ON $A \subseteq [N]$ SUCH THAT $ab+1$ IS NEVER SQUAREFREE FOR $a, b \in A$

MEHTAAB SAWHNEY

1. Statement of result

In this short note, we address the following question of Erdős [2]. We remark that substantial progress towards this problem was made by van Doorn, Weisenberg and Cambie [1].

Proposition 1.1. *There exists an integer N_0 such that if $N \geq N_0$ then the following holds. Let $A \subseteq [N]$ with the property that if $a, b \in A$ then $ab+1$ is not squarefree. Then $|A| \leq |\{n \in [N] : n \equiv 7 \bmod 25\}|$.*

Remark. Equality is achieved precisely when $A = \{n \in [N] : n \equiv 7 \bmod 25\}$ or (possibly) $A = \{n \in [N] : n \equiv 18 \bmod 25\}$. The proof in fact gives that there exists $\eta > 0$ such that if $|A| \geq (1/25 - \eta) \cdot N$ and N is sufficiently large then A is contained in either $\{n \in [N] : n \equiv 7 \bmod 25\}$ or $\{n \in [N] : n \equiv 18 \bmod 25\}$.

It may be of interest to explore further whether one can give a strong structural classification of sets A with size δN where δ is a fixed small constant.

2. Preliminary lemmas

We now require the following pair of preliminary lemmas; the first was already used by van Doorn (see [1]). In order to be self-contained, we provide proofs with suboptimal error terms.

Lemma 2.1. *Let $\mathcal{P}$ be a subset of primes with $\max \mathcal{P} \leq N^{1/2}$ and let q be a positive integer. For each $p \in \mathcal{P}$, define $\mathcal{R}_p$ to denote a set of residue classes modulo p^2 with $|\mathcal{R}_p| \leq 2$ and $\mathcal{R}_p = \emptyset$ if $(p, q) \neq 1$. Then*

$$\left|\left|\Big(\big\{n \in [N] : n \equiv t \mod q\big\} \bigcap \bigcup_{p \in \mathcal{P}} \{n \bmod p^2 \in \mathcal{R}_p\}\Big)\right| - \frac{N}{q} \cdot \Big(1 - \prod_{p \in \mathcal{P}} \Big(1 - \frac{|\mathcal{R}_p|}{p^2}\Big)\Big)\right| \ll N(\log N)^{-1/2}.$$

Proof. Let $T = \lfloor \sqrt{\log N} \rfloor$. Note that

$$\Big|[N] \bigcap \bigcup_{\substack{p \in \mathcal{P} \\ T \leq p \leq N^{1/2}}} \{n \bmod p^2 \in \mathcal{R}_p\}\Big| \ll \sum_{T \leq p \leq N^{1/2}} \frac{N}{p^2} \ll \frac{N}{T}.$$

Therefore it suffices to estimate

$$\left|\Big(\big\{n \in [N] : n \equiv t \mod q\big\} \bigcap \bigcup_{\substack{p \in \mathcal{P} \\ p \leq T}} \{n \bmod p^2 \in \mathcal{R}_p\}\Big)\right|.$$

Note that as $T \leq \sqrt{\log N}$, we have that $\prod_{p \leq T} p^2 \leq N^{o(1)}$. Via inclusion exclusion, we have

$$\begin{aligned}
&\left|\Big(\big\{n \in [N] : n \equiv t \mod q\big\} \bigcap \bigcup_{\substack{p \in \mathcal{P} \\ p \leq T}} \{n \bmod p^2 \in \mathcal{R}_p\}\Big)\right| \\
&\qquad = \sum_{\substack{S \subseteq \mathcal{P} \cap [T] \\ S \neq \emptyset}} (-1)^{|S|-1} \cdot \Big|\bigcap_{p \in S} \{n \bmod p^2 \in \mathcal{R}_p\} \cap \{n \in [N] : n \equiv t \bmod q\}\Big| \\
&\qquad = N^{o(1)} + \sum_{\substack{S \subseteq \mathcal{P} \cap [T] \\ S \neq \emptyset}} (-1)^{|S|-1} \cdot \frac{N}{q} \cdot \frac{\prod_{p \in S} |\mathcal{R}_p|}{\prod_{p \in S} p^2}
\end{aligned}$$

This research was conducted during the period that the author served as a Clay Research Fellow. The author thanks Mark Sellke for useful comments.

$$= \frac{N}{q} \cdot \Big(1 - \prod_{p \in \mathcal{P} \cap [T]} \Big(1 - \frac{|\mathcal{R}_p|}{p^2}\Big)\Big) + N^{o(1)}.$$

Completing the product, we have the desired result. □

We next require a variant of the previous lemma which allows us to incorporate the "off-diagonal" constraints that $ab+1$ is not squarefree.

Lemma 2.2. *Fix a positive integer N, residue class $t \bmod q$ with q a perfect square and an integer b such that $1 \le b \le N$. Suppose that there does not exist a prime p such that $p^2|q$ and $p^2|(bt+1)$.*

Then we have that

$$\Big|\Big|\{a \in [N] : a \equiv t \bmod q \wedge \mu(ab+1) = 0\}\Big| - \frac{N}{q} \cdot \Big(1 - \prod_{\substack{p \\ (p,qb)=1}} \Big(1 - \frac{1}{p^2}\Big)\Big)\Big| \ll \frac{N}{\sqrt{\log N}}.$$

Proof. Note that $ab + 1 \le N^2 + 1$. Therefore if $\mu(ab+1) = 0$, there exists a prime p such that $p^2|(ab+1)$. Additionally note that $(p, qb) = 1$. In particular, if $p|b$ then $p \nmid ab+1$. Furthermore if $p|q$, then observe $ab + 1 \equiv bt + 1 \not\equiv 0 \bmod p^2$ by the imposed condition.

We now let $T = \lfloor \sqrt{\log N} \rfloor$. We have that

$$\sum_{T \le p \le N} |\{a \in [N] : p^2|(ab+1)\}| \ll \sum_{T \le p \le N} \Big(\frac{N}{p^2} + 1\Big) \ll \frac{N}{T}.$$

Therefore it suffices to handle $p \le T$ with $(p, qb) = 1$. The condition that $p^2|ab+1$ gives one specified residue class (depending on each prime p) and therefore we may handle the remaining primes via Lemma 2.1 and we may conclude the proof. □

3. Completing the proof of Proposition 1.1

We now give the remainder of the analysis to conclude the proof. The proof consists of breaking the set into the parts which are 7 mod 25, 18 mod 25 and the remainder. We then apply combinations of Lemma 2.1 and Lemma 2.2 based on whether the relevant sets are nonempty. To optimize the numerical factor of $\prod_{(p,qb)=1}(1 - 1/p^2)$, we introduce casework based on whether these sets contain an even integer. We note that there exist various solutions within this framework; indeed the initial version of the solution required more extensive casework modulo numbers including 169 and 289.

Proof of Proposition 1.1. Let $\eta > 0$ be a small absolute constant, let N be sufficiently large and suppose that $|A| \ge (1/25 - \eta) \cdot N$. Define

$$A_7 = \{a \in A : a \equiv 7 \bmod 25\}$$
$$A_{18} = \{a \in A : a \equiv 18 \bmod 25\}$$
$$A^* = A \setminus (A_7 \cup A_{18}).$$

We first prove that A^* is empty.

Suppose that there exists $b \in A^*$ with $2|b$. Then for each $x \in A^*$, there is $p \equiv 1 \bmod 4$ with $p \ge 13$ such that $p^2|x^2+1$. By breaking A^* into residue classes modulo 25 (of which there are 23 possibilities) and applying Lemma 2.1, we have that

$$\frac{|A^*|}{N} \le \Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p \equiv 1 \bmod 4 \\ p \ge 13}} \Big(1 - \frac{2}{p^2}\Big)\Big) + o(1).$$

For $a \in A_7 \cup A_{18}$, we have that there exists $p \ne 2, 5$ such that $p^2|ab+1$. This implies by Lemma 2.2 that

$$\frac{|A_7 \cup A_{18}|}{N} \le \Big(\frac{2}{25}\Big)\Big(1 - \prod_{p \ne 2,5} \Big(1 - \frac{1}{p^2}\Big)\Big) + o(1).$$

Combining these bounds we have that

$$\frac{|A|}{N} \le \Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p\equiv 1 \bmod 4\\ p\ge 13}}\Big(1-\frac{2}{p^2}\Big)\Big) + \frac{2}{25}\Big(1-\prod_{p\ne 2,5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1)$$
$$\le 0.0252 + 0.0125 = 0.0377;$$

here we have absorbed the $o(1)$ error assuming that N is sufficiently large.

Therefore we have reduced to the case that A^* consists of odd elements. By breaking A^* into residue classes modulo 50 and noting there are 23 valid possibilities and applying Lemma 2.1, we have that

$$\frac{|A^*|}{N} \le \Big(\frac{1}{2}\Big)\Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p\equiv 1 \bmod 4\\ p\ge 13}}\Big(1-\frac{2}{p^2}\Big)\Big) + o(1).$$

We now split into cases based on whether or not $A_7 \cup A_{18}$ contains even elements. First suppose that no element of $A_7 \cup A_{18}$ is even; then $A_7 \subseteq \{7, 57\} \bmod 100$ and $A_{18} \subseteq \{43, 93\} \bmod 100$. Fixing $b \in A^*$, for each of A_7 and A_{18} one of the residue classes $\bmod 100$ will have 2^2 as a divisor and the other will not. Therefore applying Lemma 2.2 to these progressions $\bmod 100$, we have that

$$\frac{|A_7 \cup A_{18}|}{N} \le \frac{1}{50} + \Big(\frac{1}{50}\Big)\Big(1-\prod_{p\ne 2,5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1).$$

Therefore we have a bound of

$$\frac{|A|}{N} \le \Big(\frac{1}{2}\Big)\Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p\equiv 1 \bmod 4\\ p\ge 13}}\Big(1-\frac{2}{p^2}\Big)\Big) + \frac{1}{50} + \Big(\frac{1}{50}\Big)\Big(1-\prod_{p\ne 2,5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1)$$
$$\le 0.0126 + 0.0200 + 0.0032 = 0.0358.$$

Therefore we have reduced to the case where A^* is nonempty, consisting of odd elements and at least one of A_7 or A_{18} has an even element. We now fix $b \in A^*$ and $b' \in A_7$ with $2|b'$ (with the case where A_{18} has the even element being completely analogous). The size of $|A^*|$ is bounded as before. By using b and Lemma 2.2, we have that

$$\frac{|A_7|}{N} \le \frac{1}{25}\Big(1-\prod_{p\ne 5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1).$$

Additionally, using $b' \in A_7$ and Lemma 2.2, we have that

$$\frac{|A_{18}|}{N} \le \frac{1}{25}\Big(1-\prod_{p\ne 2,5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1).$$

Therefore we have that

$$\frac{|A|}{N} \le \Big(\frac{1}{2}\Big)\Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p\equiv 1 \bmod 4\\ p\ge 13}}\Big(1-\frac{2}{p^2}\Big)\Big) + \frac{1}{25}\Big(1-\prod_{p\ne 5}\Big(1-\frac{1}{p^2}\Big)\Big) + \frac{1}{25}\Big(1-\prod_{p\ne 2,5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1)$$
$$\le 0.0126 + 0.0147 + 0.0063 = 0.0336.$$

Therefore we have now reduced to the case where $A^* = \emptyset$. Suppose that A_7 and A_{18} are nonempty; fixing $b \in A_7$ and $b' \in A_{18}$ and applying Lemma 2.2 twice we have that

$$\frac{|A|}{N} \le \frac{2}{25}\Big(1-\prod_{p\ne 5}\Big(1-\frac{1}{p^2}\Big)\Big) + o(1) \le 0.0294.$$

Therefore A is fully contained inside one of A_7 or A_{18}. This immediately gives the desired bound. $\square$

4. Discussion of the role of AI assistance

The proof of Proposition 1.1 was obtained with assistance from ChatGPT 5 Pro. The primary discussion with GPT is

`https://chatgpt.com/share/68ec50da-cf00-8005-b5f6-b683506e5853`.

The main additional lemma Lemma 2.2 over those on [1] was suggested by ChatGPT 5 Pro. The lemma is a slight variant of standard facts giving the density of square–free integers. GPT however crucially notes that one may exclude primes depending on b and on the common difference of the progression under consideration. GPT then attempts (in various elaborate and incorrect ways) to prove that one cannot have two residue classes modulo 25 with positive density; these attempts are seen to be incorrect by using an analogous example modulo 169. However it attempts naturally suggest using a single element in the alternate residue class to sieve A_7 and A_{18}. Additionally at various steps it treats $1 - \prod_{p\neq 5}(1-1/p^2) = (1 - 6/\pi^2 \cdot 25/24)$ as 0; this "mistake" was quite useful as it suggests pushing this quantity as close to zero as possible. Giving these lemmas, and supposing that A^* is nonempty, one gets a bound of the form

$$\Big(\frac{23}{25}\Big)\Big(1 - \prod_{\substack{p\equiv 1 \bmod 4\\ p\geq 13}}\Big(1-\frac{2}{p^2}\Big)\Big) + \frac{2}{25}\Big(1-\prod_{p\neq 5}\Big(1-\frac{1}{p^2}\Big)\Big) \leq 0.0252 + 0.0294 \approx 0.0546.$$

This is already quite close to the desired bound and incorporating further primes to push the numerical factor towards zero ultimately completes the proof. (A particularly encouraging subcase is assuming that A^* is empty and then carrying out the argument; an incorrect execution of this appears in 3.3 of ChatGPT's first response but the correct numerics are sufficient.) We remark that in the discussion/initial solution of the author the auxiliary prime 3 was used; that the prime 2 could be used was realized slightly later leading to a cleaner proof.

Numerical constants and verification

We record various constants in the argument to 4 digits; we round to preserve the stated inequalities.

- $\prod_p\Big(1-\frac{1}{p^2}\Big) = \frac{6}{\pi^2} \geq 0.6079$, so $1 - \frac{6}{\pi^2} \leq 0.3921$.
- $\prod_{p\neq 5}\Big(1-\frac{1}{p^2}\Big) = \frac{25}{4\pi^2} \geq 0.6332$, so $1 - \frac{25}{4\pi^2} \leq 0.3668$.
- $\prod_{p\neq 2,5}\Big(1-\frac{1}{p^2}\Big) = \frac{25}{3\pi^2} \geq 0.8443$, so $1 - \frac{25}{3\pi^2} \leq 0.1557$.
- $1 - \prod_{\substack{p\equiv 1\ (4)\\ p\geq 13}}\Big(1-\frac{2}{p^2}\Big) \leq 0.0274$.

References

[1] Thomas F. Bloom, `https://www.erdosproblems.com/go_to/848`. 1, 4

[2] Paul Erdős, *Some of my favourite problems in various branches of combinatorics*, vol. 47, 1992, Combinatorics 92 (Catania, 1992), pp. 231–240. 1

Department of Mathematics, Columbia University, New York, NY 10027

Email address: m.sawhney@columbia.edu