352

ASTÉRISQUE

2013

SÉMINAIRE BOURBAKI
VOLUME 2011/2012
EXPOSÉS 1043-1058

1054) *Arithmetic and polynomial progressions in the primes*

Julia WOLF

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2012
64e année, 2011-2012, no 1054, p. 389 à 427

ARITHMETIC AND POLYNOMIAL PROGRESSIONS IN THE PRIMES
[after Gowers, Green, Tao and Ziegler]

by **Julia WOLF**

1. INTRODUCTION

In 2004 Green and Tao [**25**] proved the following groundbreaking result.

THEOREM 1 (Green-Tao theorem). — *The primes contain arbitrarily long arithmetic progressions. Moreover, the same is true of any subset of the primes of positive relative density.*[(1)(2)]

Theorem 1 vastly generalizes van der Corput's result [**11**] that there are infinitely many 3-term arithmetic progressions in the primes, as well as a significant strengthening due to Green [**22**], which established the existence of 3-term progressions in any subset of the primes of positive relative density. It also represents a special case of a conjecture by Erdős and Turán [**12**], dating back to 1936.

CONJECTURE 2 (Erdős-Turán conjecture). — *Any subset* $X \subseteq \mathbb{N}$ *satisfying*

$$\sum_{x \in X} \frac{1}{x} = +\infty$$

contains arbitrarily long arithmetic progressions.

What is truly remarkable about Theorem 1 is the diversity of methods which are brought together in its proof: it combines tools from arithmetic combinatorics (especially so-called higher-order Fourier analysis) with traditional analytic number

(1) It is easy to see that the primes cannot contain an infinite arithmetic progression. Suppose that $P(j) = a + jd$ for some $a, d \in \mathbb{N}$ takes prime values for $j = 0, 1, \ldots, k$. Then $P(a) \equiv 0 \bmod a$ and $P(a) > a$. But if $P(a) = ma$ for some integer $m > 1$, it is no longer prime and hence $k < a$.
(2) The longest currently known arithmetic progression in the primes, $43\ 142\ 746\ 595\ 714\ 191 + 23\ 681\ 770 \times 223\ 092\ 870 \times j$, where $j = 0, 1, \ldots 25$, was found by Périchon with software by Wróblewski and Reynolds in 2010.

theory, all while taking inspiration from ergodic theory. The Erdős-Turán conjecture suggests that the existence of arithmetic structure in the primes is in fact merely a consequence of their density, and so it is perhaps not surprising that analytic and combinatorial (rather than classical number-theoretic) methods should play a major role. Indeed, in 1975 Szemerédi [**47**] showed in a purely combinatorial fashion that any subset of the integers of positive upper density contains arbitrarily long arithmetic progressions.[(3)] Denoting by $[N]$ the set of integers $\{1, 2, \dots, N\}$, we have the following finitary version of this statement.

THEOREM 3 (Szemerédi's theorem). — *Suppose that $A \subseteq [N]$ is a subset of density α which contains no k-term arithmetic progressions. Then*

$$\alpha = o_k(1),$$

where $o_k(1)$ is a quantity that tends to zero as N tends to infinity.

However, our current understanding of the decay rate of α does not allow us to immediately deduce Theorem 1. Indeed, the best known bound on the density of a subset of $[N]$ that contains no 3-term progressions, due to recent work of Sanders [**43**], is of the form $(\log N)^{-(1-o(1))}$, falling just short of the density of the primes. For longer progressions, the discrepancy is much more alarming. For length 4, the best known bound on α is of the form $\exp(-c\sqrt{\log \log N})$ [**26**], while for longer progressions it is $(\log \log N)^{-c}$ for some small positive constant c depending on k [**17**]. On the other hand, the best known example of a 3-term progression free set has density $\exp(-c\sqrt{\log N})$ [**3**], which is believed by many to be closer to the truth. However, proving upper bounds of this shape seems very much out of reach of currently available techniques (but see a recent result of Schoen and Shkredov [**46**]).

Many excellent expository articles have been written on the proof of the Green-Tao theorem [**23, 36, 48**]; in particular, it was covered in the Séminaire Bourbaki in 2005 by Bernard Host [**33**]. In contrast to Host's ergodic theoretic perspective we adopt a more analytic viewpoint in the present exposition. Moreover, our main focus will be on the developments that have taken place since the original proof of the Green-Tao theorem, which have brought new understanding and a number of additional exciting results to the subject.

Shortly after the proof of Theorem 1, Green and Tao [**31**] extended their result from arithmetic progressions to solutions of more general systems of linear equations in the primes. This work covered essentially all systems of linear equations for which the conclusion is neither trivially false nor known to be extremely difficult (such as those systems related to Goldbach's conjecture or the twin primes problem), but was

(3) A qualitative proof was given by Furstenberg [**14**] in 1977 and initiated the long-standing and fruitful interaction between combinatorics and ergodic theory.

conditional on two conjectures: the *inverse conjecture for the uniformity norms*, and the *Möbius nilsequences conjecture*. The latter was established by Green and Tao **[28]** shortly afterwards, and the former very recently by the same authors in joint work with Ziegler **[30]**. With the completion of this very substantial research programme the authors are able to assert not only the existence of general linear patterns in the primes, but give precise asymptotics for their frequency in the interval $[N]$.

In a further step towards generalization, Tao and Ziegler **[50]** proved in 2008 that the primes contain arbitrarily long polynomial progressions.

THEOREM 4 (Tao-Ziegler theorem). — *Given polynomials $P_1, \dots, P_k \in \mathbb{Z}[m]$ such that $P_1(0) = \cdots = P_k(0) = 0$, there exist infinitely many integers x, m such that $x + P_1(m), \dots, x + P_k(m)$ are simultaneously prime. Moreover, the same is true of any subset of the primes of positive relative density.*

The first non-trivial example of such a polynomial pattern is a configuration consisting of two elements that differ by a square, which corresponds to $P_1(m) = 0$, $P_2(m) = m^2$. In dense subsets of the integers the existence of such a configuration is guaranteed by a theorem of Sárközy **[45]**, which is obtained using a sophisticated application of the circle method. In fact, and in contrast with the situation for 3-term arithmetic progressions described above, the best known bound in Sárközy's theorem is strong enough to directly imply the existence of square differences in any positive-density subset of the primes. In the case of more general polynomial configurations, however, the results from arithmetic combinatorics are very far from implying a statement resembling that of Theorem 4. Worse, there is currently *no quantitative theorem at all* in the literature asserting that if a subset $A \subseteq [N]$ is dense enough, then it contains a polynomial configuration of the above type.[(4)] [(5)] What we do have is a *qualitative* polynomial Szemerédi theorem due to Bergelson and Leibman **[6]** proved by ergodic theoretic methods, whose statement is fundamental to the proof of Theorem 4. Moreover, Tao and Ziegler rely heavily on an induction technique which allowed Bergelson and Leibman to linearize a system of polynomials in successive stages, known as *PET induction* **[4]**.

The general strategy of proof for Theorem 4 is largely the same as in the case of (linear) arithmetic progressions. There are two main novelties here: first, a *transference principle* is explicitly formulated for the first time, which was implicit in and absolutely fundamental to the proof of the Green-Tao theorem. Roughly speaking, the problem is that the von Mangoldt function (a weighted indicator function of the

(4) A paper by Green **[21]**, which proves the existence of a 3-term progression whose common difference is a sum of two squares in any dense subset of the integers, may be regarded as an exception.
(5) There are, however, colouring results of this type, see the combinatorial proof of the polynomial van der Waerden theorem by Walters **[54]**.

primes) is unbounded, while the quantitative techniques from arithmetic combinatorics only apply to bounded functions. However, it turns out that the von Mangoldt function can be majorized by a so-called *pseudorandom measure* which is quite well-behaved. In particular, it can be shown that many statements involving truly bounded functions hold also, by transference, for functions that are bounded by this pseudorandom measure. The transference principle has received significant simplifications and a new conceptual context through work done by Gowers [**18**], and it is this more recent viewpoint that we shall adopt in our exposition. A similar approach was independently discovered by Reingold, Trevisan, Tulsiani and Vadhan [**40**] in theoretical computer science, where the transference principle is known as the *dense model theorem* and has found several applications in the context of complexity theory.

The second novelty concerns a new family of norms. In his work on Szemerédi's theorem, Gowers [**17**] introduced the U^k *norms*, often called *uniformity norms* or *Gowers norms*, and showed that the $(k+1)$-term progression count of a function is approximately invariant under small perturbations in the U^k norm – in other words, the uniformity norms control long arithmetic progressions. This raises the question of what can be said about functions that are large in the U^k norm, which is answered by the so-called *inverse theorem*. It is one of the central results in arithmetic combinatorics (although for $k > 3$, its strong form was only a conjecture until very recently) and states, roughly speaking, that if the U^k norm of a function is large, then the function correlates with a polynomial structure of degree $k-1$.[(6)] The inverse theorem, together with the above-mentioned approximate invariance under small perturbations in the U^k norm, is essentially sufficient to prove Szemerédi's theorem.

It is not too difficult to see that the uniformity norms are not sufficient for controlling polynomial configurations. To put it very simply, the reason is that in a linear configuration such as $x, x+d, x+2d$, which defines a 3-term progression, the range of both variables x and d is essentially linear in N. In contrast, in a configuration such as $x, x+m^2$, which represents a square difference, the range of m has to be restricted to $\sqrt{N}$. Dealing with smaller parameter ranges required Tao and Ziegler to introduce new *local uniformity norms*, and study some of their properties.

To conclude this section we give a brief overview of the structure of this paper. In Section 2 we define the uniformity norms and develop some of the fundamental notions of higher-order Fourier analysis, following Gowers's harmonic analysis approach to Szemerédi's theorem. In Section 3 we show how one uses the existence of a pseudorandom measure and the transference principle together with Szemerédi's theorem to obtain the Green-Tao theorem on arithmetic progressions in the primes. Finally, in

(6) The uniformity norms have also appeared in the context of ergodic theory. A deep result of Host and Kra [**34**] on the structure of characteristic factors for certain multiple ergodic averages is in some sense analogous to the above-mentioned inverse theorem.