]{N}$ and $2\sqrt[3]{N}$. The intersections of $\mathscr{J}_N$ with residue classes modulo d yield a partition of $\mathscr{J}_N$ into arithmetic progressions of common difference d and each having a length equal to $\left\lfloor \frac{N}{d} \right\rfloor$ or to $\left\lfloor \frac{N}{d} \right\rfloor + 1$. It suffices to further partition each of these sequences into subsequences formed by consecutive elements of the original sequences and each subsequence having a length between $\sqrt[3]{N}$ and $2\sqrt[3]{N}$.

Let P be one the intersections of $\mathscr{J}_N$ with residue classes modulo d. Let Q_1 be the set of the $\left\lceil \sqrt[3]{N} \right\rceil$ smallest elements of P, Q_2 the set of the $\left\lceil \sqrt[3]{N} \right\rceil$ next smallest elements of P, and so on. This yields a finite sequence of arithmetic progressions $Q_1, Q_2, \ldots, Q_t \subseteq \mathscr{J}_N$ ($t \in \mathbb{Z}_+$), each having common difference d and length $\left\lceil \sqrt[3]{N} \right\rceil$. Write R for the set of remaining elements of P.

If $\#Q_t + \#R \leqslant 2\sqrt[3]{N}$, we have obtained a sought-for partition of P into the sets

$$Q_1, Q_2, \ldots, Q_{t-1}, Q_t \cup R.$$

If $\#Q_t + \#R > 2\sqrt[3]{N}$, then clearly

$$\#Q_t + \#R - 1 \leqslant \left\lceil \sqrt[3]{N} \right\rceil + \left\lceil \sqrt[3]{N} \right\rceil - 1 - 1 \leqslant 2\sqrt[3]{N},$$

so that assuming that $t \geqslant 2$ (which certainly holds if $N > 3$), we can take the smallest element x of Q_t and the sought-after partition takes the form

$$Q_1, Q_2, \ldots, Q_{t-2}, Q_{t-1} \cup \{x\}, (Q_t \setminus \{x\}) \cup R.$$

This finishes the density increment proof of Roth's theorem.

Stronger quantitative forms of Roth's theorem

We close this chapter with some references on the improvements that have been obtained for r_3 after Roth's upper bound.

In [Sz4] Szemerédi gives a proof for the following estimate

$$r_3(N) \ll \frac{N}{e^{c\sqrt{\log\log N}}}, \qquad (N \longrightarrow \infty)$$

where $c \in \mathbb{R}_+$ is a fixed constant. He found the proof in the 1980s but did not publish it back then. The proof is written in the same language of the circle method as Roth's original proof.

The 1980s saw another more drastic improvement. D. R. Heath-Brown's paper [H-B] from 1987 and Szemerédi's paper [Sz3] from 1990 both prove that

$$r_3(N) \ll \frac{N}{(\log N)^c}, \qquad (N \longrightarrow \infty)$$

for some small constant $c \in \mathbb{R}_+$. Szemerédi's paper obtains the explicit value $\frac{1}{20}$ for the constant c. The proof differs from the usual density increment proof in that it considers several Fourier coefficients at once in order to gain a larger density increment per iteration. Green has written an exposition [Gre3] on the Heath-Brown–Szemerédi approach and the section 10.4 of [T&V] also discusses it briefly.

The best upper bound for r_3 obtained so far is due to J. Bourgain. In 1999 appeared his article [Bo1] in which he proves that

$$r_3(N) \ll N\sqrt{\frac{\log\log N}{\log N}}. \qquad (N \longrightarrow \infty)$$

Bourgain's approach is based on the observation that using arithmetic progressions for the iteration is rather inefficient. He uses **Bohr sets** instead. Bourgain's proof is also discussed in depth in his recent article [Bo2], which comes with the appendix [San] due to T. Sanders. In this article Bourgain obtains the best bound for r_3 to date. More precisely, he proves that

$$r_3(N) \ll \frac{N\,(\log\log N)^2}{\sqrt[3]{(\log N)^2}}, \qquad (N \longrightarrow \infty)$$

Bourgain's proof is also discussed in Green's exposition [Gre4] and the above mentioned section 10.4 of the book [T&V].

Gowers uniformity norms

In the density increment proof of Roth's theorem, we had two cases depending on whether $\|\widehat{f_A}\|_{\ell^\infty}$ was smaller or larger than $\frac{\alpha^2}{20}$. The former case corresponded to a set somewhat uniformly distributed over $\mathbb{Z}_N$ and the latter case corresponded to a set with linear bias. Gowers' proof of Szemerédi's theorem depends crucially on a well chosen generalization of this simple notion of "uniformity" or "quasirandomness".

In this chapter we will introduce the Gowers uniformity norms, which will turn out to form an appropriate notion of quasirandomness. After giving the definition we will prove the important Cauchy–Schwarz–Bunyakovsky–Gowers inequality from which we will derive the nestedness property of the Gowers norms and the fact that Gowers norms are genuine norms on $\mathbb{C}^{\mathbb{Z}_N}$. It is then easy to consider fixed-dimensional cubes and show how Gowers uniform sets resemble random sets. We will finish our treatment by proving the generalized von Neumann theorem.

Everything we do in this chapter can be found in Gowers' original article [Go3]. We follow [Go3] quite closely except that we use different notation and our treatment of the generalized von Neumann theorem is essentially as in the chapter 11 of [T&V]. The section 4 of [Go3] discusses reasons for why the concept of Gowers uniformity does not immediately lead to a proof of Szemerédi's theorem. We will not discuss these issues here.

The definition of Gowers uniformity norms

Before we can begin our program, we need to take care of some notational issues. For the purposes of this chapter we fix some positive integer N. In order to cope with the numerous indices that come with the uniformity norms, we need to define some simple notational devices. We use the standard notation

$$\left|\varepsilon\right| \stackrel{\text{def}}{=} \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$$

for multi-indices $\varepsilon = \langle \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \rangle \in \{0,1\}^n$, where $n \in \mathbb{Z}_+$ is arbitrary. Another useful device is to write

$$\mathscr{C}^n z \stackrel{\text{def}}{=} \begin{cases} z, & \text{when } 2 \mid n, \text{ and} \\ \overline{z}, & \text{if } 2 \nmid n. \end{cases}$$

Here $z \in \mathbb{C}$ and $n \in \mathbb{Z}_+$ are again arbitrary.

We may now define Gowers' uniformity norms which made their first appearance in [Go3].

Definition. *Let $d \geqslant 2$ be an integer and let $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$. Then the* ***Gowers uniformity norm*** *of f of* ***order*** *d is defined to be*

$$\|f\|_{\mathscr{U}^d} \stackrel{\text{def}}{=} \sqrt[2^d]{\mathop{\mathscr{E}}_{x}\mathop{\mathscr{E}}_{h_1}\mathop{\mathscr{E}}_{h_2}\cdots\mathop{\mathscr{E}}_{h_d}\prod_{\varepsilon\in\{0,1\}^d}\mathscr{C}^{|\varepsilon|}f(x+\varepsilon\cdot h)},$$

where h denotes simply the vector $\langle h_1, h_2, \ldots, h_d\rangle$. It is also customary to write

$$\|f\|_{\mathscr{U}^1} \stackrel{\text{def}}{=} \sqrt{\mathop{\mathscr{E}}_{x}\mathop{\mathscr{E}}_{h_1} f(x)\,\overline{f(x+h_1)}} = \left|\mathop{\mathscr{E}}_{x} f(x)\right|,$$

though, for obvious reasons, the quantity $\|\cdot\|_{\mathscr{U}^1}$ is not a true norm.

It is easy to see that the Gowers uniformity norms of a function $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ could be defined recursively by the formulae

$$\begin{cases} \|f\|_{\mathscr{U}^1} \stackrel{\text{def}}{=} \left|\mathop{\mathscr{E}}_{x} f(x)\right|, \quad \text{and} \\ \|f\|_{\mathscr{U}^{d+1}} \stackrel{\text{def}}{=} \sqrt[2^{d+1}]{\mathop{\mathscr{E}}_{h}\|\Delta_h f\|_{\mathscr{U}^d}^{2^d}}, \quad \text{which is to hold for any } d \in \mathbb{Z}_+. \end{cases}$$

Here

$$\Delta_h f \stackrel{\text{def}}{=} x \longmapsto f(x)\,\overline{f(x+h)}\colon \mathbb{Z}_N \longrightarrow \mathbb{C}.$$

Incidentally, this observation proves that the radicand in the definition of $\|f\|_{\mathscr{U}^d}$ is always a non-negative real number.

As an example, we mention that in the special case $d = 2$ the Gowers norm takes the form:

$$\|f\|_{\mathscr{U}^2} = \sqrt[4]{\mathop{\mathscr{E}}_{x}\mathop{\mathscr{E}}_{h_1}\mathop{\mathscr{E}}_{h_2} f(x)\,\overline{f(x+h_1)\,f(x+h_2)}f(x+h_1+h_2)}$$

In fact, the Gowers norm $\|f\|_{\mathscr{U}^2}$ has the useful alternative form

$$\begin{aligned} \|f\|_{\mathscr{U}^2} &= \sqrt[4]{\mathop{\mathscr{E}}_{x}\mathop{\mathscr{E}}_{h_1}\mathop{\mathscr{E}}_{h_2} f(x)\,\overline{f(x+h_1)\,f(x+h_2)}f(x+h_1+h_2)} \\ &= \sqrt[4]{\mathop{\mathscr{E}}_{h}\left|\mathop{\mathscr{E}}_{x} f(x)\,\overline{f(x+h)}\right|^2} = \sqrt[4]{\mathop{\mathscr{E}}_{h}\left|\big(f * \overline{f(-\cdot)}\big)(-h)\right|^2} \\ &= \sqrt[4]{\sum_{\xi}\left|\widehat{f * \overline{f(-\cdot)}}(\xi)\right|^2} = \sqrt[4]{\sum_{\xi}\left|\widehat{f}(\xi)\,\widehat{\overline{f(-\cdot)}}(\xi)\right|^2} \\ &= \sqrt[4]{\sum_{\xi}|\widehat{f}\,|^4} = \|\widehat{f}\,\|_{\ell^4}. \end{aligned}$$

It is customary to say that a function $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ having a vanishing expectation is **Gowers δ-uniform** ($\delta \in \mathbb{R}_+$) **of degree** $d \in \mathbb{Z}_+$ if $\|f\|_{\mathscr{U}^{d+1}} \leqslant \delta$. We say that a subset A of $\mathbb{Z}_N$ of density $\alpha \in]0,1[$ is **δ-uniform of degree** d if its balanced function $f_A = \chi_A - \alpha\chi_{\mathbb{Z}_N}$ is δ-uniform of degree d.

Basic properties of Gowers uniformity norms

The statement of the useful Cauchy–Schwarz–Bunyakovsky–Gowers inequality, our first challenge, is easier with the following additional concept.

Definition. *Let $d \geqslant 2$ be an integer and let $\langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d}$ be a family of functions from $\mathbb{Z}_N$ to $\mathbb{C}$. Then the* ***Gowers inner product of degree*** *d of the family $\langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d}$ is defined by the formula*

$$\Big\langle \langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d} \stackrel{\text{def}}{=} \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{h_1} \mathop{\mathscr{E}}_{h_2} \cdots \mathop{\mathscr{E}}_{h_d} \prod_{\varepsilon \in \{0,1\}^d} \mathscr{C}^{|\varepsilon|} f_\varepsilon(x + \varepsilon \cdot h)\,,$$

where again h denotes $\langle h_1, h_2, \dots, h_d \rangle$.

Obviously $\big\|f\big\|_{\mathscr{U}^d} = \sqrt[2^d]{\big\langle \langle \underbrace{f, f, \dots, f}_{2^d} \rangle \big\rangle_{\mathscr{U}^d}}$ for any function $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$.

The first property we shall prove about Gowers norms is the

Cauchy–Schwarz–Bunyakovsky–Gowers inequality. *Let $d \geqslant 2$ be an integer and let $\langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d}$ be an arbitrary family of complex valued functions defined on $\mathbb{Z}_N$. Then*

$$\Big| \Big\langle \langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d} \Big| \leqslant \prod_{\varepsilon \in \{0,1\}^d} \big\|f_\varepsilon\big\|_{\mathscr{U}^d}.$$

Essentially the proof just repeatedly uses the classical Cauchy–Schwarz–Bunyakovsky inequality to exchange dependence on indices into products. For example, in the case $d = 2$ we aim for the following sequence of inequalities:

$$\begin{aligned} \Big| \big\langle \langle f_{00}, f_{10}, f_{01}, f_{11} \rangle \big\rangle_{\mathscr{U}^2} \Big| &\leqslant \sqrt{\Big| \big\langle \langle f_{00}, f_{10}, f_{00}, f_{10} \rangle \big\rangle_{\mathscr{U}^2} \Big|} \sqrt{\Big| \big\langle \langle f_{01}, f_{11}, f_{01}, f_{11} \rangle \big\rangle_{\mathscr{U}^2} \Big|} \\ &\leqslant \big\|f_{00}\big\|_{\mathscr{U}^2} \cdot \big\|f_{10}\big\|_{\mathscr{U}^2} \cdot \big\|f_{01}\big\|_{\mathscr{U}^2} \cdot \big\|f_{11}\big\|_{\mathscr{U}^2}. \end{aligned}$$

The point is that the Cauchy–Schwarz–Bunyakovsky–Gowers inequality follows trivially from the following

Lemma. *Let $d \geqslant 2$ be an integer, let $\langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d}$ be functions $\mathbb{Z}_N \longrightarrow \mathbb{C}$ and let $i \in \mathscr{J}_d$. Then*

$$\begin{aligned} \Big| \Big\langle \langle f_\varepsilon \rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d} \Big| \leqslant \sqrt{\Big| \Big\langle \big\langle f_{\langle \varepsilon_1, \dots, \varepsilon_{i-1}, 0, \varepsilon_{i+1}, \dots, \varepsilon_d \rangle} \big\rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d} \Big|} \\ \cdot \sqrt{\Big| \Big\langle \big\langle f_{\langle \varepsilon_1, \dots, \varepsilon_{i-1}, 1, \varepsilon_{i+1}, \dots, \varepsilon_d \rangle} \big\rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d} \Big|} \end{aligned}$$

where, as usual, we write $\varepsilon = \langle \varepsilon_1, \varepsilon_2, \dots, \varepsilon_d \rangle$.

Proof. As promised, this involves one application of the classical Cauchy–Schwarz–Bunyakovsky inequality and some rearranging. In order to trim the following formulas, we write $\varepsilon|_i\alpha$ for

$$\langle \varepsilon_1, \ldots, \varepsilon_{i-1}, \alpha, \varepsilon_{i+1}, \ldots, \varepsilon_d \rangle .$$

Now

$$\begin{aligned}
&\left| \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{h_1} \mathop{\mathscr{E}}_{h_2} \cdots \mathop{\mathscr{E}}_{h_d} \prod_{\varepsilon\in\{0,1\}^d} \mathscr{C}^{|\varepsilon|} f_\varepsilon(x+\varepsilon\cdot h) \right| \\
&= \left| \mathop{\mathscr{E}}_{h_1} \cdots \widehat{\mathop{\mathscr{E}}_{h_i}} \cdots \mathop{\mathscr{E}}_{h_d} \left(\mathop{\mathscr{E}}_{x} \prod_{\substack{\varepsilon\in\{0,1\}^d \\ \varepsilon_i=0}} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 0}\big(x+\varepsilon_1 h_1+\ldots+\widehat{\varepsilon_i h_i}+\ldots+\varepsilon_d h_d\big) \right) \right. \\
&\qquad \left. \cdot \left(\mathop{\mathscr{E}}_{x} \prod_{\substack{\varepsilon\in\{0,1\}^d \\ \varepsilon_i=1}} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 1}\big(x+\varepsilon_1 h_1+\ldots+\widehat{\varepsilon_i h_i}+\ldots+\varepsilon_d h_d\big) \right) \right| \\
&\leqslant \sqrt{ \mathop{\mathscr{E}}_{h_1} \cdots \widehat{\mathop{\mathscr{E}}_{h_i}} \cdots \mathop{\mathscr{E}}_{h_d} \left| \mathop{\mathscr{E}}_{x} \prod_{\substack{\varepsilon\in\{0,1\}^d \\ \varepsilon_i=0}} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 0}\big(x+\varepsilon_1 h_1+\ldots+\widehat{\varepsilon_i h_i}+\ldots+\varepsilon_d h_d\big) \right|^2 } \\
&\quad \cdot \sqrt{ \mathop{\mathscr{E}}_{h_1} \cdots \widehat{\mathop{\mathscr{E}}_{h_i}} \cdots \mathop{\mathscr{E}}_{h_d} \left| \mathop{\mathscr{E}}_{x} \prod_{\substack{\varepsilon\in\{0,1\}^d \\ \varepsilon_i=1}} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 1}\big(x+\varepsilon_1 h_1+\ldots+\widehat{\varepsilon_i h_i}+\ldots+\varepsilon_d h_d\big) \right|^2 } \\
&= \sqrt{ \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{h_1} \mathop{\mathscr{E}}_{h_2} \cdots \mathop{\mathscr{E}}_{h_d} \prod_{\varepsilon\in\{0,1\}^d} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 0}(x+\varepsilon\cdot h) } \\
&\qquad \cdot \sqrt{ \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{h_1} \mathop{\mathscr{E}}_{h_2} \cdots \mathop{\mathscr{E}}_{h_d} \prod_{\varepsilon\in\{0,1\}^d} \mathscr{C}^{|\varepsilon|} f_{\varepsilon|_i 1}(x+\varepsilon\cdot h) },
\end{aligned}$$

where the large circumflexes signify omission. q.e.d.

The following monotonicity property of the Gowers norms follows as a corollary.

Nestedness of Gowers norms. *For any function $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ we have*

$$\|f\|_{\mathscr{U}^2} \leqslant \|f\|_{\mathscr{U}^3} \leqslant \|f\|_{\mathscr{U}^4} \leqslant \cdots$$

To see this, fix some integer $d \geqslant 2$, and let $f\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ be arbitrary. Define a family $\langle f_\varepsilon \rangle_{\varepsilon\in\{0,1\}^{d+1}}$ of functions from $\mathbb{Z}_N$ to $\mathbb{C}$ as follows:

$$f_\varepsilon \stackrel{\text{def}}{=} \begin{cases} f, & \text{if } \varepsilon_{d+1} = 0, \text{ and} \\ \chi_{\mathbb{Z}_N}, & \text{if } \varepsilon_{d+1} = 1. \end{cases}$$

Then

$$\begin{aligned}
\|f\|_{\mathscr{U}^d} = \sqrt[2^d]{\big\langle \langle f \rangle_{\varepsilon\in\{0,1\}^d} \big\rangle_{\mathscr{U}^d}} &= \sqrt[2^d]{\big\langle \langle f_\varepsilon \rangle_{\varepsilon\in\{0,1\}^{d+1}} \big\rangle_{\mathscr{U}^{d+1}}} \\
&\leqslant \sqrt[2^d]{\|f\|_{\mathscr{U}^{d+1}}^{2^d} \cdot \|\chi_{\mathbb{Z}_N}\|_{\mathscr{U}^{d+1}}^{2^d}} = \|f\|_{\mathscr{U}^{d+1}}.
\end{aligned}$$

From the Cauchy–Schwarz–Bunyakovsky–Gowers inequality, it is also easy to infer the following

Theorem. *For any integer $d \geqslant 2$, the function $\|\cdot\|_{\mathscr{U}^d}$ is a norm on $\mathbb{C}^{\mathbb{Z}_N}$.*

The positive definiteness follows from the coincidence $\|\cdot\|_{\mathscr{U}^2} = \|\widehat{\cdot}\|_{\ell^4}$ and the monotonicity property just proved, and the homogeneity is obvious. It suffices to prove the triangle inequality.

Fix an integer $d \geqslant 2$ and let $f, g\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$ be two arbitrary functions. Writing the power $\|f+g\|_{\mathscr{U}^d}^{2^d}$ in the form

$$\Big\langle \langle f+g \rangle_{\varepsilon \in \{0,1\}^d} \Big\rangle_{\mathscr{U}^d}$$

and using the 2^d-linearity of the Gowers inner product of degree d, we get a sum with 2^{2^d} terms, each being the Gowers inner product of functions which are copies of f and g. Using the Cauchy–Schwarz–Bunyakovsky–Gowers inequality to each of these terms separately yields

$$\|f+g\|_{\mathscr{U}^d}^{2^d} \leqslant \sum_{\ell=0}^{2^d} \binom{2^d}{\ell} \|f\|_{\mathscr{U}^d}^{\ell} \cdot \|g\|_{\mathscr{U}^d}^{2^d-\ell} = \Big(\|f\|_{\mathscr{U}^d} + \|g\|_{\mathscr{U}^d}\Big)^{2^d},$$

which is precisely the triangle inequality. q.e.d.

The combinatorial meaning of Gowers uniformity

Now we discuss the combinatorial meaning of Gowers uniformity of a subset of $\mathbb{Z}_N$ and give some motivation for considering Gowers uniformity as a form of quasirandomness. For that purpose, we fix a subset A of $\mathbb{Z}_N$ of cardinality αN, where, as usual, $\alpha \in]0,1]$. Before proceeding, we must introduce yet another concept. The Gowers norms are clearly some kind of averages over affine cubes, but the definition of affine cubes given on p. 5 is not quite adequate for reasons related to the fact that we defined arithmetic progressions as sets but often take averages over tuples that consist of the elements of arithmetic progressions. In the same vein, we define for each positive integer d the **ordered d-dimensional affine cubes of** $\mathbb{Z}_N$ as functions from $\{0,1\}^d$ to $\mathbb{Z}_N$ of the form

$$\varepsilon = \langle \varepsilon_\ell \rangle_{\ell=1}^d \longmapsto x_0 + \sum_{\ell=1}^d \varepsilon_\ell x_\ell,$$

with $x_0, x_1, \ldots, x_d \in \mathbb{Z}_N$. We say that A **contains** the ordered affine cube φ, if A contains the image of φ. We use the qualifier "ordered" to distinguish these cubes from the ones defined on p. 5 but this terminology is non-standard.

It is obvious that A contains exactly αN zero-dimensional ordered affine cubes and exactly $\alpha^2 N^2$ one-dimensional ordered affine cubes. In general, we may prove that for each $d \in \mathbb{Z}_+$,

A contains at least $\alpha^{2^d} N^{d+1}$ ordered affine cubes of dimension d.

The proof goes smoothly through induction on d.

Suppose that A necessarily contains at least $\alpha^{2^{d-1}}N^d$ ordered affine cubes of dimension $d-1$ for some $d \in \mathbb{Z}_+$. For notational simplicity, write $\square_\partial(U)$ for the number of ∂-dimensional affine cubes in an arbitrary subset $U \subseteq \mathbb{Z}_N$, for every $\partial \in \mathbb{Z}_+ \cup \{0\}$. Of course, $\square_\partial(U)$ is simply equal to $N^{\partial+1}\|\chi_U\|_{\mathscr{U}^\partial}^{2^\partial}$. Then, by induction hypothesis and the power-mean inequality,

$$\begin{aligned}
\square_d(A) &= \sum_x \square_{d-1}\big(A \cap (A+x)\big) \geqslant \sum_x N^d \left(\frac{\#\big(A \cap (A+x)\big)}{N}\right)^{2^{d-1}} \\
&= N^{d+1} \mathop{\mathscr{E}}_x \left(\frac{\#\big(A \cap (A+x)\big)}{N}\right)^{2^{d-1}} \\
&\geqslant N^{d+1} \left(\mathop{\mathscr{E}}_x \frac{\#\big(A \cap (A+x)\big)}{N}\right)^{2^{d-1}} \\
&= N^{d+1} \left(\mathop{\mathscr{E}}_x \chi_A(x) \mathop{\mathscr{E}}_y \chi_A(y)\right)^{2^{d-1}} \\
&= N^{d+1} \left(\alpha^2\right)^{2^{d-1}} = N^{d+1}\alpha^{2^d}. \quad \text{q.e.d.}
\end{aligned}$$

For really large N, the different "corners" of a typical ordered d-dimensional affine cube φ of $\mathbb{Z}_N$, that is, the elements of the image of φ, are likely to be pairwise distinct. We also see that such a cube φ is determined by the N^{d+1} possible choices of the corresponding $x_0, x_1, \ldots, x_d \in \mathbb{Z}_N$. Therefore, for a set $A \subseteq \mathbb{Z}_N$, constructed by admitting each element of $\mathbb{Z}_N$ in A with a probability α, we would expect A to contain around $\alpha^{2^d}N^{d+1}$ d-dimensional ordered affine cubes. Now we see that a Gowers uniform set of degree $d-1$ resembles a random set in the sense that

> *If A is Gowers δ-uniform of degree $d-1$, where $d \in \mathbb{Z}_+$, then A contains at most $(\alpha+\delta)^{2^d}N^{d+1}$ d-dimensional ordered affine cubes.*

This follows directly from the triangle inequality of Gowers norms: Let f_A be the balanced function $\chi_A - \alpha\chi_{\mathbb{Z}_N}$ of A. Then, using the symbol $\square_d$ defined above, we can estimate:

$$\begin{aligned}
\square_d(A) &= N^{d+1}\|\chi_A\|_{\mathscr{U}^d}^{2^d} = N^{d+1}\|\alpha\chi_{\mathbb{Z}_N} + f_A\|_{\mathscr{U}^d}^{2^d} \\
&\leqslant N^{d+1}\Big(\|\alpha\chi_{\mathbb{Z}_N}\|_{\mathscr{U}^d} + \|f_A\|_{\mathscr{U}^d}\Big)^{2^d} \leqslant N^{d+1}\,(\alpha+\delta)^{2^d}. \quad \text{q.e.d.}
\end{aligned}$$

The generalized von Neumann theorem

As we hinted on p. 19, we will deal with generalizations of expressions of the form $\mathscr{E}_x\mathscr{E}_d f(x)\,g(x+d)\,h(x+2d)$, where $f, g, h\colon \mathbb{Z}_N \longrightarrow \mathbb{C}$. This fact calls for a

Definition. *Let $k \in \mathbb{Z}_+$ and let f_1, f_2, …, f_k be functions from $\mathbb{Z}_N$ to $\mathbb{C}$. We define the multilinear form $\Lambda_k\colon \underbrace{\mathbb{C}^{\mathbb{Z}_N} \times \mathbb{C}^{\mathbb{Z}_N} \times \cdots \times \mathbb{C}^{\mathbb{Z}_N}}_{k} \longrightarrow \mathbb{C}$ by the formula*

$$\Lambda_k\big(f_1, f_2, \ldots, f_k\big) \stackrel{\text{def}}{=} \mathop{\mathscr{E}}_x \mathop{\mathscr{E}}_d f_1(x)\, f_2(x+d)\, f_3(x+2d) \cdots f_k\big(x+(k-1)d\big)\,.$$

It turns out that for each integer $k \geqslant 2$, the Gowers uniformity norm $\|\cdot\|_{\mathscr{U}^{k-1}}$ controls the k-linear form Λ_k. The rest of the section is dedicated to the proof of this remarkable fact.

The generalized von Neumann theorem. *Let $k \geqslant 2$ be an integer, let a_1, a_2, …, a_k be integers such that their pairwise differences are coprime to N, and let f_1, f_2, …, f_k be arbitrary functions defined on $\mathbb{Z}_N$, each taking values in the closed unit disc $\mathbb{D}$ of the complex plane. Then*

$$\left| \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{d} f_1(x + a_1 d)\, f_2(x + a_2 d) \cdots f_k(x + a_k d) \right| \leqslant \|f_1\|_{\mathscr{U}^{k-1}}.$$

We shall use induction on k. First, assume that $k = 2$. Then for any two functions f and g from $\mathbb{Z}_N$ to $\mathbb{C}$ bounded by one in modulus, and for any constants $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$ for which $(a - b, N) = 1$, we have

$$\left| \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{d} f(x + ad)\, g(x + bd) \right| = \left| \left(\mathop{\mathscr{E}}_{x} f(x) \right) \left(\mathop{\mathscr{E}}_{x} g(x) \right) \right|,$$

It is obvious that the last product is at most $\left|\mathscr{E} f\right|$, or equivalently, it is at most $\|f\|_{\mathscr{U}^1}$.

Next, let $k \geqslant 2$ be such an integer that the generalized von Neumann theorem is known to hold for the k-linear functional, let a_1, a_2, …, a_{k+1} be integers such that their pairwise differences are coprime to the number N, and let f_1, f_2, …, f_{k+1} be some $\mathbb{D}$-valued functions defined on $\mathbb{Z}_N$.

First we need to make the change of variables $y = x + a_{k+1} d$. Then the left-hand side expression takes the form

$$\mathop{\mathscr{E}}_{y} \mathop{\mathscr{E}}_{d} f_1(y + b_1 d)\, f_2(y + b_2 d) \cdots f_k(y + b_k d)\, f_{k+1}(y)\,,$$

where $b_1 = a_1 - a_{k+1}$, $b_2 = a_2 - a_{k+1}$, …, and $b_k = a_k - a_{k+1}$. Clearly the pairwise differences of the numbers b_1, b_2, …, b_k are just pairwise differences formed from the numbers a_1, a_2, …, a_k and a_{k+1}, and hence also coprime to N. Now we may use the triangle inequality, the fact that $\left|f_{k+1}\right| \leqslant 1$, the classical Cauchy–Schwarz–Bunyakovsky inequality, the induction hypothesis and the power-mean inequality to get

$$\begin{aligned}
&\left| \mathop{\mathscr{E}}_{x} \mathop{\mathscr{E}}_{d} f_1(x + a_1 d)\, f_2(x + a_2 d) \cdots f_{k+1}\,(x + a_{k+1} d) \right| \\
&= \left| \mathop{\mathscr{E}}_{y} \mathop{\mathscr{E}}_{d} f_1(y + b_1 d)\, f_2(y + b_2 d) \cdots f_k(y + b_k d)\, f_{k+1}(y) \right| \\
&\leqslant \mathop{\mathscr{E}}_{y} \left| \mathop{\mathscr{E}}_{d} f_1(y + b_1 d)\, f_2(y + b_2 d) \cdots f_k(y + b_k d) \right| \\
&\leqslant \sqrt{\mathop{\mathscr{E}}_{y} \left| \mathop{\mathscr{E}}_{d} f_1(y + b_1 d)\, f_2(y + b_2 d) \cdots f_k(y + b_k d) \right|^2}
\end{aligned}$$

$$= \sqrt{\mathop{\mathscr{E}}_{y}\mathop{\mathscr{E}}_{d}\mathop{\mathscr{E}}_{e} f_1(y+b_1d)\,\overline{f_1(y+b_1e)}\cdots f_k(y+b_kd)\,\overline{f_k(y+b_ke)}}$$

$$= \sqrt{\mathop{\mathscr{E}}_{e}\mathop{\mathscr{E}}_{y}\mathop{\mathscr{E}}_{d} \bigl(\Delta_{b_1(e-d)}f_1\bigr)(y+b_1d)\cdots\bigl(\Delta_{b_k(e-d)}f_k\bigr)(y+b_kd)}$$

$$= \sqrt{\mathop{\mathscr{E}}_{\delta}\mathop{\mathscr{E}}_{y}\mathop{\mathscr{E}}_{d} \bigl(\Delta_{b_1\delta}f_1\bigr)(y+b_1d)\,\bigl(\Delta_{b_2\delta}f_2\bigr)(y+b_2d)\cdots\bigl(\Delta_{b_k\delta}f_k\bigr)(y+b_kd)}$$

$$\leqslant \sqrt{\mathop{\mathscr{E}}_{\delta} \bigl\|\Delta_{(a_1-a_{k+1})\delta}f_1\bigr\|_{\mathscr{U}^{k-1}}} = \sqrt{\mathop{\mathscr{E}}_{\delta} \bigl\|\Delta_{\delta}f_1\bigr\|_{\mathscr{U}^{k-1}}} \leqslant \sqrt[2^k]{\mathop{\mathscr{E}}_{\delta} \bigl\|\Delta_{\delta}f_1\bigr\|_{\mathscr{U}^{k-1}}^{2^{k-1}}}.$$

And according to the observation we made on page 40, the last expression is equal to the Gowers uniformity norm $\bigl\|f_1\bigr\|_{\mathscr{U}^k}$. Q.E.D.

As an immediate corollary we get the following special case which is the form in which we will actually use the result.

The generalized von Neumann theorem. *Let $k \geqslant 2$ be an integer, let f_1, f_2, …, f_k be arbitrary functions from $\mathbb{Z}_N$ to $\mathbb{D}$, and suppose that the numbers N and $(k-1)!$ are coprime. Then*

$$\bigl|\Lambda_k\bigl(f_1,f_2,\dots,f_k\bigr)\bigr| \leqslant \min\Bigl\{\bigl\|f_1\bigr\|_{\mathscr{U}^{k-1}}, \bigl\|f_2\bigr\|_{\mathscr{U}^{k-1}},\dots,\bigl\|f_k\bigr\|_{\mathscr{U}^{k-1}}\Bigr\}.$$

Gowers' Approach to Szemerédi's Theorem

In this chapter we will discuss the proof of the following theorem, the ancestry and relatives of which were already discussed in the first chapter.

Gowers' theorem. [Go3] *For all integers* $k \geqslant 4$,

$$r_k(N) \ll_k \frac{N}{(\log\log N)^{c_k}}, \qquad (N \longrightarrow \infty)$$

where $c_k \in \mathbb{R}_+$ *is a number dependent only on* k.

Our primary sources are the original article [Go3] and the lecture notes [Gre8], though the articles [Go2] and [Gre7], the lectures notes [Go4] and [Sou], and the chapter 11 of the book [T&V] have also been very useful. Since the whole proof is extremely complicated (the original article [Go3] has 124 pages), we have to restrict ourselves to the easiest special case $k = 4$. The original proof gives the explicit values $c_k = 2^{-2^{k+9}}$, but even in our special case $k = 4$, we will settle for the mere existence of c_4.

We begin by stating an analogue of the density increment strategy used in the proof of Roth's theorem (p. 32), slightly modified to suit better the case of longer progressions, and by showing that it implies Gowers' bounds. Having defined and discussed the appropriate notion of quasirandomness, that of Gowers uniformity, in the previous chapter, we then present the easy quasirandom part of the proof. That is, we will show that lack of four-term arithmetic progressions implies quadratic non-uniformity.

Then comes the difficult non-quasirandom part of the proof. After some technical preliminary steps, we introduce quickly and without proof the Freĭman–Ruzsa theorem and the Balog–Szemerédi–Gowers theorem and apply them to show that a little piece of the derivative of the balanced function of the set in question has to be linear. Then we turn this linearity into correlation with quadratic phase functions. Finally, we need to introduce, again without proof, Weyl's inequality and use it to transform the correlation with quadratics into the sought-for density increment.

We close the chapter by mentioning some recent results on r_4 and by shortly discussing the case of longer progressions.

The density increment strategy

As we have mentioned before, Gowers' proof of Szemerédi's theorem essentially generalizes the Fourier analytic density increment proof of Roth's theorem. In the proof of Roth's theorem it was sufficient to approach the situation differently depending on whether the characteristic function of the set of interest had a large Fourier coefficient or not. This question was enough to distinguish quasirandom sets from structured ones. In the case of longer progressions this is, however, insufficient and consequently Gowers' proof is significantly more complicated. The concept of a set having no large Fourier coefficients is replaced by the more technical concept of Gowers uniformity which we already have discussed in detail.

The primary purpose of the present section is to establish a density increment strategy to Gowers' theorem, one very similar to the one we used in the proof of Roth's theorem (see p. 32). We will state the density increment result we aim for and show that it really implies Gowers' bound for the r_4.

The stepping stone to Gowers' bounds will be the fact that

Lack of progressions implies density increment. *Let $\alpha \in]0,1]$ and suppose that N is an integer greater than $C\alpha^{-C}$. Furthermore, let A be a subset of $\mathscr{J}_N$ of cardinality αN. Then A contains an arithmetic progression of length four or $\mathscr{J}_N$ contains a proper arithmetic subprogression P of length at least $e^{-\alpha^{-C}}N^{\alpha^C}$ in which the set A has density greater than $\alpha + c\alpha^C$, i.e. in which*

$$\frac{\#\big(A \cap P\big)}{\#P} > \alpha + c\alpha^C.$$

Here each C signifies a real constant greater than one and c denotes a positive real constant smaller than one, both dependent only on k.

We make two remarks. The first one is that obviously the set $\mathscr{J}_N$ could be replaced by any proper $\mathbb{Z}$-arithmetic progression of length N in the above statement. The second remark is that the exact values of the constants are difficult to manage and hence we will avoid them altogether. The notation becomes cleaner when it is observed that all the constants C may be taken to be equal.

The derivation of Gowers' bound via density increment

Let α, N, A and C be as in the statement of the above proposition and suppose that A does not contain a proper arithmetic progression of length four. Then we may iterate the density increment argument as many times as possible to get a sequence

$$\mathscr{J}_N \supseteq Q_1 \supseteq Q_2 \supseteq \dots$$

of proper arithmetic progressions with lengths satisfying the inequalities

$$\begin{cases} \#Q_1 \geqslant e^{-\alpha^{-C}} N^{\alpha^C}, \\ \#Q_2 \geqslant e^{-\alpha_1^{-C}} (\#Q_1)^{\alpha_1^C} \geqslant e^{-\alpha^{-C}} (\#Q_1)^{\alpha^C}, \\ \#Q_3 \geqslant e^{-\alpha^{-C}} (\#Q_2)^{\alpha^C}, \\ \dots, \end{cases}$$

respectively, and in which the set A has densities

$$\begin{cases} \alpha_1 > \alpha + c\alpha^C, \\ \alpha_2 > \alpha_1 + c\alpha_1^C > \alpha + 2c\alpha^C, \\ \alpha_3 > \alpha + 3c\alpha^C, \\ \dots, \end{cases}$$

again respectively.

Now of course, the density of A doubles in at most $\frac{\alpha^{1-C}}{c}$ steps and consequently reaches one in at most

$$\frac{1}{c\alpha^{C-1}} + \frac{1}{2^{C-1}c\alpha^{C-1}} + \frac{1}{4^{C-1}c\alpha^{C-1}} + \ldots = \frac{1}{1-2^{1-C}} \cdot \frac{1}{c\alpha^{C-1}}$$

steps. We conclude that the sequence $\mathscr{J}_N \supseteq Q_1 \supseteq Q_2 \supseteq \ldots$ must be finite, the last element being, say, Q_m, where $m \in \mathbb{Z}_+$. Here $m \ll \alpha^{1-C}$. The only assumption of the density increment argument that may fail after m steps is the lower bound for the length of $\#Q_m$. That is, we must have $\#Q_m \ll \alpha^{-C}$. Iterating the lower bounds for $\#Q_1$, $\#Q_2$, ... we get

$$e^{-m\alpha^{-C}} N^{(\alpha^C)^m} \leqslant \underbrace{e^{-\alpha^{-C}\left(e^{-\alpha^{-C}}\left(\cdots\left(e^{-\alpha^{-C}} N^{\alpha^C}\right)^{\alpha^C}\cdots\right)^{\alpha^C}\right)^{\alpha^C}}}_{m \text{ exponentiations to power } \alpha^C} \leqslant \#Q_m \ll \alpha^{-C},$$

from which we infer

$$\alpha^{-C} > \log \alpha^{-C} \gg -m\alpha^{-C} + \left(\alpha^C\right)^m \log N \gg -\alpha^{1-2C} + \left(\alpha^C\right)^m \log N.$$

In other words, $\alpha^{1-2C} \gg \left(\alpha^C\right)^m \log N$. We may continue to get

$$\alpha^{1-2C} > \log \alpha^{1-2C} \gg m\alpha^C + \log\log N,$$

which means that $$\alpha \ll \frac{1}{\sqrt[2^{C-1}]{\log\log N}}.$$ Q.E.D.

From now on, let A, N and α be as in the statement of the above density increment strategy except that A is now supposed to contain no non-trivial four-term arithmetic progressions. Before continuing, we need to handle some technicalities.

In the various steps of the proof, we will need to do things that require N to be sufficiently large. In each case, the requirement will be of the form $N > C\alpha^{-C}$, as in the statement of the density increment strategy. In each of these cases, we will tacitly assume that N satisfies the relevant inequality. This should not cause any real confusion.

Following Green's presentation in [Gre8], we observe that we may assume that $\alpha \leqslant \frac{4}{5}$ provided that $N > 20$, for if $N > 20$ and $\alpha > \frac{4}{5}$ then A obviously must contain four consecutive integers. This can be seen e.g. by partitioning $\mathscr{J}_N$ into sets, each consisting of four or five consecutive integers. Then A must have density greater than $\frac{4}{5}$ in at least one of these. It is easy to convince oneself of the fact that this does not break the above deduction of Gowers' bounds from the density increment strategy. The reasons for why we make this assumption are notational: Given any absolute constants $C, C' \in \mathbb{R}_+$, we

can find an absolute constant $C'' \in \mathbb{R}_+$ such that $C\alpha^{-C'} \leqslant \alpha^{-C''}$ and we can thereby let the exponent of α swallow any leading constants that may occur. For instance, this allows us to write $e^{-\alpha^{-C}} N^{\alpha^C}$ instead of $e^{-C\alpha^{-C}} N^{c\alpha^C}$.

A more profound technicality is that we need to embed $\mathscr{J}_N$ into a cyclic group. However, this time $\mathbb{Z}_N$ is not good enough since we will need many linear changes of variables which simply can not be made in $\mathbb{Z}_N$ unless N is a prime number. Another reason for why the primality of N would be useful is that then all arithmetic progressions of $\mathbb{Z}_N$ of length at most N would be either trivial or have pairwise distinct elements. In order to do the embedding we need to invoke the following truly classical result of elementary number theory.

Bertrand's postulate. *Let n be an integer greater than one. Then there is at least one prime number in the interval $]n, 2n[$.*

This was first conjectured by J. Bertrand who verified the first few million special cases and it was first proved by P. Chebyshev in 1850 [A&Z, ch. 2]. For a proof, see for instance [Chand, §3 of ch. VII] or [A&Z, ch. 2].

Bertrand's postulate allows us to choose a prime number p from the interval $]2N, 4N[$. As one might expect, we will embed $\mathscr{J}_N$ into $\mathbb{Z}_p$ via the canonical surjection $n \longmapsto n + p\mathbb{Z}\colon \mathbb{Z} \longrightarrow \mathbb{Z}_p$ and furthermore, to simplify notation, we will identify $\mathscr{J}_N$ and $\mathscr{J}_p$ with their images under the canonical surjection. This should not cause any real confusion since we will spend the rest of the proof in $\mathbb{Z}_p$ and only in the very end we will revert back to $\mathbb{Z}$. From now on, all sums and expectations are over $\mathbb{Z}_p$ unless otherwise stated.

A remark is in order. Even though $p \asymp N$, the density of A has dropped significantly as its habitat has grown. However, we will see that we can circumvent this problem by considering the balanced function $f_A = \chi_A - \alpha\chi_{\mathscr{J}_N}$.

The quasirandom case

Having already done all the necessary preparations, the quasirandom part of the proof is rather simple. To be precise, our goal in this section is to prove that the fact that the set A has no proper four-term progressions implies that its balanced function has a large Gowers $\mathscr{U}^3$-norm, provided that N is sufficiently large.

We begin by considering the functional $\Lambda_4\big(\alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}\big)$. It is equal to

$$
\begin{aligned}
&\frac{\alpha^4}{p^2} \#\Big\{ \langle x, d\rangle \in \mathbb{Z}_p^2 \Big| x, x+d, x+2d, x+3d \in \mathscr{J}_N \Big\} \\
\geqslant &\frac{\alpha^4}{p^2} \#\Big\{ \langle x, d\rangle \in \mathscr{J}_p^2 \Big| x, x+d, x+2d, x+3d \in \mathscr{J}_N \Big\} \\
\geqslant &\frac{\alpha^4}{p^2} \left(\left\lfloor \frac{N-1}{3} \right\rfloor + \left\lfloor \frac{N-2}{3} \right\rfloor + \ldots + 1 \right) \geqslant \frac{\alpha^4}{p^2} \left(\frac{N-4}{3} + \frac{N-5}{3} + \ldots + \frac{1}{3} \right) \\
= &\frac{\alpha^4}{p^2} \cdot \frac{(N-3)(N-4)}{6} = \frac{\alpha^4}{p^2} \cdot \frac{N^2 - 7N + 12}{6} \\
> &\alpha^4 \left(\frac{1}{96} - \frac{7}{96N} + \frac{1}{8N^2} \right) > \frac{\alpha^4}{192},
\end{aligned}
$$

provided that $\frac{1}{192} > \frac{7}{96N}$, or equivalently that $N > 14$.

On the other hand, the functional $\Lambda_4(\chi_A, \chi_A, \chi_A, \chi_A)$ is equal to $\frac{1}{p^2}$ times the number of four-term $\mathbb{Z}_p$-arithmetic progressions $\langle x, x+d, x+2d, x+3d\rangle$ of A. Since all such progressions are also $\mathbb{Z}$-arithmetic progressions, we have

$$0 < \Lambda_4(\chi_A, \chi_A, \chi_A, \chi_A) = \frac{\#A}{p^2} < \frac{\alpha}{4N} < \frac{\alpha^4}{384},$$

provided that $N > 96\alpha^{-3}$.

Using the generalized von Neumann theorem, the triangle inequality, the multilinearity of Λ_4 and the above observations, we may bound the quantity $15\|f_A\|_{\mathscr{U}^3}$ from below by

$$\begin{aligned}
&\big|\Lambda_4(f_A, f_A, f_A, f_A)\big| + \big|\Lambda_4(f_A, f_A, f_A, \alpha\chi_{\mathscr{J}_N})\big| + \big|\Lambda_4(f_A, f_A, \alpha\chi_{\mathscr{J}_N}, f_A)\big| \\
&+ \big|\Lambda_4(f_A, \alpha\chi_{\mathscr{J}_N}, f_A, f_A)\big| + \ldots + \big|\Lambda_4(\alpha\chi_{\mathscr{J}_N}, f_A, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N})\big| \\
&+ \big|\Lambda_4(\alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, f_A, \alpha\chi_{\mathscr{J}_N})\big| + \big|\Lambda_4(\alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, f_A)\big| \\
&\qquad \geqslant \big|\Lambda_4(\chi_A, \chi_A, \chi_A, \chi_A) - \Lambda_4(\alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N}, \alpha\chi_{\mathscr{J}_N})\big| \\
&\qquad > \frac{\alpha^4}{192} - \frac{\alpha^4}{384} = \frac{\alpha^4}{384}.
\end{aligned}$$

That is, we have $\|f_A\|_{\mathscr{U}^3} \gg \alpha^4$.

The non-quasirandom case, part I

We have shown that the balanced function of A is non-uniform in the sense that $\|f_A\|_{\mathscr{U}^3} \gg \alpha^4$, provided that N satisfies an inequality of the form $N > C\alpha^{-C}$, where $C > 1$ is an absolute real constant. Now we have to turn this non-uniformity into a density increment by finding a long $\mathbb{Z}$-arithmetic progression of $\mathscr{J}_N$ in which A has a density $\alpha + \Omega(\alpha^C)$ for some absolute constant $C \in]1, \infty[$. The proof has two main parts, the first of which is described in the present section.

The first part is to prove that f_A correlates locally with **quadratic phase factors** which are functions of the form $e\big(\frac{-\psi(\cdot)}{p}\big)$, where ψ is a quadratic polynomial of $\mathbb{Z}_p$. This task breaks into several substeps. We begin by considering large Fourier-coefficients of the functions $\Delta_h f_A$ ($h \in \mathbb{Z}_p$). We recall that the symbol Δ_h appearing here was defined in page 40. These functions are often called **derivatives** of f_A because they resemble directed partial derivatives. We will see that $\Delta_h f_A$ has a large Fourier-coefficient for many different values of h. More precisely, we show that there exists a relatively large subset S of $\mathbb{Z}_p$ and a function $\varphi\colon S \longrightarrow \mathbb{Z}_p$ such that $\big|\widehat{\Delta_h f_A}(\varphi(h))\big|$ is large for every $h \in S$.

Then we will see that this function φ is not quite arbitrary, but has many **additive quadruples**, which are solutions $\langle a, b, c, d\rangle$ of the equations

$$a + b = c + d \quad \text{and} \quad \varphi(a) + \varphi(b) = \varphi(c) + \varphi(d)$$

with $a, b, c, d \in S$. Then comes the result that φ is almost a linear function in a long $\mathbb{Z}_p$-arithmetic progression P. This observation depends on the Balog–Szemerédi–Gowers theorem and the Freĭman–Ruzsa theorem. Finally a technical but elementary argument is needed to turn this "almost linearity" of φ

into a family of quadratic phase factors with which f_A is then seen to correlate locally.

The second part of the non-quasirandom case is concerned with turning this correlation with quadratics into the sought-for density increment and will be discussed in the next section.

Large Fourier-coefficients of the derivatives of f_A

We have shown that

$$\sqrt[8]{\mathop{\mathscr{E}}_h \big\|\Delta_h f_A\big\|_{\mathscr{U}^2}^4} = \big\|f_A\big\|_{\mathscr{U}^3} \geqslant c\alpha^4,$$

where $c \in \mathbb{R}_+$ is a small absolute constant, provided that N satisfies an inequality of the form $N > C\alpha^{-C}$ (where $C > 1$ denotes an absolute real constant). Since f_A is everywhere bounded by one in modulus, so are the norms $\big\|\Delta_h f_A\big\|_{\mathscr{U}^2}$ of its derivatives for all $h \in \mathbb{Z}_p$. We must have at least $\frac{c^8\alpha^{32}p}{2}$ values of h with

$$\big\|\Delta_h f_A\big\|_{\mathscr{U}^2} \geqslant \frac{c^2\alpha^8}{\sqrt[4]{2}},$$

for otherwise

$$c\alpha^4 \leqslant \sqrt[8]{\mathop{\mathscr{E}}_h \big\|\Delta_h f_A\big\|_{\mathscr{U}^2}^4} < \sqrt[8]{\frac{c^8\alpha^{32}}{2}\cdot 1^4 + \Big(1 - \frac{c^8\alpha^{32}}{2}\Big)\Big(\frac{c^2\alpha^8}{\sqrt[4]{2}}\Big)^4} < c\alpha^4.$$

For these values of h, we have

$$\begin{aligned}\big\|\widehat{\Delta_h f_A}\big\|_{\ell^\infty} &\geqslant \big\|\widehat{\Delta_h f_A}\big\|_{\ell^\infty}\cdot\big\|\Delta_h f_A\big\|_{\mathscr{L}^2} = \big\|\widehat{\Delta_h f_A}\big\|_{\ell^\infty}\cdot\big\|\widehat{\Delta_h f_A}\big\|_{\ell^2}\\ &\geqslant \big\|\widehat{\Delta_h f_A}\big\|_{\ell^4}^2 = \big\|\Delta_h f_A\big\|_{\mathscr{U}^2}^2 \geqslant \frac{c^4\alpha^{16}}{\sqrt{2}}.\end{aligned}$$

Therefore these exists a subset S of $\mathbb{Z}_p$ having cardinality at least $\frac{c^8\alpha^{32}p}{2}$ and a function $\varphi\colon S \longrightarrow \mathbb{Z}_N$ such that for each $h \in S$, we have

$$\Big|\widehat{\Delta_h f_A}\big(\varphi(h)\big)\Big| \geqslant \frac{c^4\alpha^{16}}{\sqrt{2}}.$$

The additive structure of the large Fourier-coefficients

In the next section we want to apply the Balog–Szemerédi–Gowers theorem to the graph of the function φ and for this purpose we need to verify that it really does satisfy the necessary conditions, namely that it has enough additive quadruples.

Additive structure of the large Fourier-coefficients. *There are at least* $\frac{c^{64}\alpha^{256}p^3}{256}$ *quadruples* $\langle a,b,c,d\rangle \in S^4$ *such that*

$$a + b = c + d \quad \textit{and} \quad \varphi(a) + \varphi(b) = \varphi(c) + \varphi(d)\,,$$

where c *is as in the previous subsection.*

Proof. The number M of such quadruples is of course

$$\frac{1}{p}\sum_{a\in S}\sum_{b\in S}\sum_{c\in S}\sum_{d\in S}\mathop{\mathscr{E}}_{x} e\Big(\frac{x\,(\varphi(a)+\varphi(b)-\varphi(c)-\varphi(d))}{p}\Big)\sum_{\xi} e\Big(\frac{-(a+b-c-d)\,\xi}{p}\Big)\,.$$

Some rearranging and a single application of the Cauchy–Schwarz–Bunyakovsky inequality gives

$$\begin{aligned}\frac{M}{p^3} &= \mathop{\mathscr{E}}_{x}\sum_{\xi}\Big|\mathop{\mathscr{E}}_{h}\chi_S(h)\,e\Big(\frac{x\varphi(h)-h\xi}{p}\Big)\Big|^4\\ &= \mathop{\mathscr{E}}_{x}\sum_{\xi}\Big|\widehat{\chi_S(\cdot)\,E_x(\cdot)}(\xi)\Big|^4\\ &\geqslant \Big(\mathop{\mathscr{E}}_{x}\sqrt{\sum_{\xi}\Big|\widehat{\chi_S(\cdot)\,E_x(\cdot)}(\xi)\Big|^4}\Big)^2,\end{aligned}$$

where E_x denotes the function defined by the expression $e\big(\frac{x\varphi(\cdot)}{p}\big)$. The derivative $\Delta_x f_A$ is bounded by one in modulus for any $x\in\mathbb{Z}_p$ and thus

$$\sum_{\xi}\Big|\widehat{\Delta_x f_A}\Big|^4 = \Big\|\big(\widehat{\Delta_x f_A}\big)^2\Big\|_{\ell^2}^2 = \Big\|\Delta_x f_A * \Delta_x f_A\Big\|_{\mathscr{L}^2}^2 \leqslant 1.$$

We infer that

$$\frac{M}{p^3} \geqslant \Big(\mathop{\mathscr{E}}_{x}\sqrt{\sum_{\xi}\Big|\widehat{\chi_S(\cdot)\,E_x(\cdot)}(-\xi)\Big|^4}\sqrt{\sum_{\xi}\Big|\widehat{\Delta_x f_A}(\xi)\Big|^4}\Big)^2$$

Using the Cauchy–Schwarz–inequality once more and then using the basic properties of the discrete Fourier transform gives

$$\begin{aligned}\frac{M}{p^3} &\geqslant \Big(\mathop{\mathscr{E}}_{x}\sum_{\xi}\Big|\widehat{\chi_S(\cdot)\,E_x(\cdot)}(-\xi)\Big|^2\Big|\widehat{\Delta_x f_A}(\xi)\Big|^2\Big)^2\\ &= \Big(\mathop{\mathscr{E}}_{x}\Big\|\big(\chi_S(-\cdot)\,E_x(-\cdot)*\Delta_x f_A\big)^{\wedge}\Big\|_{\ell^2}^2\Big)^2\\ &= \Big(\mathop{\mathscr{E}}_{x}\Big\|\chi_S(-\cdot)\,E_x(-\cdot)*\Delta_x f_A\Big\|_{\mathscr{L}^2}^2\Big)^2\end{aligned}$$

Continuing in a similar fashion, using the Cauchy–Schwarz–Bunyakovsky inequality and the fact that f_A is real-valued and bounded by one in modulus, we

get

$$\begin{aligned}
\frac{M}{p^3} &\geqslant \left(\mathop{\mathscr{E}}_x\mathop{\mathscr{E}}_y\left|\mathop{\mathscr{E}}_h \chi_S(h)\, f_A(y+h)\,\overline{f_A(x+y+h)}\, e\Big(\frac{x\varphi(h)}{p}\Big)\right|^2\right)^2 \\
&\geqslant \left(\mathop{\mathscr{E}}_x\mathop{\mathscr{E}}_y\left|\mathop{\mathscr{E}}_h \chi_S(h)\, f_A(y+h)\,\overline{f_A(x+y+h)}\, e\Big(\frac{x\varphi(h)}{p}\Big)\right|\right)^4 \\
&\geqslant \left(\mathop{\mathscr{E}}_h\mathop{\mathscr{E}}_x\mathop{\mathscr{E}}_y \chi_S(h)\, f_A(y)\,\overline{f_A(y+h)\, f_A(x+y)} f_A(x+y+h)\right. \\
&\qquad\qquad \left.\cdot\, e\Big(\frac{-y\varphi(h)}{p}\Big)\, e\Big(\frac{(x+y)\,\varphi(h)}{p}\Big)\right)^4 \\
&= \left(\mathop{\mathscr{E}}_h \chi_S(h)\left|\widehat{\Delta_h f_A}\big(\varphi(h)\big)\right|^2\right)^4.
\end{aligned}$$

The claim now follows from the simple observation that:

$$\sum_{h\in S}\left|\widehat{\Delta_h f_A}\big(\varphi(h)\big)\right|^2 \geqslant \frac{c^8\alpha^{32}p}{2}\cdot\Big(\frac{c^4\alpha^{16}}{\sqrt{2}}\Big)^2 = \frac{c^{16}\alpha^{64}p}{4}. \qquad \text{q.e.d.}$$

The Balog–Szemerédi–Gowers theorem and the Freĭman–Ruzsa theorem

Let Γ be the graph of φ, that is, let

$$\Gamma = \Big\{\big\langle h, \varphi(h)\big\rangle\Big| h\in S\Big\} \subseteq \mathbb{Z}_p^2.$$

We know that Γ has $\Omega\big(\alpha^C p^3\big) \gg \alpha^C\,(\#\Gamma)^3$ additive quadruples, that is quadruples $\langle x,y,z,w\rangle \in \Gamma^4$ with $x+y=z+w$. We also know that $\#\Gamma = \#S \gg \alpha^C p$. Here $C>1$ is some absolute real constant. Our goal is to find a sufficiently long $\mathbb{Z}_p$-arithmetic progression Q such that φ coincides with a linear function on a large subset of Q. For this we use some well-known results of additive combinatorics and some simple averaging arguments.

In 1994 A. Balog and Szemerédi published the article [Bal&S] in which they prove a theorem which, from our point of view, basically says that if a non-empty finite subset of an Abelian group has many additive quadruples, then it has a large subset X such that $\frac{\#(X+X)}{\#X}$ is small. Soon after this little variations of this result were considered in different applications, resulting in a number of different precise formulations of the same phenomena. Gowers' proof requires a result of this type but the original proof of Balog and Szemerédi invoked Szemerédi's regularity lemma which has a tendency of producing weak quantitative bounds [Go1]. However, he found a way to replace the application of Szemerédi's regularity lemma by a simple averaging argument, presented in [Go2] and [Go3]. This allowed significantly better quantitative bounds.

The Balog–Szemerédi–Gowers theorem is discussed for instance in sections 2.5 and 6.4 of the book [T&V]. For a more thorough coverage with quantitatively strong versions of the theorem and numerous references, Balog's article [Bal] may be consulted.

We will apply the theorem in the following form, the proof of which, together with explicit expressions for the constants, may be found in M.-C. Chang's article [Chang2].

The Balog–Szemerédi–Gowers theorem. *Let X be a non-empty and finite subset of some ambient Abelian group. Suppose that X has more than $K\,(\#X)^3$ additive quadruples. Then the set X has a subset Y with*

$$\#Y > K^c \#X, \quad \textit{and} \quad \#(Y+Y) < CK^{-C}\#Y,$$

where $c \in]0,1[$ and $C > 1$ are absolute real constants.

We immediately conclude that the set Γ has a subset Γ' for which

$$\#\Gamma' \gg \alpha^C \#\Gamma \quad \text{and} \quad \#(\Gamma'+\Gamma') \ll \alpha^{-C}\#\Gamma',$$

where each C is another absolute real constant greater than one.

In order to be able to apply the Freĭman–Ruzsa theorem, we need to change our context from that of $\mathbb{Z}_p^2$ to that of $\mathbb{Z}^2$. For this purpose we define the maps

$$\varrho = x \longmapsto x \cap \mathscr{J}_p \colon \mathbb{Z}_p \longrightarrow \mathbb{Z}$$

and

$$\varrho \times \varrho = \langle x, y\rangle \longmapsto \langle \varrho(x)\,, \varrho(y)\rangle : \mathbb{Z}_p^2 \longrightarrow \mathbb{Z}^2,$$

and the set $\Gamma'' = (\varrho \times \varrho)[\Gamma']$, for which $\#\Gamma'' = \#\Gamma' \gg \alpha^C \#\Gamma$ and

$$\#(\Gamma''+\Gamma'') \ll \alpha^{-C}\#\Gamma'',$$

with implicit constant four times as large as in the corresponding estimate for $\#(\Gamma'+\Gamma')$.

We also need to introduce a new concept. Let G be an Abelian group. A **generalized G-arithmetic progression** is a sum of ordinary G-arithmetic progressions. If P_1, P_2, …, P_d ($d \in \mathbb{Z}_+$) are some arithmetic progressions in G having lengths ℓ_1, ℓ_2, …, ℓ_d, respectively, then the generalized arithmetic progression $P = P_1 + P_2 + \ldots + P_d$ is said to have **dimension** equal to d if it can not be represented as the sum of fewer than d ordinary arithmetic progressions. We write $\dim P = d$. The generalized arithmetic progression P is said to be **proper** if its cardinality is $\#P_1 \#P_2 \cdots \#P_d$. It is a useful observation that generalized G-arithmetic progressions could be equivalently defined as sets of the form

$$\left\{ x + k_1 d_1 + k_2 d_2 + \ldots + k_n d_n \,\middle|\, k_1 \in \mathscr{J}_{\ell_1}, k_2 \in \mathscr{J}_{\ell_2}, \ldots, k_n \in \mathscr{J}_{\ell_n} \right\},$$

with $x \in G$, $d_1, d_2, \ldots, d_n \in G$, $\ell_1, \ell_2, \ldots, \ell_n \in \mathbb{Z}_+$ and $n \in \mathbb{Z}_+$.

If P is a generalized arithmetic progression of dimension $d \in \mathbb{Z}_+$, then obviously $\#(P+P) \leqslant 2^d \#P$. In 1964 G. A. Freĭman proved a famous theorem which says that if Y is a non-empty and finite set of integers satisfying $\#(Y+Y) \leqslant K\#Y$, where $K > 0$ is some absolute constant, then Y is contained in a generalized arithmetic progression $P \subseteq \mathbb{Z}$ which satisfies the estimates

$$\dim P \ll_K 1, \quad \text{and} \quad \#P \ll_K \#Y.$$

The original proof [Fr1, Fr2, Fr3] has a reputation of being quite difficult, and it does not yield very strong bounds for the implicit constants. A cleaner simpler presentation of the proof can be found in Yu. Bilu's article [Bi]. In the early 1990s I. Z. Ruzsa [Ru1, Ru2, Ru3, Ru4] found a different and much stronger proof, which gave polynomial bounds for the implicit constants and allowed the generalized arithmetic progression to be assumed to be proper. This proof gained a wider audience with the book [Nat] of M. Nathanson. The bounds were further improved by Chang [Chang1]. Green and Ruzsa have generalized the result to arbitrary Abelian groups [Gr&R].

Recall that an Abelian group is called **torsion-free** if its only element having a finite order is its neutral element. Even though the Freĭman–Ruzsa theorem was originally stated only for $\mathbb{Z}$, and we are presently living inside $\mathbb{Z}^2$, this does not pose any problems, as the proof goes through without any major changes for any ambient torsion-free Abelian group. We could apply the Freĭman–Ruzsa theorem in the following general and yet quantitatively strong form, which is proved for instance in [Chang1] and [Gre1] in the integer case and in [T&V, ch. 5] in the general case.

The Freĭman–Ruzsa theorem for torsion-free groups. *Let Y be a non-empty finite subset of some ambient torsion-free Abelian group G. Suppose that $\#(Y+Y) \leqslant K\#Y$, where K is a positive real number. Then the set Y is contained in a proper generalized G-arithmetic progression P of dimension at most K and cardinality at most $e^{CK^C}\#Y$, where $C > 1$ is an absolute positive real constant.*

Since $\mathbb{Z}^2$ is a torsion-free Abelian group we immediately conclude that there is a proper generalized $\mathbb{Z}^2$-arithmetic progression P such that

$$\Gamma'' \subset P, \quad \dim P \leqslant C\alpha^{-C} \quad \text{and} \quad \#P \leqslant e^{C\alpha^{-C}}\#\Gamma'',$$

where $C > 1$ is an absolute real constant. However, following this line of thinking to the very end would only yield the inferior upper bound $r_4(N) \ll \frac{N}{(\log\log\log N)^c}$, the reason being that in the end of a typical proof of Freĭman–Ruzsa theorem a lot of structure is obtained but discarded immediately in our application of it. On the bright side, Ruzsa's approach obtains a lemma which says that when $\#(Y+Y) \leqslant K\#Y$, then $2Y-2Y$ contains a generalized arithmetic progression of characteristics similar to those in the conclusion of Freĭman–Ruzsa theorem. Here and also in the rest of this subsection the expression $2Y-2Y$, and other similar expressions, signifies the sumset $Y+Y-Y-Y$, or analogously other such sumsets, respectively. The precise formulation of the result we shall use is as follows. This is presented in the book [T&V] as theorem 5.30.

The Ruzsa–Chang lemma. *Let Y be a non-empty and finite subset of some ambient torsion-free Abelian group G. Suppose that $\#(Y+Y) \leqslant K\#Y$ for some real number $K > 1$. Then the sumset $2Y-2Y$ contains a proper generalized arithmetic progression P such that*

$$\dim P \ll K\,(1+\log K) \quad \textit{and} \quad \#P \geqslant e^{-CK(1+\log K)}\#Y,$$

where $C > 1$ is an absolute real constant.

We may use this result to get a stronger conclusion than we would get from a direct application of the Freĭman–Ruzsa theorem, though we also use the special case $k = 3, \ell = 2$ of the following theorem. This is not strictly necessary since an elementary argument would imply a weaker but fully sufficient sumset inequality.

Plünnecke–Ruzsa estimates. *Let Y be a non-empty and finite subset of some ambient Abelian group G. Suppose that $\#(Y+Y) \leqslant K\#Y$ for some real number $K > 1$. Then*

$$\#(kY - \ell Y) \leqslant K^{k+\ell}\#Y$$

for all non-negative integers k and ℓ.

This theorem is also a classic in additive combinatorics. This is given in [Gre1] as theorem 3 and it also follows immediately from the corollary 6.29 of [T&V].

Now let us apply the Ruzsa–Chang lemma to the set $\Gamma'' \subseteq \mathbb{Z}^2$. We immediately conclude that the set $2\Gamma'' - 2\Gamma''$ contains a proper generalized arithmetic progression P of dimension at most $C\alpha^{-C}$ and cardinality at least $e^{-C\alpha^{-C}}\#\Gamma''$. The set Γ'' is certainly contained in

$$\bigcup_{x\in\mathbb{Z}^2}(x+P)\,,$$

and on the other hand

$$\sum_{x\in\mathbb{Z}^2}\#\big(\Gamma''\cap(x+P)\big) = \#\Gamma''\#P,$$

where the number of indices of x, for which the corresponding term in the above sum is non-vanishing, is at most

$$\#(3\Gamma'' - 2\Gamma'') \ll \alpha^{-C}\#\Gamma'',$$

for yet another absolute real constant $C > 1$, since such elements x are obviously contained in the sumset $3\Gamma'' - 2\Gamma''$. Therefore we must have a translate of P, say P', such that $\#(\Gamma''\cap P') \gg \alpha^C\#P'$.

Let P_1, P_2, …, P_n, where $n \in \mathbb{Z}_+$ and $n \leqslant C\alpha^{-C}$, be some $\mathbb{Z}^2$-arithmetic progressions of lengths ℓ_1, ℓ_2,…, ℓ_n, respectively, such that $P' = P_1 + P_2 + \ldots + P_n$. Choose the longest of these ordinary arithmetic progressions. Without loss of generality we may assume that it is P_1. Then P' can be partitioned into $\ell_2\ell_3\cdots\ell_n$ translates of P_1. Therefore

$$\#P_1 \geqslant \sqrt[n]{\#P'} \geqslant \sqrt[n]{e^{-C\alpha^{-C}}\#\Gamma''} \gg e^{-cC\alpha^{-C}}\big(\alpha^C p\big)^{c\alpha^C} \geqslant e^{-C\alpha^{-C}}p^{c\alpha^C}.$$

Since $\#(\Gamma''\cap P') \gg \alpha^C\#P'$, we must have at least one translate of P_1, say P'', for which

$$\#(\Gamma''\cap P'') \gg \alpha^C\#P''.$$

This is greater than one provided that $N > C\alpha^{-C}$ for some suitable absolute real constant $C > 1$, simply because $\#P'' \geqslant (\#\Gamma'')^{c\alpha^C}$, and $\#\Gamma'' \gg \alpha^C N$. Therefore the progression P'' can not be contained in a translate of the set $\{0\}\times\mathbb{Z}$, that is, the progression P'' can not be "vertical" in $\mathbb{Z}^2$. But then P'' is

the graph of a linear function defined on a $\mathbb{Z}$-arithmetic progression P''' of the same cardinality as P''. Reverting to $\mathbb{Z}_p$ via the canonical projection we get a $\mathbb{Z}_p$-arithmetic progression Q and a linear function

$$\widetilde{\varphi} = x \longmapsto \lambda x + \mu \colon \mathbb{Z}_p \longrightarrow \mathbb{Z}_p,$$

where $\lambda, \mu \in \mathbb{Z}_p$, for which

$$\#Q \geqslant e^{-\alpha^{-C}} p^{\alpha^C},$$

for some absolute $C \in]1, \infty[$, where we have applied our notation simplifying assumption $\alpha \leqslant \frac{4}{5}$ (p. 49), and φ coincides with $\widetilde{\varphi}$ for at least $\alpha^C \#Q$ elements of Q. Especially we must have

$$\mathop{\mathscr{E}}_{h \in Q} \left| \widehat{\Delta_h f_A}(\lambda h + \mu) \right| \gg \alpha^C.$$

We remark that we may safely assume that $\#Q \leqslant \sqrt{p}$, a bound which will prove to be useful in the next section. If this is not the case, the upper bound can be achieved e.g. by breaking $\#Q$ into subprogressions having cardinalities $\lfloor \sqrt{\#Q} \rfloor$ and $\lceil \sqrt{\#Q} \rceil$. Then in one of these subprogressions, say in Q', we must have $\mathscr{E}_{h \in Q'} \left| \widehat{\Delta_h f_A}(\lambda h + \mu) \right| \gg \alpha^C$, and clearly $\#Q' \geqslant e^{-\alpha^{-C'}} N^{\alpha^{C'}}$ for some absolute $C' \in]1, \infty[$.

A local inverse theorem for $\left\| \cdot \right\|_{\mathscr{U}^3}$

We have finally reached the stage in which we can obtain the goal of this section, Gowers' impressive local inverse theorem for the $\mathscr{U}^3$-norm. Let Q be as in the conclusion of the previous subsection. Then

There exists a family $\langle \psi_s \rangle_{s \in \mathbb{Z}_p}$ of quadratic polynomials such that

$$\mathop{\mathscr{E}}_{s} \left| \mathop{\mathscr{E}}_{h \in s+Q} f_A(h)\, e\Big(\frac{-\psi_s(h)}{p}\Big) \right| \gg \alpha^C,$$

where, as always, $C > 1$ is an absolute real constant.

Given the earlier considerations of this section, the proof is a bit technical but fully elementary, the most complicated prerequisite being the Cauchy–Schwarz–Bunyakovsky inequality.

We begin by expanding the last conclusion we obtained, which was that

$$\mathop{\mathscr{E}}_{h \in Q} \left| \widehat{\Delta_h f_A}(\lambda h + \mu) \right|^2 \geqslant c\alpha^C,$$

where Q is a $\mathbb{Z}_p$-arithmetic progression of length at least $e^{-C\alpha^{-C}} N^{c\alpha^C}$, λ and μ are some elements of $\mathbb{Z}_p$ and $c < 1$ and $C > 1$ are positive real constants. We

recall that f_A takes only real values. We get

$$\begin{aligned}
c\alpha^C &\leqslant \mathop{\mathscr{E}}_{h\in Q} \left|\widehat{\Delta_h f_A}(\lambda h+\mu)\right|^2 \\
&= \mathop{\mathscr{E}}_{h\in Q}\mathop{\mathscr{E}}_{s} f_A(s)\, f_A(s+h)\, e\Big(\frac{-s\,(\lambda h+\mu)}{p}\Big) \\
&\qquad\qquad \cdot \overline{\mathop{\mathscr{E}}_{t} f_A(t)\, f_A(t+h)\, e\Big(\frac{-t\,(\lambda h+\mu)}{p}\Big)} \\
&= \mathop{\mathscr{E}}_{h\in Q}\mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{t} f_A(s)\, f_A(s+h)\, \overline{f_A(t)\, f_A(t+h)}\, e\Big(\frac{(t-s)(\lambda h+\mu)}{p}\Big) \\
&= \mathop{\mathscr{E}}_{h\in Q}\mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{u} f_A(s)\, f_A(s+h)\, \overline{f_A(s+u)\, f_A(s+u+h)}\, e\Big(\frac{u\,(\lambda h+\mu)}{p}\Big).
\end{aligned}$$

Let the terms of Q be $x+d$, $x+2d$, …, $x+Td$, for some $x, d\in\mathbb{Z}_p$ and $T\in\mathbb{Z}_+$. We get

$$\begin{aligned}
c\alpha^C &\leqslant \mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{u} f_A(s)\, f_A(s+x+id)\, \overline{f_A(s+u)\, f_A(s+u+x+id)} \\
&\qquad\qquad e\Big(\frac{u\,(\lambda x+\lambda id+\mu)}{p}\Big) \\
&= \mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{j\in\mathscr{J}_T}\mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{v} f_A(s)\, f_A(s+x+id)\, \overline{f_A(s+v+jd)} \\
&\qquad\qquad \cdot\overline{f_A(s+v+jd+x+id)}\, e\Big(\frac{(v+jd)(\lambda x+\lambda id+\mu)}{p}\Big) \\
&\leqslant \mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{v}\Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{j\in\mathscr{J}_T} f_A(s+x+id)\, \overline{f_A(s+v+jd)} \\
&\qquad\qquad \cdot\overline{f_A(s+v+jd+x+id)}\, e\Big(\frac{(v+jd)(\lambda x+\lambda id+\mu)}{p}\Big)\Bigg|.
\end{aligned}$$

Write $\gamma_{s,v}$ for the last expression in $\left|\cdot\right|$ together with the bars. Clearly $\mathscr{E}_s\mathscr{E}_v\gamma_{s,v}\geqslant c\alpha^C$.

We temporarily fix $s\in\mathbb{Z}_p$ and $v\in\mathbb{Z}_p$. To simplify notation, we write

$$\begin{cases} f_1(\varsigma) = f_A(s+x+\varsigma d)\,, \quad f_2(\varsigma) = f_A(s+v+\varsigma d)\,, \quad \text{and} \\ f_3(\varsigma) = f_A(s+v+x+\varsigma d)\,, \end{cases}$$

for all $\varsigma\in\mathbb{Z}_p$. We also write

$$a=\lambda dv, \quad b=d\,(\lambda x+\mu)\,, \quad \text{and} \quad c=-\frac{\lambda d^2}{2},$$

where the division is, of course, in $\mathbb{Z}_p$. Then by rewriting the inequalities obtained above, and forgetting the unimodular factor $e\big(\frac{v\lambda x+v\mu}{p}\big)$, we get

$$\begin{aligned}
\gamma_{s,v} &= \Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{j\in\mathscr{J}_T} f_1(i)\, f_2(j)\, f_3(i+j)\, e\Big(\frac{ai+bj-2cij}{p}\Big)\Bigg| \\
&= \Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{j\in\mathscr{J}_T} f_1(i)\, e\Big(\frac{ci^2+ai}{p}\Big)\, f_2(j)\, e\Big(\frac{cj^2+bj}{p}\Big)\, f_3(i+j)\, e\Big(\frac{-c\,(i+j)^2}{p}\Big)\Bigg|.
\end{aligned}$$

The last expression is easily seen to be equal to

$$\begin{aligned}&\Bigg|\mathop{\mathscr{E}}_{\xi}\mathop{\mathscr{E}}_{i\in\mathscr{J}_T}\mathop{\mathscr{E}}_{j\in\mathscr{J}_T}\sum_{k=1}^{2T} f_1(i)\,e\Big(\frac{ci^2+ai}{p}\Big)\,f_2(j)\,e\Big(\frac{cj^2+bj}{p}\Big)\\&\qquad\qquad\cdot f_3(k)\,e\Big(\frac{-ck^2}{p}\Big)\,e\Big(\frac{-\xi\,(i+j-k)}{p}\Big)\Bigg|\\=&\Bigg|\mathop{\mathscr{E}}_{\xi}\mathop{\mathscr{E}}_{i\in\mathscr{J}_T} f_1(i)\,e\Big(\frac{ci^2+ai}{p}\Big)\,e\Big(\frac{-i\xi}{p}\Big)\mathop{\mathscr{E}}_{j\in\mathscr{J}_T} f_2(j)\,e\Big(\frac{cj^2+bj}{p}\Big)\,e\Big(\frac{-j\xi}{p}\Big)\\&\qquad\qquad\cdot\sum_{k=1}^{2T} f_3(k)\,e\Big(\frac{-ck^2}{p}\Big)\,e\Big(\frac{k\xi}{p}\Big)\Bigg|.\end{aligned}$$

By further simplifying notation by defining

$$\begin{cases} g_1(\varsigma)=\chi_{\mathscr{J}_T}(\varsigma)\,f_1(\varsigma)\,e\Big(\dfrac{c\varsigma^2+a\varsigma}{p}\Big),\ \ g_2(\varsigma)=\chi_{\mathscr{J}_T}(\varsigma)\,f_2(\varsigma)\,e\Big(\dfrac{c\varsigma^2+b\varsigma}{p}\Big),\ \text{ and}\\ g_3(\varsigma)=\chi_{\mathscr{J}_{2T}}(\varsigma)\,f_3(\varsigma)\,e\Big(\dfrac{-c\varsigma^2}{p}\Big),\end{cases}$$

for all $\varsigma\in\mathbb{Z}_p$, and by applying the Cauchy–Schwarz–Bunyakovsky inequality, we see that

$$\begin{aligned}\frac{\gamma_{s,v}T^2}{p^2}=\Bigg|\sum_{\xi}\widehat{g_1}(\xi)\,\widehat{g_2}(\xi)\,\widehat{g_3}(-\xi)\Bigg|&\leqslant\big\|\widehat{g_1}\big\|_{\ell^\infty}\cdot\big\|\widehat{g_2}\big\|_{\ell^2}\cdot\big\|\widehat{g_3}\big\|_{\ell^2}\\&=\big\|\widehat{g_1}\big\|_{\ell^\infty}\cdot\big\|g_2\big\|_{\mathscr{L}^2}\cdot\big\|g_3\big\|_{\mathscr{L}^2}\leqslant\big\|\widehat{g_1}\big\|_{\ell^\infty}\sqrt{\frac{T}{p}}\sqrt{\frac{2T}{p}}\\&\leqslant\frac{T\sqrt{2}}{p^2}\max_{\xi\in\mathbb{Z}_p}\Bigg|\sum_{i=1}^{T} f_1(i)\,e\Big(\frac{ci^2+(a-\xi)\,i}{p}\Big)\Bigg|.\end{aligned}$$

Hence

$$\Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T} f_1(i)\,e\Big(\frac{ci^2+(a-\xi)\,i}{p}\Big)\Bigg|\geqslant\frac{\gamma_{s,v}}{\sqrt{2}}$$

for some $\xi\in\mathbb{Z}_p$. We conclude that for each $s\in\mathbb{Z}_p$ and each $v\in\mathbb{Z}_p$, we have a quadratic polynomial $\psi_{s,v}$ such that

$$\Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T} f_A(s+x+id)\,e\Big(\frac{-\psi_{s,v}(i)}{p}\Big)\Bigg|\geqslant\frac{\gamma_{s,v}}{\sqrt{2}}.$$

Summing over v yields

$$\mathop{\mathscr{E}}_{v}\Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T} f_A(s+x+id)\,e\Big(\frac{-\psi_{s,v}(i)}{p}\Big)\Bigg|\geqslant\frac{1}{\sqrt{2}}\mathop{\mathscr{E}}_{v}\gamma_{s,v}.$$

Therefore we may choose v in such a way that

$$\Bigg|\mathop{\mathscr{E}}_{i\in\mathscr{J}_T} f_A(s+x+id)\,e\Big(\frac{-\psi_{s,v}(i)}{p}\Big)\Bigg|\geqslant\frac{1}{\sqrt{2}}\mathop{\mathscr{E}}_{v}\gamma_{s,v}.$$

Finally, a simple linear change of variables, namely $\psi_{s,v}(i) = \widetilde{\psi}_s(s+x+id)$, and a summation over s give the required

$$\mathop{\mathscr{E}}_{s}\left|\mathop{\mathscr{E}}_{h\in s+Q} f_A(h)\, e\Big(\frac{-\widetilde{\psi}_s(h)}{p}\Big)\right| \geqslant \frac{1}{\sqrt{2}}\mathop{\mathscr{E}}_{s}\mathop{\mathscr{E}}_{v}\gamma_{s,v} \geqslant \frac{c\alpha^C}{\sqrt{2}}. \qquad \text{Q.E.D.}$$

We point out that the arguments given in this section add up to give a proof of the following theorem.

Gowers' inverse theorem for $\|\cdot\|_{\mathscr{U}^3}$. *Let N be a sufficiently large odd prime number and let f be a $[-1,1]$-valued function defined on $\mathbb{Z}_N$ such that*

$$\mathop{\mathscr{E}}_{x} f(x) = 0, \quad \textit{and} \quad \big\|f\big\|_{\mathscr{U}^3} \geqslant \delta \in \mathbb{R}_+.$$

Then there exists a proper $\mathbb{Z}_N$-arithmetic progression Q of length at least $e^{-\delta^{-C}}N^{\delta^C}$ and a family $\langle\psi_s\rangle_{s\in\mathbb{Z}_N}$ of quadratic polynomials such that

$$\mathop{\mathscr{E}}_{s}\left|\mathop{\mathscr{E}}_{h\in s+Q} f(h)\, e\Big(\frac{-\psi_s(h)}{N}\Big)\right| \gg \delta^C,$$

where $C \in \,]1,\infty[$ is an absolute constant.

The non-quasirandom case, part II

Having obtained the inverse theorem for the $\mathscr{U}^3$-norm, we now have to turn its conclusion into a density increment. Our situation is quite similar to the one we were in the beginning of the non-quasirandom part of the proof of Roth's theorem (see p. 35 onwards). There the balanced function of the set under consideration correlated globally with a linear phase factor. With the help of Dirichlet's lemma, the ambient set $\mathscr{J}_N$ was dissected into arithmetic progressions in which the phase factor was approximately constant. Then a simple averaging argument yielded the density increment in restriction to one of these smaller arithmetic progressions.

Now we have local correlation with quadratic phase factors. The locality poses no problems since we will work inside one small arithmetic progression anyway. On the other hand, the quadraticity requires much stronger tools than Dirichlet's lemma. Besides Dirichlet's lemma, the rest of the argument will crucially depend on a suitable corollary of the famous inequality of Weyl.

We will begin the last stage of our journey by first changing the translates of Q into a partition of $\mathbb{Z}_p$ into translates of Q and translates of Q extended by one term. We then get rid of the quadratic phase factors by using a proper corollary of Weyl's inequality and Dirichlet's lemma to break each of the $\mathbb{Z}_p$-progressions into much smaller subprogressions in which the phase factors are approximately constant. Then we will refine this partition to get a partition of $\mathbb{Z}_p$ into $\mathbb{Z}$-arithmetic progressions. Finally, we will only need a simple averaging argument to pick a $\mathscr{J}_N$-arithmetic progression in which the density increment is finally reached.

Partitioning $\mathbb{Z}_p$ into $\mathbb{Z}_p$-arithmetic progressions

Out starting point is the fact that we have a $\mathbb{Z}_p$-arithmetic progression Q of length at least $e^{-\alpha^{-C}}N^{\alpha^C}$ and a family $\langle\psi_s\rangle_{s\in\mathbb{Z}_p}$ of quadratic polynomials such that

$$\sum_s \left|\sum_{h\in s+Q} f_A(h)\, e\Big(\frac{-\psi_s(h)}{p}\Big)\right| \geqslant c\alpha^C p\#Q,$$

where $C\in\,]1,\infty[$ is an absolute real constant. Provided that $\#Q\leqslant\sqrt{p}$, which is the case by the remark made on p. 58, we can obviously represent p as a sum in which each term is either $\#Q$ or $1+\#Q$, and furthermore in such a way that the number $\#Q$ is used at least once. First just write a sum of $\left\lfloor\frac{p}{\#Q}\right\rfloor\geqslant\#Q$ times $\#Q$ and then increase at most $\#Q-1$ terms by one. The point of this observation is that we can partition $\mathbb{Z}_p$ into $\mathbb{Z}_p$-arithmetic progressions P_1, P_2, …, P_M ($M\in\mathbb{Z}_+$) with each progression being a translate of Q or a translate of Q with one additional term. Why this can be done becomes obvious after a linear change of variables that maps Q to an arithmetic progression of common difference one. We write P_j' for the arithmetic progression formed by the first $\#Q$ terms of P_j.

Plugging all this new notation to the conclusion of Gowers' inverse theorem, we get

$$\sum_{s\in\mathbb{Z}_p}\sum_{j=1}^{M}\left|\sum_{h\in s+P_j'} f_A(h)\, e\Big(\frac{-\psi_{s,j}(h)}{p}\Big)\right| \geqslant c\alpha^C pM\#Q,$$

where the quadratic polynomials have been naturally reindexed. For some value of the index $s\in\mathbb{Z}_p$, we must have

$$\sum_{j=1}^{M}\left|\sum_{h\in s+P_j'} f_A(h)\, e\Big(\frac{-\psi_{s,j}(h)}{p}\Big)\right| \geqslant c\alpha^C M\#Q.$$

When the translate $s'+Q$ of Q (where $s'\in\mathbb{Z}_p$) is extended by an additional term h_0, so that $Q\cup\{h_0\}$ is still an arithmetic progression, the sum

$$\left|\sum_{h\in s'+Q} f_A(h)\, e\Big(\frac{-\psi_{s'}(h)}{p}\Big)\right|$$

decreases by at most one since the summation changes by one term and each term is bounded by one. Less than $\#Q\leqslant\sqrt{p}$ of the arithmetic progressions $s+P_1$, $s+P_2$, …, $s+P_M$ are extended copies of Q. Therefore

$$\sum_{j=1}^{M}\left|\sum_{h\in s+P_j} f_A(h)\, e\Big(\frac{-\psi_{s,j}(h)}{p}\Big)\right| \geqslant c\alpha^C M\#Q-\sqrt{p}.$$

We have $M\#Q>p-\sqrt{p}$ and thus the last expression is at least

$$c\alpha^C M\#Q-\sqrt{p}\geqslant c\alpha^C p-c\alpha^C\sqrt{p}-\sqrt{p}>\frac{c\alpha^C p}{2},$$

provided that $N\asymp p>16\left(c\alpha^C\right)^{-2}$.

Eliminating the quadratic phase factors

In 1913 H. Weyl published his deep paper [Wey] on uniform distribution. There he introduces an important inequality which has corollaries relevant to our subject. The inequality itself does not admit a fully transparent proof and the precise statement is not used here either. Instead we settle for the following statement of the corollary we will actually use.

A corollary to Weyl's inequality. *For any sufficiently large positive integer n and any $t \in \mathscr{J}_n$ and $a \in \mathscr{J}_n$, there exists a number $d \in \mathscr{J}_{\lfloor \sqrt[4]{t} \rfloor}$ such that*

$$\left\| \frac{ad^2}{n} \right\| \leqslant \frac{\varsigma}{t^{\varkappa}},$$

where $\varkappa \in]0,1[$ is a (small) absolute constant and ς is an absolute positive real constant.

This follows trivially from the theorem 16 of [Go2]. A standard reference for Weyl's inequality is [Vau], in which it is fully proved and used in applications of the circle method. For discussions from the point of view of Gowers' proof, the section 5 of the article [Go3] or any of the lecture notes [Go4, Gre8, Sou] may be consulted.

Let us revise our situation. We have obtained $\mathbb{Z}_p$-arithmetic progressions P_1, P_2, …, P_M and quadratic polynomials ψ_1, ψ_2, …, ψ_M for which

$$\sum_{j=1}^{M} \left| \sum_{h \in P_j} f_A(h)\, e\Big(\frac{-\psi_j(h)}{p}\Big) \right| \geqslant c\alpha^C p,$$

where $c \in]0,1[$ and $C \in]1,\infty[$ are absolute real constants. If we could partition each P_j into subprogressions P_{j1}, P_{j2}, …, P_{jm_j} such that for any $s = 1, 2, \ldots, m_j$,

$$\left| e\Big(\frac{-\psi_j(x)}{p}\Big) - e\Big(\frac{-\psi_j(y)}{p}\Big) \right| \leqslant \frac{c\alpha^C}{2},$$

whenever $x, y \in P_{js}$, then we could reason as in the density increment proof of Roth's theorem. That is, we could choose an element x_{js} from each P_{js} ($j \in \mathscr{J}_M$, $s \in \mathscr{J}_{m_j}$) and use the triangle inequality to get

$$\begin{aligned}
c\alpha^C p &\leqslant \sum_{j=1}^{M} \left| \sum_{x \in P_j} f_A(x)\, e\Big(\frac{-\psi_j(x)}{p}\Big) \right| \\
&\leqslant \sum_{j=1}^{M} \sum_{s=1}^{m_j} \left(\left| \sum_{x \in P_{js}} f_A(x)\, e\Big(\frac{-\psi_j(x_{js})}{p}\Big) \right| \right. \\
&\qquad\qquad \left. + \left| \sum_{x \in P_{js}} f_A(x) \left(e\Big(\frac{-\psi_j(x)}{p}\Big) - e\Big(\frac{-\psi_j(x_{js})}{p}\Big) \right) \right| \right) \\
&\leqslant \sum_{j=1}^{M} \sum_{s=1}^{m_j} \left| \sum_{x \in P_{js}} f_A(x) \right| + \frac{c\alpha^C}{2} \sum_{j=1}^{M} \sum_{s=1}^{m_j} \#P_{js},
\end{aligned}$$

thereby obtaining

$$\sum_{j=1}^{M}\sum_{s=1}^{m_j}\left|\sum_{x\in P_{js}} f_A(x)\right| \geqslant \frac{c\alpha^C p}{2}.$$

Now we diverge slightly from our approach in the density increment proof of Roth's theorem. There we aimed to control the length of the subprogressions from below. This time we aim to control the number of the subprogressions. The reason for this is that in the next phase we will break our newly obtained subprogressions into even smaller sub-$\mathbb{Z}$-progressions, and this procedure could conceivably produce very short subprogressions. Even if some of the subprogressions could be very short, the upper bound for the number of subprogressions gives a lower bound for the **average length** of the subprogressions. In the end, we will see that the final $\mathbb{Z}$-arithmetic progression in which the density increment is obtained can be found because the too short progressions simply do not contribute so much in the sums we will obtain later.

Fix the value of $j \in \mathscr{J}_M$ and write $\psi_j(x) = ax^2 + bx$ with $a, b \in \mathbb{Z}_p$. Obviously we can disregard the constant term of ψ_j if it happens to have one. We will split the $\mathbb{Z}_p$-arithmetic progression P_j into at most $e^{-\alpha^{-C}}N^{1-\alpha^C}$ arithmetic subprogressions such that if $Q = \{x, x+d, x+2d, \ldots, x+(\ell-1)d\}$, where $x, d \in \mathbb{Z}_p$ and $\ell \in \mathbb{Z}_+$, is one of them, then

$$\left|e\Big(\frac{-\psi_j(x)}{p}\Big) - e\Big(\frac{-\psi_j(x+kd)}{p}\Big)\right| \leqslant \frac{c\alpha^C}{4},$$

for any $k \in \mathscr{J}_{\ell-1}$. Here $C > 1$ is yet another absolute constant. A simple linear change of variables allows us to assume, without loss of generality, that the common difference of P_j is equal to one.

We begin by estimating the above difference:

$$\begin{aligned}
&\left|e\Big(\frac{-\psi_j(x)}{p}\Big) - e\Big(\frac{-\psi_j(x+kd)}{p}\Big)\right| \\
&\leqslant \left|e\Big(\frac{-ax^2-bx}{p}\Big) - e\Big(\frac{-ax^2-ak^2d^2-2akdx-bx-bkd}{p}\Big)\right| \\
&= \left|1 - e\Big(\frac{-ak^2d^2-2akdx-bkd}{p}\Big)\right| = 2\left|\sin\Big(\pi\cdot\frac{-ak^2d^2-2akdx-bkd}{p}\Big)\right| \\
&\leqslant 2\pi\left\|\frac{ak^2d^2+2akdx+bkd}{p}\right\| \leqslant 2\pi\ell^2\left\|\frac{ad^2}{p}\right\| + 2\pi\left\|\frac{(2ax+b)dk}{p}\right\|.
\end{aligned}$$

Now we split P_j into arithmetic subprogressions in which the first term is at most $\frac{c\alpha^C}{8}$, and then we split these subprogressions further to get the same upper bound for the second term.

Choose

$$\ell = \left\lfloor \Big(\frac{c\alpha^C}{16\pi\varsigma}\Big)^{\frac{1}{2(6\varkappa+1)}} \cdot (\#P_j)^{\frac{2\varkappa}{6\varkappa+1}} \right\rfloor.$$

Our corollary to Weyl's inequality allows us to choose $d \in \mathscr{J}_{\left\lfloor \sqrt[4\varkappa]{\frac{16\pi\varsigma\ell^2}{c\alpha^C}} \right\rfloor}$ in such a way that

$$\left\|\frac{ad^2}{p}\right\| \leqslant \frac{c\alpha^C}{16\pi\ell^2},$$

provided that $N \asymp p \gg 1$. Of course we now have

$$\left\lfloor \frac{\#P_j}{d} \right\rfloor \geqslant \left\lfloor \#P_j \sqrt[4\varkappa]{\frac{c\alpha^C}{16\pi\varsigma\ell^2}} \right\rfloor \geqslant \left\lfloor \ell^{\frac{6\varkappa+1}{2\varkappa}} \left(\frac{1}{\ell^2}\right)^{\frac{1}{4\varkappa}} \right\rfloor = \left\lfloor \ell^{3+\frac{1}{2\varkappa}} \ell^{-\frac{1}{2\varkappa}} \right\rfloor \geqslant \left\lfloor \ell^3 \right\rfloor \geqslant \ell^2.$$

We may now first split P_j into $\mathbb{Z}_p$-arithmetic progressions of common difference d, yielding arithmetic progressions of length at least $\left\lfloor \frac{\#P_j}{d} \right\rfloor \geqslant \ell^2$, and clearly we can further partition these subprogressions into subprogressions of common difference d and each having length ℓ or $\ell+1$.

Now suppose that Q is one these progressions. Let $\ell' = \left\lfloor \sqrt[3]{\frac{c\alpha^C\ell}{16\pi}} \right\rfloor$. Dirichlet's lemma gives us a number $d' \in \mathscr{J}_{\left\lfloor \frac{16\pi\ell'}{c\alpha^C} \right\rfloor}$ such that

$$\left\| \frac{(2ax+b)dd'}{p} \right\| \leqslant \frac{c\alpha^C}{16\pi\ell'}.$$

We break Q into arithmetic subprogressions of common difference dd'. The length of such a subprogression is at least $\left\lfloor \frac{\#Q}{d'} \right\rfloor$, and since

$$\left\lfloor \frac{\#Q}{d'} \right\rfloor \geqslant \left\lfloor \frac{\ell}{d'} \right\rfloor \geqslant \left\lfloor \frac{\ell}{\left(\frac{16\pi\ell'}{c\alpha^C}\right)} \right\rfloor \geqslant \left\lfloor \frac{\ell}{\frac{16\pi}{c\alpha^C}\sqrt[3]{\frac{c\alpha^C\ell}{16\pi}}} \right\rfloor \geqslant \left\lfloor \sqrt[3]{\frac{\ell^2\left(c\alpha^C\right)^2}{\left(16\pi\right)^2}} \right\rfloor \geqslant \left(\ell'\right)^2,$$

we can break these subprogressions into subprogressions each having common difference dd', and each having length ℓ' or $\ell'+1$.

It is clear that the last subprogressions obtained are the ones in which the corresponding quadratic phase factors are approximately constant precisely in the required way. Since $\#P_j \geqslant e^{-\alpha^{-C}}N^{\alpha^C}$ for each $j \in \mathscr{J}_M$, we originally had at most $e^{\alpha^{-C}}N^{1-\alpha^C}$ arithmetic progressions. Each of these was broken into subprogressions of length at least ℓ, and the partition was refined into subprogressions of length at least ℓ'. It is obvious that

$$\ell' = \left\lfloor \sqrt[3]{\frac{c\alpha^C\ell}{16\pi}} \right\rfloor = \left\lfloor \sqrt[3]{\frac{c\alpha^C}{16\pi}\left\lfloor \left(\frac{c\alpha^C}{16\pi\varsigma}\right)^{\frac{1}{2(6\varkappa+1)}} \cdot \left(\#P_j\right)^{\frac{2\varkappa}{6\varkappa+1}} \right\rfloor} \right\rfloor \geqslant e^{-\alpha^{C'}}N^{\alpha^{C'}},$$

for some constant $C' \in \mathbb{R}_+$. Thus we have split $\mathbb{Z}_p$ into at most $e^{\alpha^{-C'}}N^{1-\alpha^C}$ $\mathbb{Z}_p$-arithmetic progressions in which the corresponding quadratic phase factors are approximately constant. q.e.f.

Partitioning $\mathbb{Z}_p$ into $\mathbb{Z}$-arithmetic progressions

We have a partition of $\mathbb{Z}_p$ into $\mathbb{Z}_p$-arithmetic progressions $Q_1, Q_2, \ldots, Q_L$, for which $L \leqslant e^{\alpha^{-C}}N^{1-\alpha^C}$ and

$$\sum_{j=1}^{L}\left|\sum_{h\in Q_j} f_A(x)\right| \gg \alpha^C N,$$

where each C is an absolute real constant greater than one. We would rather like to have a partition into $\mathbb{Z}$-arithmetic progressions for ultimately we want a density increment in a $\mathbb{Z}$-arithmetic progression. For this purpose we introduce the following simple lemma.

Lemma. *A $\mathbb{Z}_p$-arithmetic progression P of length $\ell \in \mathscr{J}_p$ may be partitioned into at most $3\sqrt{\ell}$ $\mathbb{Z}$-arithmetic progressions.*

Proof. Let d denote the common difference between the terms of P. Abusing notation slightly, we may use Dirichlet's lemma to find an integer $r \in \mathscr{J}_{\lfloor\sqrt{\ell}\rfloor}$ such that

$$\left\| \frac{rd}{p} \right\| \leqslant \frac{1}{\sqrt{\ell}}.$$

We now partition P into $\mathbb{Z}_p$-arithmetic progressions of common difference rd, thereby obtaining at most r progressions. Because of the way in which r was chosen, each of these new progressions may be partitioned into $\mathbb{Z}$-arithmetic progressions, each of which, with at most two possible exceptions, has length at least $\sqrt{\ell}$. The progression P has been partitioned into at most $\frac{\ell}{\sqrt{\ell}} + 2r \leqslant 3\sqrt{\ell}$ $\mathbb{Z}$-arithmetic progressions. q.e.d.

Applying this lemma to each of the $\mathbb{Z}_p$-arithmetic progressions $Q_1, Q_2, \ldots, Q_L$ yields a partition of $\mathbb{Z}_p$ into $\mathbb{Z}$-arithmetic progressions $R_1, R_2, \ldots, R_H$ ($H \in \mathbb{Z}_+$), and in such a way that

$$\sum_{j=1}^{H} \left| \sum_{h \in R_j} f_A(x) \right| \gg \alpha^C N.$$

Since $\sum_{j=1}^{L} \#Q_j = N$, we have $H \leqslant 3\sum_{j=1}^{L} \sqrt{\#Q_j} \leqslant 3\sqrt{LN} \leqslant e^{\alpha^{-C'}} N^{1-\alpha^{C'}}$, for some even worse absolute real constant $C' > 1$.

Obtaining the density increment

We are now almost done. We have partitioned $\mathbb{Z}_p$ into $\mathbb{Z}$-arithmetic progressions $R_1, R_2, \ldots, R_H$ satisfying $H \leqslant e^{\alpha^{-C}} N^{1-\alpha^C}$ and

$$\sum_{j=1}^{H} \left| \sum_{h \in R_j} f_A(h) \right| \geqslant c\alpha^C N,$$

where $c \in]0,1[$ and $C \in]1,\infty[$ are absolute real constants. We now finish the proof in a way similar to the proof of Roth's theorem.

Since $\sum_x f_A(x) = 0$, we have

$$c\alpha^C N \leqslant \sum_{j=1}^{H} \left(\left| \sum_{h \in R_j} f_A(h) \right| + \sum_{h \in R_j} f_A(h) \right).$$

Thus the contribution of the arithmetic progressions R_j in which the density is at least α is at least $\frac{c\alpha^C N}{2}$. Among these, the contribution from those progressions in which the density increases by less than $\frac{c\alpha^C}{4}$ is clearly less than $\frac{c\alpha^C N}{4}$. On the other hand, since the function f_A is supported on $\mathscr{J}_N$, the contribution from those progressions R_j for which $R_j \cap \mathscr{J}_N$ is less than $\frac{c\alpha^C N}{4H}$ is clearly less than $H \cdot \frac{c\alpha^C N}{4H} = \frac{c\alpha^C N}{4}$. Therefore we have at least one progression R_j with both

$$\sum_{h \in R_j} f_A(h) \geqslant \frac{c\alpha^C \#R_j}{4}$$

and

$$\#(R_j \cap \mathscr{J}_N) \geqslant \frac{c\alpha^C N}{4H} \geqslant \frac{c\alpha^C}{4} e^{-\alpha^{-C}} N^{\alpha^C} \geqslant e^{-\alpha^{-C'}} N^{\alpha^{C'}},$$

where C' is an even larger absolute real constant.

We have obtained a proper $\mathbb{Z}$-arithmetic progression $R = R_j \cap \mathscr{J}_N$ which is contained in $\mathscr{J}_N$ and which has length at least $e^{-\alpha^{-C}} N^{\alpha^C}$, for some absolute real constant $C > 1$, and for which

$$\sum_{h \in R} f_A(h) = \sum_{h \in R_j} f_A(h) \gg \alpha^C \# R_j \gg \alpha^C \# R,$$

for some other absolute real constant $C > 1$, so that A has density at least $\alpha + \Omega\big(\alpha^C\big)$ in R. Q.E.D.

Some recent results on r_4

In [Gr&T3] Green and Tao generalize Gowers' bounds for r_4 for arbitrary finite Abelian groups. This is achieved through a better understanding of the functions with vanishing expectation and large $\mathscr{U}^3$-norm. They obtain sharper inverse results for the $\mathscr{U}^3$-norm. The following results depend on these improvements.

In [Gr&T4] it is proved that for finite fields F of characteristic not equal to two or three, and for dimensions $n \in \mathbb{Z}_+$, one has

$$r_4(F^n) \ll_F \frac{(\#F)^n}{n^c},$$

with $c \in \mathbb{R}_+$ being an absolute constant. In fact, the explicit value $c = 2^{-21}$ is provided.

The paper [Gr&T5] deals with the classical Erdős–Turán constants and proves that

$$r_4(N) \ll \frac{N}{e^{c\sqrt{\log \log N}}}, \qquad (N \longrightarrow \infty)$$

where $c \in \mathbb{R}_+$ is yet another absolute constant. The improvement over Gowers' bound is obtained by adapting the Heath-Brown–Szemerédi idea.

In [Gr&T5] Green and Tao announce that with more effort one can improve the estimate to the Heath-Brown–Szemerédi type bound

$$r_4(N) \ll \frac{N}{(\log N)^c}, \qquad (N \longrightarrow \infty)$$

with $c \in \mathbb{R}_+$being an absolute constant.

Longer progressions

We have proved Gowers' theorem for four-term progressions and so it remains to very briefly discuss what additional ingredients are needed to handle longer arithmetic progressions. We begin with the simple parts of the proof: The density increment strategy works in exactly the same way, and the quasi-random

part of the proof generalizes quite easily. As the generalized von Neumann theorem was proven in full for arbitrary $k \in \mathbb{Z}_+$, the quasi-random part would only require a few mostly cosmetic changes.

The non-quasirandom part of Gowers' proof for longer progressions follows the same plan as in the case of four-term progressions. Given a set whose balanced function is not Gowers uniform of appropriate degree, it will turn out that the balanced function correlates locally with higher degree phase functions. Proving this is, however, much more difficult than in the case of four-term progressions. Let us consider the case of five-term progressions. No new major ideas are required for even longer progressions.

In the case $k = 4$, we concluded that $\Delta_h f_A$ has a large Fourier-coefficient for many values of h, giving rise to a function φ from a large subset B of $\mathbb{Z}_N$ to $\mathbb{Z}_N$ such that $\widehat{\Delta_h f_A}\big(\varphi(h)\big)$ is large for every $h \in B$. In the case $k = 5$, we can instead conclude that $\Delta_{h_1}\Delta_{h_2} f_A$ has a large Fourier-coefficient for many values of the variables h_1 and h_2. This gives rise to a function φ from a large subset B of $\mathbb{Z}_N^2$ to $\mathbb{Z}_N$ with $\widehat{\Delta_{h_1}\Delta_{h_2} f_A}\big(\varphi(h_1, h_2)\big)$ being large for every $\langle h_1, h_2\rangle \in B$.

In the case of four-term progressions, φ was shown to coincide with a linear function on a long arithmetic progression through applications of the Balog–Szemerédi–Gowers theorem and arguments from a proof of the Freĭman–Ruzsa theorem. For $k > 4$, the function φ will depend on several variables, thereby requiring multidimensional versions of the tools used when $k = 4$. This is probably the most difficult obstacle in adapting the proof for longer progressions. Overcoming these obstacles allows us to conclude that a large piece of φ is multilinear. A technical induction argument on k will turn this multilinearity into a correlation with some high-degree phase factors. Once the local correlation is obtained, the density increment is obtained in more or less the same way as in the case of four-term progressions. Except for now the arguments are more technical and we need a corollary to Weyl's inequality for polynomials of arbitrary degree.

Roth's theorem and Gowers' proof of Szemerédi's theorem for four-term progressions have become part of a "standard" course in additive combinatorics, as is witnessed by the lecture notes [Go4, Gre8, Sou]. However, the original article [Go3] remains so far the only presentation of Gowers' Fourier-analytic proof of the entire Szemerédi theorem.

References and Sources

[A&Z] AIGNER, M., and G. M. ZIEGLER: *Proofs from the Book*, Springer, Berlin, 2004.

[Bal] BALOG, A.: *Many Additive Quadruples*, in [Gra&al.], 39–49.

[Bal&S] BALOG, A., and E. SZEMERÉDI: *A Statistical Theorem of Set Addition*, Combinatorica, **14** (1994), 263–268.

[Bar&al.] BARAK, B., L. TREVISAN, A. WIGDERSON (lecturers), et al.: *A Mini Course on Additive Combinatorics*, the first draft of the lecture notes available at
`http://www.cs.princeton.edu/theory/index.php/Main/AdditiveCombinatoricsMinicourse`

[Beh1] BEHREND, F. A.: *On Sequences of Integers Containing No Arithmetic Progressions*, Časopis Pěst. Mat., **67** (1938), 235–239.

[Beh2] BEHREND, F. A.: *On Sets of Integers Which Contain No Three Terms in Arithmetical Progression*, Proc. Nat. Acad. Sci., **32** (1946), 331–332.

[Ber1] BERGELSON, V.: *Ergodic Ramsey Theory — an Update*, in [Polli&S], 1–61.

[Ber2] BERGELSON, V.: *Ergodic Theory and Diophantine Problems*, in [Bl&al.], 167–205.

[Ber3] BERGELSON, V.: *Combinatorial and Diophantine Applications of Ergodic Theory* (with appendices by A. Leibman, A. Quas and M. Wierdl), in [Ha&K], 745–841.

[Ber4] BERGELSON, V.: *Ergodic Ramsey Theory: a Dynamical Approach to Static Theorems*, in [S-S&al.], 1655–1678.

[Ber&L] BERGELSON, V., and A. LEIBMAN: *Polynomial Extensions of van der Waerden's and Szemerédi's Theorems*, J. Amer. Math. Soc., **9** (1996), 725–753.

[Bi] BILU, YU.: *Structure of Sets with Small Sumset*, Astérisque, **258** (1999), 77–108.

[Bl&al.] BLANCHARD, F., A. MAASS, and A. NOGUEIRA (editors): *Topics in Symbolic Dynamics and Applications*, London Math. Soc. Lecture Note Ser., **279**, Cambridge University Press, Cambridge, 2000.

[Bo1] Bourgain, J.: *On Triples in Arithmetic Progression*, Geom. Funct. Anal., **9** (1999), 968–984.

[Bo2] Bourgain, J.: *Roth's Theorem on Arithmetic Progressions Revisited*, J. Analyse Math., **104** (2008), 155–192.

[Chand] Chandrasekharan, K.: *Introduction to Analytic Number Theory*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, **148**, Springer-Verlag, Berlin, 1968.

[Chang1] Chang, M.-C.: *A Polynomial Bound in Freĭman's Theorem*, Duke Math. J., **113** (2002), 399–419.

[Chang2] Chang, M.-C.: *On Problems of Erdős and Rudin*, J. Funct. Anal., **207** (2004), 444–460.

[Che&al.] Chen, W. W. L., W. T. Gowers, H. Halberstam, W. M. Schmidt, and R. C. Vaughan (editors): *Analytic Number Theory: Essays in Honour of Klaus Roth*, Cambridge University Press, New York, 2009.

[Chi] Chipeniuk, K. O.: *Exposing Roth's Theorem in the Primes*, exposition available at
http://www.math.ubc.ca/∼karstenc/RothinPrimes_Final.pdf

[Cho] Chowla, S.: *There Exists an Infinity of 3-Combinations of Primes in A. P.*, Proc. Lahore Philos. Soc, **2** (1944), 15–16.

[Chuda] Chudakov, N. G.: *On the Density of the Set of Even Numbers which are not Representable as a Sum of Two Odd Primes* (in Russian), Izv. Akad. Nauk. SSSR Ser. Math., **1** (1938), 25–40.

[Chudn&al.] Chudnovsky, D. V., G. Chudnovsky, and M. B. Nathanson (editors): *Number Theory: New York Seminar, 1991–1995*, Springer Verlag, New York, 1996.

[vdC] van der Corput, J. G.: *Über Summen von Primzahlen and Primzahlquadraten*, Math. Ann., **116** (1939), 1–50.

[Cr] Croot, E.: *Roth's Theorem on 3-Term Arithmetic Progressions*, available at
http://www.math.gatech.edu/∼ecroot/roth.pdf.

[Cr&L] Croot, E. S., and V. F. Lev: *Open Problems in Additive Combinatorics*, in [Gra&al.], 207–233.

[Cr&S] Croot, E., and O. Sisask: *A New Proof of Roth's Theorem on Arithmetic Progressions*, available at
http://www.math.gatech.edu/∼ecroot/noniterativeroth3.pdf.

[El] Elkin, M.: *An Improved Construction of Progression-Free Sets*, arXiv:0801.4310v1.

[Er1] Erdős, P.: *Some Unsolved Problems*, Michigan Math. J., **4** (1957), 291–300.

[Er2] Erdős, P.: *Some Unsolved Problems*, Magyar tudomanyos akademia mathematikai kutato intezetenek közlemenyel, **6** ser. A (1–2) (1961), 221–254.

[Er&T] Erdős, P., and P. Turán: *On Some Sequences of Integers*, J. London Math. Soc., **11** (1936), 261–264.

[Fr1] Freĭman, G. A.: *On the Addition of Finite Sets* (in Russian), Dokl. Akad. Nauk. SSSR, **158** (1964), 1038–1041.

[Fr2] Freĭman, G. A.: *Elements of a Structural Theory of Set Addition* (in Russian), Kazan. Gosudarstv. Ped. Inst., Elabuž. Gosudarstv. Ped. Inst., Kazan, 1966.

[Fr3] Freĭman, G. A.: *Foundations of a Structural Theory of Set Addition*, Trans. Math. Monogr., **37**, Amer. Math. Soc., Providence, Rhode Island, USA, 1973.

[Fu] Furstenberg, H.: *Ergodic Behaviour of Diagonal Measures and a Theorem of Szemerédi on Arithmetic Progressions*, J. Analyse Math., **31** (1977), 204–256.

[Fu&K1] Furstenberg, H., and Y. Katznelson: *An Ergodic Szemerédi Theorem for Commuting Transformations*, J. Analyse Math., 34 (1978), 275–291.

[Fu&K2] Furstenberg, H., and Y. Katznelson: *A Density Version of the Hales–Jewett Theorem*, J. Analyse Math., **57** (1991), 64–119.

[Fu&al.] Furstenberg, H., Y. Katznelson, and D. Ornstein: *The Ergodic Theoretical Proof of Szemerédi's Theorem*, Bull. Amer. Math. Soc., **7** (1982), 527–552.

[Ge&R] Geroldinger, A., and I. Z. Ruzsa: *Combinatorial Number Theory and Additive Group Theory*, Advanced Courses in Mathematics — CRM Barcelona, Birkhäuser Verlag, Basel, 2009.

[Go1] Gowers, W. T.: *Lower Bounds of Tower Type for Szemerédi's Uniformity Lemma*, Geom. Funct. Anal., **7** (1997), 322–337.

[Go2] Gowers, W. T.: *A New Proof of Szemerédi's Theorem for Arithmetic Progressions of Length Four*, Geom. Funct. Anal., **8** (1998), 529–551.

[Go3] Gowers, W. T.: *A New Proof of Szemerédi's Theorem*, Geom. Funct. Anal., **11**, (2001), 465–588.

[Go4] Gowers, W. T.: *Additive and Combinatorial Number Theory*, online lecture notes at
`http://www.dpmms.cam.ac.uk/~wtg10/addnoth.notes.dvi`.

[Go5] Gowers, W. T.: *Quasirandomness, Counting and Regularity for 3-Uniform Hypergraphs*, Combin. Probab. Comput., **15** (2006), 143–184.

[Go6] Gowers, W. T.: *Hypergraph Regularity and the Multidimensional Szemerédi Theorem*, arXiv:0710.3032v1.

[Gra] Granville, A.: *An Introduction to Additive Combinatorics*, in [Gra&al.], 1–27.

[Gra&al.] Granville, A., M. B. Nathanson, and J. Solymosi (editors): *Additive Combinatorics*, CRM Proc. & Lecture Notes, **43**, Amer. Math. Soc., Providence, Rhode Island, USA, 2007.

[Gre1] Green, B. J.: *Structure Theory of Set Addition*, expository notes available at
http://www.dpmms.cam.ac.uk/∼bjg23/papers/icmsnotes.pdf

[Gre2] Green, B. J.: *Finite Field Models in Additive Combinatorics*, in [Web], 1–27.

[Gre3] Green, B. J.: *Progressions of Length 3 Following Szemerédi*, exposition available at
http://www.dpmms.cam.ac.uk/∼bjg23/papers/szemeredi-roth.pdf.

[Gre4] Green, B. J.: *On Triples in Arithmetic Progression*, exposition available at
http://www.dpmms.cam.ac.uk/∼bjg23/papers/bourgain-roth.pdf.

[Gre5] Green, B. J.: *Long Arithmetic Progressions of Primes*, arXiv:math/0508063v1.

[Gre6] Green, B. J.: *Roth's Theorem in the Primes*, Annals of Math., **161** (2005), 1609–1636.

[Gre7] Green, B. J.: *Montréal Notes on Quadratic Fourier Analysis*, in [Gra&al.], 69–102.

[Gre8] Green, B. J.: *Additive Combinatorics*, online lectures notes at
http://www.dpmms.cam.ac.uk/∼bjg23/add-combinatorics.html.

[Gr&R] Green, B. J., and I. Ruzsa: Freĭman's Theorem in an Arbitrary Abelian Group, J. London Math. Soc., **75** (2007), 163–175.

[Gr&T1] Green, B. J., and T. Tao: *Szemerédi's theorem*, Scholarpedia, 2(7):3446. Available at
http://www.scholarpedia.org/article/Szemeredi%27s_theorem.

[Gr&T2] Green, B. J., and T. Tao: *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.

[Gr&T3] Green, B. J., and T. Tao: *An Inverse Theorem for the Gowers U^3 Norm*, Proc. Edin. Math. Soc., **51** (2008), 73–153.

[Gr&T4] Green, B. J., and T. Tao: *New Bounds for Szemerédi's Theorem, I: Progressions of Length 4 in Finite Field Geometries*, Proc. Lond. Math. Soc., **98** (2009), 365–392.

[Gr&T5] Green B. J., and T. Tao: *New Bounds for Szemerédi's Theorem, II: A New Bound for $r_4(N)$*, in [Che&al.], 180–204.

[Gr&T6] Green, B. J., and T. Tao: *Linear Equations in Primes*, to appear in Annals of Math.

[Gr&W] Green, B. J., and J. Wolf: *A Note on Elkin's Improvement of Behrend's Construction*, arXiv:0810.0732v1.

[Ha&K] Hasselblatt, B., and A. Katok (editors): *Handbook of Dynamical Systems, Volume 1B*, Elsevier, Amsterdam, 2006.

[H-B] Heath-Brown, D. R.: *Integer Sets Containing no Arithmetic Progressions*, J. London Math. Soc., **35** (1987), 385–394.

[Hi] Hilbert, D.: *Über die Irreduzibilatät ganzer rationaler Functionen mit ganzzahligen Koeffizienten*, J. Reine Angew. Math., **110** (1892), 104–129.

[Ho] : Host, B.: *Progressions arithmétiques dans les nombres premiers, d'après B. Green et T. Tao*, arXiv:math/0609795v1.

[Kh] Khinchin, A. Y.: *Three Pearls of Number Theory*, Dover Publications, New York, 1998.

[Kn] Knopp, M. I. (editor): *Number Theory: Proceedings, Philadelphia, 1980*, Lecture Notes in Mathematics, **899**, Springer-Verlag, New York, 1981.

[Ko&S] Komlós, J., and M. Simonovits: *Szemerédi's Regularity Lemma and its Applications in Graph Theory*, in [Mik&al], 295–352.

[Ko&al.] Komlós, J., A. Shokoufandeh, M. Simonovits, and E. Szemerédi: *Szemerédi's Regularity Lemma and its Applications in Graph Theory*, in [Mik&al], 295–352.

[Kr] Kra, B.: *Ergodic Methods in Additive Combinatorics*, in [Gra&al.], 103–143.

[Ł] Łaba, I.: *From Harmonic Analysis to Arithmetic Combinatorics*, Bull. Amer. Math. Soc., **45** (2008), 77–115.

[Ł&L] Łaba, I., and M. T. Lacey: *On Sets of Integers Not Containing Long Arithmetic Progressions*, arXiv:math/0108155v1.

[L&al.] Landman, B., A. Robertson, and C. Culver: *Some New Exact van der Waerden Numbers*, arXiv:math/0507019v1.

[Mik&al] Miklós, D., V. T. Sós, and T. Szönyi: *Combinatorics, Paul Erdős is Eighty (Volume 2)*, János Bolyai Math. Soc., Budapest, 1996.

[Mir] Mirsky, L. (editor): *Papers in Combinatorial Theory, Analysis, Geometry, and the Theory of Numbers presented to Richard Rado on the occasion of his sixty-fifth birthday*, Academic Press, London and New York, 1971.

[Mo] Moser, L.: *On Non-Averaging Sets of Integers*, Canadian J. Math., **5** (1953), 245–252.

[Nag&al.] Nagle, B., V. Rödl, and M. Schacht: *The Counting Lemma for Regular k-Uniform Hypergraph*, Random Structures & Algorithms, **28** (2006), 113–179.

[Nat] Nathanson, M. B.: *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Graduate Texts in Mathematics, **165**, Springer-Verlag, New York, 1996.

[Ne] Newman, D. J.: *Sequences Without Arithmetic Progressions*, in [Kn], 311–314.

[Polla] Pollack, P.: *Not Always Buried Deep*, online lectures notes at http://www.math.dartmouth.edu/∼ppollack/notes.pdf.

[Polli&S] Pollicott, M., and K. Schmidt (editors): *Ergodic Theory of Z^d-actions*, London Math. Soc. Lecture Note Ser., **228**, Cambdridge University Press, Cambridge, 1996.

[Ra] Rankin, R. A.: *Sets of Integers Containing not more than a Given Number of Terms in Arithmetical Progression*, Proc. Roy. Soc. Edinburgh Sect A., **65** (1960), 332–344.

[Rö&Sc1] Rödl, V., and M. Schacht: *Regular Partitions of Hypergraphs: Regularity Lemmas*, Combin. Probab. Comput., **16** (2007), 833–885.

[Rö&Sc2] Rödl, V., and M. Schacht: *Regular Partitions of Hypergraphs: Counting Lemmas*, Combin. Probab. Comput., **16** (2007), 887–901.

[Rö&Sk1] Rödl, V., and J. Skokan: *Regularity Lemma for Uniform Hypergraphs*, Random Structures & Algorithms, **25** (2004), 1–42.

[Rö&Sk2] Rödl, V., and J. Skokan: *Applications of the Regularity Lemma for Uniform Hypergraphs*, Random Structures & Algorithms, **28** (2006), 180–194.

[Ro1] Roth, K. F.: *Sur quelques ensembles d'entiers*, Comptes Rendus, **234** (1952), 388–390.

[Ro2] Roth, K. F.: *On Certain Sets of Integers*, J. London Math. Soc., **28** (1953), 104–109.

[Ro3] Roth, K. F.: *On Certain Sets of Integers (II)*, J. London Math. Soc., **29** (1954), 20–26.

[Ro4] Roth, K. F.: *Irregularities of Sequences Relative to Arithmetic Progressions*, Math. Ann., **169** (1967), 1–25.

[Ro5] Roth, K. F.: *Irregularities of Sequences Relative to Arithmetic Progressions III*, J. Number Theory, **2** (1970), 125–142.

[Ro6] Roth, K. F.: *Irregularities of Sequences Relative to Arithmetic Progressions IV*, Period. Math. Hungar., **2** (1972), 301–326.

[Ru1] Ruzsa, I. Z.: *An Application of Graph Theory to Additive Number Theory*, Scientia, **3** (1989), 97–109.

[Ru2] Ruzsa, I. Z.: *Arithmetic Progressions and the Number of Sums*, Period. Math. Hungar., **25** (1992), 105–111.

[Ru3] Ruzsa, I. Z.: *Generalized Arithmetic Progressions and Sumsets*, Acta Math. Hungar., **65** (1994), 379–388.

[Ru4] Ruzsa, I. Z.: *Sums of Finite Sets*, in [Chudn&al.], 281–293.

[Ru&Sz] Ruzsa, I. Z., and E. Szemerédi: *Triple Systems with no Six Points Carrying Three Triangles*, Colloq. Math. Soc. J. Bolyai, **18** (1978), 939–945.

[San] Sanders, T.: *Appendix to "Roth's Theorem on Progressions Revisited" by J. Bourgain*, J. Anal. Math., **104** (2008), 193–206.

[S-S&al.] Sanz-Solé, M., J. Soria, J. L. Varona, and J. Verdera (editors): *Proceedings of the International Congress of Mathematicians Madrid, August 22–30, 2006, Volume II*, European Math. Soc. Publishing House, Zürich, 2007.

[Sal&S1] Salem, R., and D. C. Spencer: *On Sets of Integers Which Contain No Three Terms in Arithmetical Progression*, Proc. Nat. Acad. Sci., **28** (1942), 561–563.

[Sal&S2] Salem, R., and D. C. Spencer: *On Sets Which Do Not Contain a Given Number of Terms in Arithmetical Progression*, Nieuw Archief voor Wiskunde, **23** (1950), 133–143.

[Sc] Schur, I.: *Über die Kongruenz* $x^m + y^m \equiv z^m \pmod{p}$, Jber. Deutsch. Math.-Verein., **25** (1916), 114–116.

[Sh] Shkredov, I. D.: *Szemerédi's Theorem and Problems on Arithmetic Progressions*, Russian Math. Surveys, **61** (2006), 1101–1166.

[Soi] Soifer, A.: *The Mathematical Coloring Book: Mathematics of Coloring and the Colorful Life of its Creators*, Springer, New York, 2009.

[Sol] Solymosi, J.: *Elementary Additive Combinatorics*, in [Gra&al.], 29–38.

[Sou] Soundararajan, K.: *Additive Combinatorics*, online lecture notes available at `http://math.stanford.edu/~ksound/Notes.pdf`.

[Sz1] Szemerédi, E.: *On Sets of Integers Containing No Four Elements in Arithmetic Progression*, Acta Math. Acad. Sci. Hungar., **20** (1969), 89–104.

[Sz2] Szemerédi, E.: *On Sets of Integers Containing No k Elements in Arithmetic Progression*, Acta Arith., **27** (1975), 299–345.

[Sz3] Szemerédi, E.: *Integer Sets Containing no Arithmetic Progressions*, Acta Math. Hung., **56** (1990), 155–158.

[Sz4] Szemerédi, E.: *An Old New Proof of Roth's Theorem*, in [Gra&al.], 51–54.

[T1] Tao, T.: *Gowers' Proof of the Szemerédi Theorem on Arithmetic Progressions of Length 4*, exposition available at
`http://www.math.ucla.edu/~tao/preprints/Expository/ap4.dvi`

[T2] Tao, T.: *Szemerédi's Proof of Szemerédi's Theorem*, exposition available at
`http://www.math.ucla.edu/~tao/preprints/Expository/szemeredi_theorem.dvi`

[T3] Tao, T.: *A Quantitative Ergodic Theory Proof of Szemerédi's Theorem (Abridged)*, exposition available at
`http://www.math.ucla.edu/~tao/preprints/Expository/quantitative_AP.dvi`

[T4] Tao, T.: *Arithmetic Progressions and the Primes — El Escorial Lectures*, arXiv:math/0411246v1.

[T5] Tao, T.: *The Dichotomy Between Structure and Randomness, Arithmetic Progressions, and the Primes*, arXiv:math/0512114v2.

[T6] Tao, T.: *A Quantitative Ergodic Theory Proof of Szemerédi's Theorem*, Electron. J. Combin., **13** (2006), 1–49.

[T7] Tao, T.: *The Gaussian Primes Contain Arbitrarily Shaped Constellations*, J. Analyse Math., **99** (2006), 109–176.

[T8] Tao, T.: *The Ergodic and Combinatorial Approaches to Szemerédi's Theorem*, in [Gra&al.], 145–193.

[T&V] Tao, T., and V. H. Vu: *Additive Combinatorics*, Cambridge studies in advanced mathematics, **105**, Cambridge University Press, Cambridge, 2007.

[T&Z] Tao, T., and T. Ziegler: *The Primes Contain Arbitrarily Long Polynomial Progressions*, Acta Math., **201** (2008), 213-305.

[Var1] Varnavides, P.: *Note on a Theorem of Roth*, J. London Math. Soc., **30** (1955), 325–326.

[Var2] Varnavides, P.: *On Certain Sets of Positive Density*, J. London Math. Soc., **34** (1959), 358–360.

[Vau] Vaughan, R. C.: *The Hardy–Littlewood Method*, Cambridge Tracts in Mathematics, **125**, Cambridge University Press, Cambridge, 1997.

[vdW1] van der Waerden, B. L.: *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wisk., **15** (1927), 212–216.

[vdW2] van der Waerden, B. L.: *How the Proof of Baudet's Conjecture was Found*, in [Mir], 251–260. Also reproduced in [Soi], chapter 33.

[Web] Webb, B. S. (editor): *Surveys in Combinatorics, 2005*, London Math. Soc. Lecture Note Series, **327**, Cambridge University Press, Cambridge, 2005.

[Wey] Weyl, H.: *Über die Gleichverteilung von Zahlen mod Eins*, Math. Ann., **77** (1913), 313–352.