ACTA ARITHMETICA
163.4 (2014)

Product sets cannot contain long arithmetic progressions

by

Dmitrii Zhelezov (Gothenburg)

1. Introduction. Sum-product estimates are among the most important questions in modern additive combinatorics. In general, one wants to show that if there is enough additive structure in a set A (for example if it has small doubling constant $|A+A|/|A|$), then the *product set* $A.A = \{aa' \mid a, a' \in A\}$ is large. The most famous conjecture in this area, posed by Erdős and Szemerédi [4], says that for any set A of complex numbers,

$$\max(|A.A|, |A+A|) \geq c|A|^{2-\epsilon}$$

for arbitrary $\epsilon > 0$ and some $c > 0$ that may depend on ϵ. The state of the art exponent $4/3 - o(1)$ was obtained by Solymosi in a very elegant way [9]. It is worth noting that each new bound for the exponent required a substantial new idea and attracted considerable attention from experts in the field.

In this note we investigate a different sort of relationship between the additive structure and the size of a product set. Namely, we show that a product set cannot contain extremely long arithmetic progressions. The result is the following.

Theorem 1. *Suppose that B is a set of n natural numbers. Then the longest arithmetic progression in $B.B$ has length at most $O(n\log^2 n/\log\log n)$.*

A lower bound is provided by

Theorem 2. *Given an integer $n > 0$ there is a set B of n natural numbers such that $B.B$ contains an arithmetic progression of length $\Omega(n\log n)$.*

In the fourth section of this note we will extend Theorem 1 to sets of complex numbers, but with a considerably weaker bound $O(n^{3/2})$.

2010 *Mathematics Subject Classification*: Primary 11B25.
Key words and phrases: product sets, arithmetic progressions, convex sets.

DOI: 10.4064/aa163-4-1

2. Notation. Let $f, g : \mathbb{N} \to \mathbb{R}_+$. The following standard notation will be used:

- $f(n) = O(g(n))$ means that $\limsup_{n\to\infty} f(n)/g(n) < \infty$.
- $f(n) = \Omega(g(n))$ means that $g(n) = O(f(n))$.
- Let H be a fixed graph. Then $\mathrm{ex}(n, H)$ denotes the maximal number of edges among all graphs with n vertices which do not contain H as a subgraph. In particular, $\mathrm{ex}(n, C_k)$ denotes the maximal number of edges in a graph with n vertices and no cycles of length k.
- Let p be a prime. Then $d = \mathrm{ord}_p(n)$ denotes the maximal power of p such that $p^d \mid n$.

3. Main result. Let $A = \{r + di\}$, $i = 0, \ldots, N$, be an arithmetic progression in the product set $B.B$ of a set B of size n. We start with the observation that by taking absolute values of B the longest arithmetic progression in $B.B$ can be shortened by a factor of at most two, so we may assume that all elements in B are positive.

We proceed with the following technical lemma.

Lemma 1. *We may assume that* $A = \{D(r' + d'i)\}$ *for some* $D > 0$ *such that* $\gcd(d', Dr') = 1$.

Proof. Let p be a prime such that $\mathrm{ord}_p(d) > \mathrm{ord}_p(r)$. If there is no such p then $D = \gcd(r, d), d' = d/D, r' = r/D$ provides the desired factorization. If $k' = \mathrm{ord}_p(r) = 1$ then every number in A is a product $b_i b_j$ such that $p \mid b_i$ but $p \nmid b_j$ and thus we can reduce B to

$$B' = \{b_i \mid b_i \in B,\, p \nmid b_i\} \cup \{b_i/p \mid b_i \in B,\, p \mid b_i\}$$

and iterate the lemma again.

So, now we assume that $k = \mathrm{ord}_p(d) > k' > 1$. We divide B into three sets B_1, B_2, B_3 such that $b_i \in B_1$ if $p \nmid b_i$, $b_i \in B_2$ if $0 < \mathrm{ord}_p(b_i) < k'$ and finally $b_i \in B_3$ if $p^{k'} \mid b_i$. Since $\mathrm{ord}_p(d) > k'$ for every $a \in A$ we have $\mathrm{ord}_p(a) = k'$ and a can be either a product of two numbers from B_2 or a product $b_1 b_3$ where $b_1 \in B_1$ and $b_3 \in B_3$. Thus, we can reduce B to

$$B' = \{b_i \mid b_i \in B_1\} \cup \{b_i/p \mid b_i \in B_2\} \cup \{b_i/p^2 \mid b_i \in B_3\}$$

such that $B'.B'$ contains an arithmetic progression A/p^2 of the same length as A, and then iterate the lemma. ∎

From now on we will assume the factorization $A = \{D(r + di)\}$ such that $\gcd(Dr, d) = 1$. By N we will always denote the length of A and $a_k = D(r + dk)$ will be the kth element of A (if not stated explicitly).

Lemma 2. *For* $i \neq j$,

$$\gcd(a_i, a_j) \le DN.$$

Proof. For $i > j$ we have

$$\begin{aligned}\gcd(a_i, a_j) &= \gcd(a_i - a_j, a_j) = \gcd(Dd(i-j), D(di+r)) \\ &= D\gcd(d(i-j), di+r) = D\gcd(i-j, di+r) \le DN.\end{aligned}$$

The last equality follows from $\gcd(d, Dr) = 1$. ∎

Let us fix a single pair $b_i, b_j \in B$ for each $a \in A$ such that $b_i b_j = a$ and make a graph G with $b \in B$ as vertices, such that for every $a \in A$ there is a unique edge between b_i and b_j which has been previously fixed for such a (for each edge we can simply take the first representation of a in lexicographical order). We will have $n = |B| = |V(G)|$ and $N = |A| = |E(G)|$. It turns out that our further analysis significantly simplifies if G is simple (without loops) and bipartite. However, we can always achieve this sacrificing just a constant factor by simply taking two copies of B, say B_1 and B_2, that are going to be the color classes of G, such that for each edge $e = \langle b_i, b_j\rangle \in G$, $i \le j$, we place an edge between $b_i \in B_1$ and $b_j \in B_2$, so the resulting graph is bipartite and simple.

As we will see from our example, which provides a lower bound $N = \Omega(n \log n)$, it is safe to assume $N > 2n$, a very weak yet convenient bound, as it guarantees, for example, that G contains a cycle.

LEMMA 3. *If G contains an even cycle of size $2k$, then $r \le N^k$ and $d \le N^k$.*

Proof. Let $C = b_1 \ldots b_{2k}$ be a simple cycle in G of length $2k \le n$, so $b_i b_{i+1} \in A$, $i = 1, \ldots, 2k$ (hereafter we assume addition of indices modulo $2k$). By simple algebra we have

$$b_{2k} b_1 = \frac{b_1 b_2}{b_2 b_3} \frac{b_3 b_4}{b_4 b_5} \cdots \frac{b_{2k-3} b_{2k-2}}{b_{2k-2} b_{2k-1}} b_{2k-1} b_{2k}, \tag{1}$$

and since for each i there is some j such that $b_i b_{i+1} = D(r + j_i d)$ we can rewrite (1) as

$$\prod_{i=1}^{k} (r + j_{2i} d) = \prod_{i=1}^{k} (r + j_{2i-1} d), \tag{2}$$

where all j_i are distinct (since for every $a \in A$ we have chosen only a single representation). Expanding the brackets, we obtain the equation

$$c_0 r^k + c_1 r^{k-1} d + \cdots + c_{k-1} r d^{k-1} + c_k d^k = 0 \tag{3}$$

for integer coefficients c_i which depend only on indices j. First, let us note that it cannot happen that all c_i are zero since then (2) would hold for *any* r, d, which contradicts the fact that all js are distinct. Let l and m be respectively the smallest and largest indices such that $c_l, c_m \neq 0$. Obviously,

$l < m$ and dividing (3) by $r^l d^{k-m}$ we arrive at

$$c_l r^{m-l} + \cdots + c_m d^{m-l} = 0. \tag{4}$$

Since r and d are coprime, $r \mid c_m$ and $d \mid c_l$ (all the terms in the middle are divisible by rd), and the claim of the lemma follows if the bound $c_i \le N^k$ holds for all coefficients. But on the other hand, c_t is a sum of $2\binom{k}{t}$ t-fold products of js. Since each index j is less than N, for $t \le k/2$ we have

$$c_t \le 2k^t N^t < n^t N^t < N^k,$$

and analogously, for $t \ge k/2$,

$$c_t \le 2k^{k-t} N^t < n^{k-t} N^t < N^k.$$

Here we used the trivial bound $2k \le n$. ∎

LEMMA 4. *If $d < N^k$, $r < N^k$, $3^k < N/9$ then $N \le 36kn \log n$ for sufficiently large n.*

Proof. Suppose for contradiction that $N > 36(k+1)n \log n$. Let $p_1, \ldots, p_K$ be primes such that $N/3 < p_i < N/2$ and $p_i \nmid d$. By the Prime Number Theorem there are more than $N/(6 \log N) > 3(k+1)n$ primes in $[N/3, N/2]$ (1) (for N large enough) and at most k of them may divide d (since $d < N^k$ and $3^{k+1} < N$), so $K > 3(k+1)n$.

Recall the graph G with $b \in B$ as vertices and edges that correspond to the relation $b_i b_j \in A$, with each representation of $a \in A$ being unique. Let us call an edge of G *regular* if

$$\gcd(b_i b_j / D, p_1 \ldots p_K) = 1,$$

or, in words, if $b_i b_j$ does not have any additional power of the aforementioned $p_1, \ldots, p_K$ in its prime decomposition. Otherwise, if $\operatorname{ord}_p(b_i b_j) > \operatorname{ord}_p(D)$ let us call an edge (b_i, b_j) *p-irregular*. Further, by an "irregular edge", we mean an edge that is p-irregular for at least one $p \in \{p_1, \ldots, p_K\}$. Note that it can be irregular for some primes, but regular with respect to others.

Let $p \in P_K = \{p_1, \ldots, p_K\}$. Since $p \nmid d$, dj covers the full system of residues modulo p when j goes from 0 to N. Hence, since $p \in [N/3, N/2]$, there are either two or three indices j such that $p \mid dj + r$, and thus two or three p-irregular edges in G.

By the pigeonhole principle, we can pick a set S of at least $n+1$ distinct irregular edges such that for every $p \in P_K$ there is at most one p-irregular edge in S. Indeed, every element in A can have at most $k+1$ divisors in P_K (due to the bounds $d < N^k, r < N^k$ we have $r + id < N^{k+1}$ for $0 \le i \le n$). On the other hand, each $p \in P_K$ divides at most three elements in A.

(1) This is the only place where we use the technical bound $3^k < N/9$, but as we will see later, this restriction does not affect the final bound, as k is going to be $o(\log n)$.

The next step is to clean up our original graph G by removing all edges except those not in S. We will refer to the resulting graph as G'. Of course, it is simple and bipartite as was G. Now we claim that it contains no cycles. Indeed, let e_p be a (unique) p-irregular edge in G' and $e_p = a_1 \dots a_{2l}$ be a cycle it lies on (of course, here indices of a's indicate just the ordering in the cycle, not in A). Note that now we write the cycle as a set of edges rather than vertices, meaning that $a_i \in A$ and each a_i is a product of two consecutive vertices of the cycle. Thus, arguing exactly as in Lemma 3 it is easy to see that

$$\prod_{i\,\mathrm{odd}} a_i = \prod_{i\,\mathrm{even}} a_i.$$

But this cannot happen. Indeed, for each $a_i \neq e_p = a_1$ we have $\mathrm{ord}_p(a_i) = \mathrm{ord}_p(D)$ since e_p is the only p-irregular edge in G', and the p-order of the RHS is strictly less than that of the LHS. Thus, G' cannot contain more than n edges, a contradiction. ∎

Putting it all together, we obtain the main result of this note.

Proof of Theorem 1. If G does not contain even cycles of length up to $2k$ the result of Bondy and Simonovits [1] from extremal combinatorics gives

$$N \leq \mathrm{ex}(n, C_{2k}) < 100kn^{1+1/k}. \tag{5}$$

But otherwise Lemmas 3 and 4 apply and we obtain $N \ll (k+1)n\log n$, so finally we have

$$N \leq O(\max\{kn^{1+1/k}, kn\log n\}).$$

This can be optimized by taking $k = \log n/\log\log n$, which gives the desired bound $N = O(n\log^2 n/\log\log n)$. ∎

Now we present a construction for the lower bound of Theorem 2.

Proof of Theorem 2. Consider a set B which consists of all natural numbers from 1 to n plus all primes in the interval $[n, \lfloor n\log n\rfloor]$. By the Prime Number theorem, $|B| \leq 2n$ for large n and $B.B$ contains all natural numbers in the interval $[1, \lfloor n\log n\rfloor]$ which is an arithmetic progression of size $\Omega(n\log n)$.

Indeed, suppose $x \in [n, \lfloor n\log n\rfloor]$. If the maximal prime p that divides x is greater than $\log n$, then $x/p \leq n$ and $x = p\cdot\frac{x}{p}$ is clearly in $B.B$, since all primes in the interval $[1, \lfloor n\log n\rfloor]$ are in B. Otherwise, we run the following algorithm. Let p_1 be an arbitrary prime divisor of x and set $d_1 = p_1$, $d_2 = x/p_1$. Then choose the smallest prime divisor p' of d_2, set $d_1 := d_1p'$, $d_2 := d_2/p'$ and iterate this procedure until $d_2 = 1$. If there is a moment when both $d_1, d_2 \leq n$ then of course $x \in B.B$ and we are done. Otherwise, at some step $d_1 < n$, $d_2 > n$, but $d_1p' > n$, $d_2/p' < n$. But since every prime divisor of x is less than $\log n$ we have

$$x = d_1 d_2 \geq n^2/\log n,$$

which contradicts $x \in [n, \lfloor n \log n \rfloor]$. ∎

4. The case of complex numbers

Theorem 3. *Suppose that B is a set of n complex numbers. Then the longest arithmetic progression in $B.B$ has length at most $O(n^{3/2})$.*

Our strategy will be to show that if $B.B$ contains an arithmetic progression A of size $\Omega(n^{3/2})$ then in fact one can take a new set B' of only rational numbers, perhaps twice as big as the original set B, such that $B'.B'$ contains a progression of the same length. Unfortunately, we can prove that such a reduction exists only if the arithmetic progression A in the original set has length at least $\Omega(n^{3/2})$, so the resulting bound is much weaker than what Theorem 1 gives for sets of natural numbers.

So let $A = \{r + di\}$ be an arithmetic progression of length N in $B.B$. The first step is to scale A by simply dividing each element in B by $\sqrt{d}$, and from now on we will assume that $A = \{r + i\}$.

Recall the graph G which provides a one-to-one correspondence between elements of A and its edges, namely an edge $e_a = \langle b_i, b_j \rangle$ corresponds to the element $a = b_i b_j$.

Lemma 5. *If G contains a 4-cycle then r is rational and so are all elements of $A = \{r + i\}$.*

Proof. Let $\langle b_1 b_2 b_3 b_4 \rangle$ be a 4-cycle in G. Then both $b_1(b_2 - b_4)$ and $b_3(b_2 - b_4)$ are non-zero integers as they are differences of two distinct elements of A. Thus, b_1/b_3 is rational, and so is $q = b_1 b_2 / b_2 b_3 \neq 1$. On the other hand, writing $b_1 b_2 = r + i_1$ and $b_2 b_3 = r + i_2$, we have

$$\frac{r + i_1}{r + i_2} = q,$$

so

$$r = \frac{i_1 - q i_2}{q - 1}$$

is rational since i_1, i_2 are integers. ∎

Corollary 1. *If $A = \{r + i\}$ is contained in a product set $B.B$ with $|B| = n$ and $|A| > n^{3/2}$ then it consists of rational numbers.*

Proof. The claim follows from the well-known fact that a graph with more than $n^{3/2}$ edges contains a 4-cycle ([2]) together with Lemma 5. ∎

([2]) In fact, $\mathrm{ex}(n, C_4) \leq \frac{n}{4}(1 + \sqrt{4n - 3})$ (see [8]).

While the condition that all elements in A are rational is strong, it still does not guarantee that elements in B are rational as well, so some additional tweaks are needed in order to invoke Theorem 1. We will construct a slightly different set B' of only rational numbers such that $B'.B'$ contains A. Our main observation is the following.

LEMMA 6. *Assume A consists of rational numbers. Then if b_i and b_j are connected in G by a path of even length, the quotient b_i/b_j is rational. If they are connected by a path of odd length, the product b_ib_j is rational.*

Proof. Indeed, if there is a path $L = \langle b_i b_{i+1} \dots b_{i+2k+1} = b_j\rangle$ of even length we have

$$\frac{b_i}{b_j} = \frac{(b_ib_{i+1})(b_{i+2}b_{i+3})\dots(b_{i+2k-1}b_{i+2k})}{(b_{i+1}b_{i+2})\dots(b_{i+2k}b_{i+2k+1})}, \tag{6}$$

which is rational. The second claim follows in exactly the same way. ∎

Our next step is to make elements in B rational while preserving the property that A is contained in $B.B$. Remember that from the very beginning we assume our graph G is simple bipartite (which one can always do without loss of generality).

LEMMA 7. *Let A be a subset of $B.B$ consisting of only rational numbers and suppose the corresponding incidence graph G is bipartite. Then there is a set B' of rational numbers of size $|B|$ such that $A \subset B'.B'$.*

Proof. Let $K_1, \dots, K_l$ be the connected components of the bipartite graph G. We will treat them separately one by one. So let K be one of the components. As K does not contain odd cycles, we can color its vertices black and white so that there are edges only between white and black vertices.

By Lemma 6 the quotient b_i/b_j is rational for the vertices of the same color, and so is the product of any two vertices of different color. Thus, we can take an arbitrary white element b_w from K and modify our set B as follows:

- For all white $b \in K$ set $b := b/b_w$.
- For all black $b \in K$ set $b := bb_w$.

As K is bipartite, this procedure will keep the set A unchanged. On the other hand, it makes all the elements in K rational.

Iterating the procedure above for all components, we finally obtain the set B' with the desired properties. ∎

Proof of Theorem 3. Now the theorem follows as an immediate corollary of Corollary 1 and Theorem 1 since multiplying our new set B' by a sufficiently composite number we obtain a set of integers whose product set contains an arithmetic progression of the same length. It remains to note that by taking absolute values of B' the longest arithmetic progression in $B'.B'$ can be shortened by a factor of at most two. ∎

5. Discussion. The motivation for asking how long an arithmetic progression in a product set can be stems from the question of Hegarty [6].

QUESTION 1. *Let B be a set of n integers and let A be a strictly convex* (*or strictly concave*) *subset of $B+B$. Must $|A|$ be $o(n^2)$?*

Recall that a sequence of numbers $A = \{a_1, \ldots, a_n\}$ is called *strictly convex* (resp. *concave*) if the consecutive differences $a_i - a_{i-1}$ are strictly increasing (resp. decreasing).

It is not difficult to see that it does not matter whether the numbers in Question 1 are reals or integers. Now suppose that $B = \{\log b'_i\}$ for some b'_i, so $B+B = \{\log(b'_i b'_j)\}$. If $B'.B' = \{b'_i b'_j\}$ contains a long arithmetic progression, we immediately obtain a convex set of the same size in $B+B$. If we assume that b'_i are natural numbers then Theorem 1 shows that the longest convex set we can possibly get in this way is of size $O(n^{1+o(1)})$. Apart from Hegarty's original inquiry, we now ask the following question that might be simpler.

QUESTION 2. *Can one construct an example of a set of size n such that the sumset $B+B$ contains a convex* (*or concave*) *set of size $n^{1+\delta}$ for some $\delta > 0$ and arbitrarily large n?*

REMARK. Erdős and Newman [2] gave an example of a set B of size $n/\log^M n$ such that $B+B$ covers $\{1, 2^2, \ldots, n^2\}$ for arbitrary $M > 0$, which is better than our construction above, but still this lower bound is very weak.

REMARK. Erdős and Pomerance [3] asked if it is true that for a large enough c, every interval of length cn contains a number divisible by precisely one prime in $(n/2, n]$. While the question remains open, a positive answer would give an essentially sharp upper bound $O(n \log n)$ for Theorem 1.

An obvious direction of research is to match the bound for the case of complex numbers to the one of Theorem 1. Moreover, we believe that the lower bound $O(n \log n)$ is sharp for Theorem 1 and perhaps for Theorem 3 as well.

Another interesting twist is to ask the question of the current note for subsets of finite fields $\mathbb{F}_p$. By a recent result of Grosu [5], the bound of Theorem 3 translates to subsets $B \subset \mathbb{F}_p$ of size $O(\log\log\log p)$. While there are sets B of size $O(\sqrt{p})$ such that $B.B$ covers the whole field $\mathbb{F}_p$ and thus contains an AP of size $\Omega(|B|^2)$, we conjecture that for smaller sets the bound $|B|^{1+o(1)}$ holds.

Conjecture 1. *There is an absolute constant $c > 0$ such that for any $B \subset \mathbb{F}_p$ with $|B| < c\sqrt{p}$ the product set $B.B$ contains no arithmetic progression of size greater than $|B|^{1+o(1)}$. Here we assume p and $|B|$ are large.*

A lot of related questions arise if we continue the general idea of asking how large a set with additive structure can be if it is contained in a product set. For example, instead of arithmetic progressions one may ask about generalized arithmetic progressions or just sumsets of an arbitrary set.

Acknowledgements. I am grateful to my supervisor Professor Peter Hegarty for helpful discussions and constant support. I also thank Boris Bukh for comments during the poster session at the Erdős 100 conference in Budapest which helped to improve the exposition. Finally, I thank the anonymous referee for valuable comments and especially for pointing out that applying the result of Bondy and Simonovits [1] to our case actually gives a better bound than a more recent result of Lam and Verstraëte [7].

References

[1] J. A. Bondy and M. Simonovits, *Cycles of even length in graphs*, J. Combin. Theory Ser. B 16 (1974), 97–105.

[2] P. Erdős and D. J. Newman, *Bases for sets of integers*, J. Number Theory 9 (1977), 420–425.

[3] P. Erdős and C. Pomerance, *Matching the natural numbers up to n with distinct multiples in another interval*, Nederl. Akad. Wetensch. Proc. Ser. A 83 (1980), 147–161.

[4] P. Erdős and E. Szemerédi, *Sums and products of integers*, in: Studies in Pure Mathematics, Birkhäuser, 1983, 213–218.

[5] C. Grosu, *$\mathbb{F}_p$ is locally like $\mathbb{C}$*, J. London Math. Soc. (2014) (online).

[6] P. Hegarty, http://mathoverflow.net/questions/106817/convex-subsets-of-sumsets.

[7] T. Lam and J. Verstraëte, *A note on graphs without short even cycles*, Electron. J. Combin. 12 (2005), note 5, 6 pp.

[8] I. Reiman, *Über ein Problem von K. Zarankiewicz*, Acta Math. Acad. Sci. Hungar. 9 (1958), 269–273.

[9] J. Solymosi, *Bounding multiplicative energy by the sumset*, Adv. Math. 222 (2009), 402–408.

Dmitrii Zhelezov
Department of Mathematical Sciences
Chalmers University of Technology
and
University of Gothenburg
41296 Gothenburg, Sweden
E-mail: zhelezov@chalmers.se

Received on 27.3.2013
and in revised form on 19.5.2014 (7387)