Squares in arithmetic progressions and infinitely many primes

Andrew Granville

Abstract. We give a new proof that there are infinitely many primes, relying on van der Waerden's theorem for coloring the integers, and Fermat's theorem that there cannot be four squares in an arithmetic progression. We go on to discuss where else these ideas have come together in the past.

1. INFINITELY MANY PRIMES Levent Alpoge recently gave a rather different proof [**1**] that there are infinitely many primes. His starting point was the famous result of van der Waerden (see, e.g., [**9**]):

van der Waerden's Theorem. *Fix integers $m \geq 2$ and $\ell \geq 3$. If every positive integer is assigned one of m colors, in any way at all, then there is an ℓ-term arithmetic progression of integers which each have the same color.*

Using a clever coloring in van der Waerden's theorem, and some elementary number theory, Alpoge deduced that there are infinitely many primes. We proceed from van der Waerden's theorem a little differently, employing a famous result of Fermat (see, e.g., [**6**]):

Fermat's Theorem. *There are no four-term arithmetic progressions of distinct integer squares.*

From these two results we deduce the following:

Theorem 1. *There are infinitely many primes.*

Proof. If there are only finitely many primes $p_1, \dots, p_k$, then every integer n can be written as $p_1^{e_1} \cdots p_k^{e_k}$ for some integers $e_1, e_2, \dots, e_k \geq 0$. We can write each of these exponents e_j as

$$e_j = 2q_j + r_j, \text{ where } r_j \text{ is the "remainder" when dividing } e_j \text{ by 2,}$$

and equals 0 or 1. Therefore if we let

$$R = p_1^{r_1} \cdots p_k^{r_k}$$

then R is a squarefree integer that divides n, and

$$n/R \text{ is the square of an integer.}$$

(In fact, $n/R = Q^2$ where $Q = p_1^{q_1} \cdots p_k^{q_k}$.)

We will use 2^k colors to color the integers: Integer n is colored by the vector $(r_1, \dots, r_k)$. By van der Waerden's theorem there are four integers in arithmetic progression

$$A, A + D, A + 2D, A + 3D, \text{ with } D \geq 1,$$

which all have the same color $(r_1, \ldots, r_k)$. Now $R = p_1^{r_1} \cdots p_k^{r_k}$ divides each of these numbers, so also divides $D = (A + D) - A$. Letting $a = A/R$ and $d = D/R$, we see that

$$a, a + d, a + 2d, a + 3d \text{ are four squares in arithmetic progression,}$$

contradicting Fermat's theorem. ∎

These ideas have come together before to make a rather different, not-too-obvious deduction:

2. THE NUMBER OF SQUARES IN A LONG ARITHMETIC PROGRESSION

Let $Q(N)$ denote the maximum number of squares that there can be in an arithmetic progression of length N. A slight refinement of the Erdős–Rudin conjecture states that the maximum number is attained by the arithmetic progression

$$\{24n + 1 : \ 0 \le n \le N - 1\}$$

which contains $\sqrt{\frac{8}{3}N}$ squares, plus or minus one. From Fermat's theorem one easily sees that

$$Q(N) \le \frac{3N + 3}{4},$$

but it is difficult to see how to improve the bound to, say, $Q(N) \le \delta N + b$ for some constant $\delta < \frac{3}{4}$.

It was this problem that inspired one of the most influential results [**8**] in combinatorics and analysis (see, e.g., [**5**]):

Szemerédi's Theorem. *Fix $\delta > 0$ and integer $\ell \ge 3$. If N is sufficiently large (depending on δ and ℓ) then any subset A of $\{1, 2, \ldots, N\}$ with $\ge \delta N$ elements, must contain an ℓ-term arithmetic progression.*

van der Waerden's theorem is a consequence of Szemerédi's theorem, because if we let $\delta = 1/m$ and we color the integers in $\{1, 2, \ldots, N\}$ with m colors, then at least one of the colors is used for at least N/m integers. We apply Szemerédi's theorem to this subset A of $\{1, 2, \ldots, N\}$, to obtain an ℓ-term arithmetic progression of integers which each have the same color.

In [**7**], Szemerédi applied his result to the question of squares in arithmetic progressions:

Theorem 2 (Szemerédi). *For any constant $\delta > 0$, if N is sufficiently large, then $Q(N) < \delta N$.*

Proof. Suppose that there are at least δN squares in the arithmetic progression $\{r + ns : \ n = 1, 2, \ldots, N\}$ with $s \ge 1$; that is, there exists a subset A of $\{1, 2, \ldots, N\}$ with at least δN elements for which

$$r + ns \text{ is a square, whenever } a \in A.$$

Szemerédi's theorem with $\ell = 4$ then implies that A contains a four-term arithmetic progression, say $u + jv$ for $j = 0, 1, 2, 3$. For these values of n, we have $r + ns = a + jd$, where $a = r + us$ and $d = vs > 1$. That is, we have shown that

$$a, a + d, a + 2d, a + 3d \text{ are four squares in arithmetic progression,}$$

contradicting Fermat's theorem. ∎

3. MORE HEAVY MACHINERY One day over lunch, in late 1989, Bombieri showed me a completely different proof of Theorem 2, this time relying on one of the most influential results in algebraic and arithmetic geometry, Faltings' theorem [**3**]. Faltings' theorem is not easy to state, requiring a general understanding of an algebraic curve and its genus. The basic idea is that an equation in two variables with rational coefficients has only finitely many rational solutions (that is, solutions in which the two variables are rational numbers), unless the equation "boils down to" an equation of degree 1, 2 or 3. To be precise about "boiling down" involves the concept of *genus*, which is too complicated to explain here (see, e.g., [**3**]). Here we only need a simple consequence of Faltings' theorem.

Corollary to Faltings' Theorem. *Let $b_1, b_2, \ldots, b_k$ be distinct integers with $k \geq 5$. Then there are only finitely many rational numbers x for which*

$$(x + b_1)(x + b_2) \cdots (x + b_k) \text{ is the square of a rational number.}$$

Another proof of Theorem 2. Fix an integer $M > 6/\delta$. Let $B(M)$ be the total number of rational numbers x and integer 6-tuples $b_1 = 0 < b_2 < \ldots < b_6 \leq M - 1$ for which $(x + b_1)(x + b_2) \cdots (x + b_6)$ is the square of a rational number. Faltings' theorem implies that $B(M)$ is some finite number, as there are only finitely many choices for the b_j. We let N be any integer $\geq M(B(M) + 5)$.

The interval $[0, N - 1]$ is covered by the sub-intervals I_j for $j = 0, 1, 2, \ldots, k - 1$, where I_j denotes the interval $[jM, (j + 1)M)$, and kM is the smallest multiple of M that is greater than N.

Let $\mathcal{N} := \{n : \ 0 \leq n \leq N - 1 \text{ and } a + nd \text{ is a square}\}$, where the arithmetic progression is chosen so that $|\mathcal{N}| = Q(N)$. Let $\mathcal{N}_j = \{n \in \mathcal{N} : \ n \in I_j\}$ for each integer j. Let J be the set of integers j for which $\mathcal{N}_j$ has six or more elements.

Now if $n_1 < n_2 < \ldots < n_6$ all belong to $\mathcal{N}_j$, write $x = a/d + n_1$ and $b_i = n_i - n_1$ for $i = 1, \ldots, 6$, so that

$$b_1 = 0 < b_2 < \ldots < b_6 \leq M - 1$$

and each $x + b_i = a/d + n_i = (a + n_i d)/d$, which implies that

$$(x + b_1)(x + b_2) \cdots (x + b_6) = \frac{(a + n_1 d)(a + n_2 d) \cdots (a + n_6 d)}{d^6}$$

is the square of a rational number.

This gives rise to one of the $B(M)$ solutions counted above, and all the solutions created in this way are distinct (since given $x, d, b_1, \ldots, b_6$ we have each $a + n_j d = d(x + b_j)$). Therefore the set $\mathcal{N}_j$ gives rise to $\binom{|\mathcal{N}_j|}{6}$ such solutions, and so in total we have

$$\sum_{j \in J} \binom{|\mathcal{N}_j|}{6} \leq B(M).$$

It is easy to verify that $r \leq 5 + \binom{r}{6}$ for all integers $r \geq 1$, and so

$$Q(N) = |\mathcal{N}| = \sum_{j=0}^{k-1} |\mathcal{N}_j| \leq \sum_{j=0}^{k-1} 5 + \sum_{j \in J} \binom{|\mathcal{N}_j|}{6} \leq 5k + B(M),$$

as $|\mathcal{N}_j| \le 5$ if $j \notin J$. Finally, as $k \le N/M + 1$ we have

$$Q(N) \le 5k + B(M) \le \frac{5N}{M} + (B(M) + 5) \le \frac{6N}{M} < \delta N,$$

as desired. ∎

Bombieri [**2**] went on, together with Granville and Pintz, to combine these two proofs (along with much more arithmetic geometry machinery), to prove that

$$Q(N) < N^c$$

for any $c > \frac{2}{3}$, for sufficiently large N. Bombieri and Zannier [**4**] improved this to $c > \frac{3}{5}$ with a rather simpler proof. The conjecture that $Q(N)$ behaves more like a constant times $N^{1/2}$ remains open.

REFERENCES

1. L. Alpoge, van der Waerden and the primes, *Amer. Math. Monthly* **122** (2015), 784–785.
2. E. Bombieri, A. Granville, and J. Pintz, Squares in arithmetic progressions, *Duke Math. J.* **66** (1992), 369–385.
3. E. Bombieri and W. Gubler, *Heights in Diophantine Geometry*, New Mathematical Monographs, Vol. 4, Cambridge Univ. Press, Cambridge 2006.
4. E. Bombieri and U. Zannier, A note on squares in arithmetic progressions. II, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* **13** (2002), 69–75.
5. W.T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.
6. J.H. Silverman, *The arithmetic of elliptic curves*, Springer Verlag, New York, 1986.
7. E. Szemerédi, The number of squares in an arithmetic progression, *Studia Sci. Math. Hungar.* **9** (1974), 417.
8. E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.
9. T. Tao and Van Vu, *Additive Combinatorics*, Cambridge Univ. Press, Cambridge, 2006.

ANDREW GRANVILLE *Département de mathématiques et de statistique, Université de Montréal, CP 6128 succ. Centre-Ville, Montréal, QC H3C 3J7, Canada*
andrew@dms.umontreal.ca
and
Department of Mathematics, University College London, Gower Street, London, WC1E 6BT, United Kingdom
a.granville@ucl.ac.uk