PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 149, Number 9, September 2021, Pages 3669–3682
https://doi.org/10.1090/proc/15511
Article electronically published on June 16, 2021

SETS WHOSE DIFFERENCES AVOID SQUARES MODULO m

KEVIN FORD AND MIKHAIL R. GABDULLIN

(Communicated by Amanda Folsom)

ABSTRACT. We prove that if $\varepsilon(m) \to 0$ arbitrarily slowly, then for almost all m and any $A \subset \mathbb{Z}_m$ such that $A - A$ does not contain non-zero quadratic residues we have $|A| \leqslant m^{1/2-\varepsilon(m)}$.

1. Introduction

Let $R_m = \{a^2 : a \in \mathbb{Z}_m\}$ be the set of quadratic residues modulo m. In this paper we find an upper bound for the sets $A \subset \mathbb{Z}_m$ with

$$(A - A) \cap R_m = \{0\}, \tag{1.1}$$

where $A - A = \{a - b : a, b \in A\}$, for a large set of m. This question originated from the corresponding problem in $\mathbb{Z}$: Ruzsa [R] constructed a set of integers $B \subset [1, N]$ such that $B - B$ avoids squares with $|B| \gg N^\gamma$, where $\gamma = \frac{1}{2}(1 + \frac{\log 7}{\log 65}) = 0.73\ldots$, and this construction was based on a 7-element subset of $\mathbb{Z}_{65}$ obeying (1.1) (see [Le] for further improvements). As for upper bounds in this integer setting, we just note that such a set B must obey $|B| = o(N)$ (see [Sa], [PSS], and also [BPPS]), but no bounds with power saving are known.

Now we begin a discussion of (1.1). In the case of a prime $m \equiv 3 \pmod 4$, $(\frac{-1}{m}) = -1$ and thus any set $A \subset \mathbb{Z}_m$ with (1.1) is a singleton or empty at all, and so the problem is trivial. In contrast, in the case of a prime $m = p \equiv 1 \pmod 4$ it should be very hard to obtain good bounds, since the problem is related to two other famous questions. The first one is the clique number problem for the Paley graph. Recall that the Paley graph is the graph $G_p = (V, E)$ with $V = \mathbb{Z}_p$ and $\{a, b\} \in E$ iff $a - b$ is a quadratic residue modulo p, and a clique of an undirected graph is a subset of its vertices such that every two distinct vertices in the clique are adjacent (that is, its induced subgraph is complete). The clique number of a graph is the size of its maximum clique. Fix any quadratic non-residue $\xi \in \mathbb{Z}_p \backslash R_p$; then $C \subset \mathbb{Z}_p$ is a clique of G_p if and only if ξC obeys (1.1), and so any bound for the clique number is a bound for our sets, and vice versa. While it is not hard to show that any clique in this graph (and, hence, any set $A \subset \mathbb{Z}_p$ with (1.1)) has size at most $p^{1/2}$ (see Section 2 for a short proof), any improvement of it requires non-trivial ideas; currently the best upper bound is $\sqrt{p/2}+1$ (see [HP]). The second

Received by the editors July 11, 2020, and, in revised form, January 9, 2021.

2020 *Mathematics Subject Classification.* Primary 11A15, 11P83, 11T23.

The first author was supported by National Science Foundation Grant DMS-1802139.

The work of the second author was performed at the Steklov International Mathematical Center and supported by the Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-15-2019-1614).

problem related to our sets is finding an upper bound for the least quadratic non-residue. If we denote it by $n(p)$, then the set $\xi \cdot \{1, \dots, n(p)\}$, where $\xi \in \mathbb{Z}_p \backslash R_p$ is again any quadratic non-residue, has size $n(p)$ and obeys (1.1), and so any bound for sets with (1.1) is a bound for $n(p)$, but the best we know is $n(p) \ll p^{1/4\sqrt{e}+o(1)}$, due to the work [Bu] (see also the classical papers [V] and [Li]).

We turn to the case of composite m. Matolcsi and Ruzsa [MR] proved that $|A| \leqslant m^{1/2}$ for all $A \subset \mathbb{Z}_m$ with (1.1) and square-free m which have prime divisors of the form $4k+1$ only, and that $|A| \leqslant m \exp(-c\sqrt{\log m})$ for all square-free m (throughout the paper we denote by c absolute positive constants which may vary from line to line). The second author [G] proved that for any square-free m and such A we have

$$|A| \leqslant \min\left\{ m^{1/2}(3\omega(m))^{3\omega(m)/2},\ m\exp\left(-\frac{c\log m}{\log\log m}\right)\right\},$$

where $\omega(m)$ is the number of prime divisors of m. Since $\omega(m) \leqslant 2\log\log m$ for almost all m (due to Hardy and Ramanujan [HR]) we deduce that $|A| \leqslant m^{1/2+o(1)}$ for almost all m (that is, for a set of density 1; here and in what follows we consider the lower asymptotic density of a set $M \subseteq \mathbb{N}$, which is defined as $\underline{\lim}_{N\to\infty} \frac{\#(M\cap\{1,\dots,N\})}{N}$).

In this paper we overcome this square-root barrier for almost all moduli. For a positive integer m, we denote by $\omega_3(m)$ the number of its prime divisors of the form $4k+3$. Our main results are the following.

Theorem 1.1. *Let m be square-free and let $A \subset \mathbb{Z}_m$ obey* (1.1). *Then*

$$|A| \leqslant m^{1/2}q^{-1/2}(10\omega(m))^{2\omega(m)},$$

where q denotes the least prime divisor of m of the form $4k+3$ if $\omega_3(m)$ is odd, and $q = 1$ otherwise.

It is an improvement (due to the factor $q^{-1/2}$) of the mentioned result of [G] (note that we obtain worse constants 10 and 2 instead of 3 and 3/2, but it is not so important). While it is not useful directly for all moduli, it allows us (using some "truncation" trick) to obtain a bound $o(m^{1/2})$ for almost all m.

Theorem 1.2. *Let $\varepsilon \in [(\log x)^{-1/2}, 1]$ and $c(\varepsilon) = \exp(-(\log \varepsilon^{-1})^{1/10})$. Then for all but $O(c(\varepsilon)x)$ numbers $m \leqslant x$ and any $A \subset \mathbb{Z}_m$ with* (1.1) *we have*

$$|A| \leqslant m^{1/2-\varepsilon/5}.$$

We immediately conclude the following.

Corollary 1.3. *Let $\varepsilon(m) \to 0$ arbitrarily slowly. Then*

$$|A| \leqslant m^{1/2-\varepsilon(m)}$$

for almost all m and $A \subset \mathbb{Z}_m$ with (1.1).

We now discuss possible improvements of this bound. For $\eta \in (0,1)$, we set

$$M_\eta = \{m \in \mathbb{N} : \text{ for any } A \subset \mathbb{Z}_m \text{ with (1.1) the bound } |A| \leqslant m^\eta \text{ holds }\}.$$

Using this notation, we can reformulate the mentioned corollary from [G] as follows: for any $\varepsilon > 0$ the set $M_{1/2+\varepsilon}$ has density 1. Theorem 1.2 can also be presented in terms of these sets M_η: it means that the density of the set $M_{1/2-\varepsilon}$ tends to 1 as $\varepsilon \to 0$.

Note that in the case $m = p^2$, where p is a prime, $\mathbb{Z}_{p^2}$ contains the set $\{0, p, 2p, \dots, (p-1)p\}$ which has size $m^{1/2}$ and obeys (1.1) (see also Proposition 5.1 of [Y] for a more general statement). Nevertheless, it is believed that for any square-free m and $A \subset \mathbb{Z}_m$ with (1.1) the bound $|A| \ll_\varepsilon m^\varepsilon$ holds for any $\varepsilon > 0$, and, hence, that the set M_ε has density 1. While this hypothesis seems to be far beyond the reach of current methods, one can prove the following weak form of it.

Theorem 1.4. *For any $\varepsilon > 0$ the set M_ε has positive density.*

Finally, we mention a lower bound for almost all moduli.

Theorem 1.5. *For almost all m there exists a set $A \subset \mathbb{Z}_m$ with $(A-A) \cap R_m = \{0\}$ and*

$$|A| \geqslant \exp(0.375(\log\log m)^2(1 + o(1))).$$

In Section 2 we prove Theorem 1.1; we closely follow the proof of the main result of [G] with some modifications. In Section 3 we use Theorem 1.1 and some "truncation" argument to reduce Theorem 1.2 to Lemma 3.2, which concerns with the properties of large prime divisors of the form $4k+3$ of typical integers. In Section 4 we prove Lemma 3.2 and thus finish the proof of Theorem 1.2. The proof of Lemma 3.2 relies on the fact that if $T_1, \dots, T_r$ are disjoint subset of primes in the interval $[y, z] \subset [2, x]$, where $y \to \infty$ and $\log x/\log z \to \infty$, then $\omega(n, T_j) = \#\{p|n, p \in T_j\}$ behave like independent Poisson random variables with parameters $H(T_j) = \sum_{p \in T_j} p^{-1}$. Section 5 is devoted to Theorems 1.4 and 1.5. The proof of Theorem 1.4 relies on the observation that we have the bound $|A| \leqslant m^\varepsilon$ for any $A \subset \mathbb{Z}_m$ with (1.1) whenever m has a prime divisor $q \equiv 3 \pmod 4$ such that $q \geqslant m^{1-\varepsilon}$. Finally, Theorem 1.5 is obtained by a "product" argument from the corresponding lower bounds for the cases of a prime $m = p \equiv 1 \pmod 4$ and $m = q_1 q_2$, where q_1 and q_2 are primes $3 \pmod 4$.

2. Proof of Theorem 1.1

In what follows, we use the words "residue"and "non-residue" for "quadratic residue" and "quadratic non-residue" respectively.

Firstly, we show that it is enough to prove the theorem for odd m. Suppose that m is even and write $m = 2m_1$; then $\mathbb{Z}_m = \mathbb{Z}_2 \oplus \mathbb{Z}_{m_1}$. Set

$$A_1 = A \pmod{m_1} = \{x \in \mathbb{Z}_{m_1} : \text{ there exists } a \in \mathbb{Z}_2 \text{ with } (a, x) \in A\}.$$

Note that for any $x \in \mathbb{Z}_{m_1}$ at most one of the elements $(0, x)$ and $(1, x)$ belongs to A (since their difference $(1, 0)$ is a residue modulo m); denote this element, if it exists, by (a_x, x). Hence, $|A| = |A_1|$. Further, for any distinct $x, y \in A_1$ the difference $x - y$ is a non-residue modulo m_1 (since otherwise the difference $(a_x - a_y, x - y)$ would be a non-zero residue modulo m), and so without loss of generality we may assume that m is odd.

Now we prove the theorem for odd m. We induct on $n = \omega(m)$. Let $n = 1$, that is, $m = p$ is a prime. If $p \equiv 3 \pmod 4$, then $|A| \leqslant 1$, since -1 is a non-residue modulo p and, hence, for any $a \neq b$ one of the differences $a - b$ or $b - a$ is a residue modulo p. If $p \equiv 1 \pmod 4$, we have the bound $|A| \leqslant p^{1/2}$. We give an elegant and folklore proof for that. Let us assume that $|A| > p^{1/2}$ and fix a non-residue $\xi \in \mathbb{Z}_p$. Consider the map $\varphi\colon A^2 \to \mathbb{Z}_p$ defined by $\varphi(a, b) = a + \xi b$. By the pigeonhole principle, there are two distinct pairs (a_1, b_1) and (a_2, b_2) with

$\varphi(a_1,b_1) = \varphi(a_2,b_2)$, that is, $\xi = (a_1-a_2)(b_2-b_1)^{-1}$. It follows that one of the differences a_1-a_2 and b_1-b_2 is a non-residue modulo p, and we are done.

Now assume that $n \geqslant 2$ and the claim is true for all square-free m' with $\omega(m') < n$. Let $p_1 < p_2 < \ldots < p_n$ be the prime divisors of m. Denote by χ_j the Legendre symbol modulo p_j. Since each difference $a-b$ of distinct elements of A is a non-residue modulo p_j for at least one p_j, we have

$$|A| = \sum_{a,b\in A}\prod_{j=1}^{n}(1+\chi_j(a-b)) = |A|^2 + \sum_{D}\sum_{a,b\in A}\chi_D(a-b), \tag{2.1}$$

where D runs over all non-empty subsets of $[n]=\{1,\ldots,n\}$ and $\chi_D(x)=\prod_{j\in D}\chi_j(x)$. Set $p_D = \prod_{j\in D}p_j$. The key observation which makes possible our improvement of the main result of [G] is that we may restrict the outer summation over those (non-empty) D for which $\omega_3(p_D)$ is even (since otherwise $\chi_D(-1) = -1$ and $\chi_D(a-b)+\chi_D(b-a) = 0$ for any $a,b \in \mathbb{Z}_{p_D}$). In what follows, we denote the summation over such D by $\sum'_D$.

Denote $\sigma = 1-|A|^{-1}$. Then we may rewrite (2.1) as follows:

$$|A|^2\sigma = -\sum_D{}' \sum_{a,b\in A}\chi_D(a-b).$$

Using Cauchy-Schwarz, we see that

$$|A|^2\sigma \leqslant \sum_D{}' |A|^{1/2}S_D^{1/2},$$

where

$$S_D = \sum_{a\in A}\left|\sum_{b\in A}\chi_D(a-b)\right|^2.$$

Thus

$$|A|^{3/2}\sigma \leqslant \sum_D{}' S_D^{1/2}. \tag{2.2}$$

Now we need to estimate the sums S_D. For $D \subseteq [n]$ we set

$$H_D = \max\{|A| : A \subset \mathbb{Z}_{mp_D^{-1}},\ A \text{ obeys } (1.1)\}.$$

The following bound is crucial for the induction step.

Lemma 2.1. *For any non-empty $D \subseteq [n]$ we have*

$$S_D \leqslant |A|^2H_D + |A|H_D\sum_{D'\subseteq D}H_{D'}p_{D'}$$

(here and in what follows the summation is over non-empty D').

Proof. For each residue x modulo p_D we set

$$A_x = \{a \in A : a \equiv x \pmod{p_D}\}.$$

One can think of elements of A_x modulo mp_D^{-1}, and the difference of distinct elements of A_x is a non-residue modulo mp_D^{-1}. Then by the definition of H_D we have

$|A_x| \leqslant H_D$; further, obviously, $A = \bigsqcup_{x\in\mathbb{Z}_{p_D}} A_x$ and elements of A_x give the same contribution to S_D. We thus see that

$$\begin{aligned} S_D &= \sum_{x\in\mathbb{Z}_{p_D}} \sum_{a\in A_x} \left|\sum_{b\in A} \chi_D(x-b)\right|^2 = \sum_{x\in\mathbb{Z}_{p_D}} |A_x| \left|\sum_{b\in A} \chi_D(x-b)\right|^2 \\ &\leqslant H_D \sum_{b_1,b_2\in A} \sum_{x\in\mathbb{Z}_{p_D}} \prod_{j\in D} \chi_j(x-b_1)\chi_j(x-b_2) \\ &= H_D \sum_{b_1,b_2\in A} \prod_{j\in D} \sum_{x_j\in\mathbb{Z}_{p_j}} \chi_j(x_j-b_1)\chi_j(x_j-b_2). \end{aligned}$$

Let us compute the inner sum. For the sake of brevity we introduce the following definition: a pair (b_1, b_2) is said to be *special* modulo p if $b_1 \equiv b_2 \pmod p$. We have

$$\sum_{x\in\mathbb{Z}_{p_j}} \chi_j(x-b_1)\chi_j(x-b_2) = p_j - 1$$

if (b_1, b_2) is special modulo p_j, and

$$\sum_{x\in\mathbb{Z}_{p_j}} \chi_j(x-b_1)\chi_j(x-b_2) = \sum_{x\neq b_2} \chi_j\left(1 + \frac{b_2-b_1}{x-b_2}\right) = \sum_{\substack{x\in\mathbb{Z}_{p_j}\\ x\neq 1}} \chi_j(x) = -1$$

otherwise.

Denote by $B_{D'}$ the set of pairs $(b_1, b_2) \in A^2$ which are special modulo each prime p_j with $j \in D'$, and not special modulo every prime p_j with $j \in D\backslash D'$. In particular, $(b_1, b_2) \in B_{D'}$ implies that $b_1 \equiv b_2 \pmod{p_{D'}}$ and thus $|B_{D'}| \leqslant |A|H_{D'}$. We thus have

$$\begin{aligned} S_D &\leqslant H_D \left((-1)^{|D|}|B_\varnothing| + \sum_{D'\subseteq D} (-1)^{|D|-|D'|}\phi(p_{D'})|B_{D'}| \right) \\ &\leqslant H_D|A|^2 + |A|H_D \sum_{D'\subseteq D} p_{D'}H_{D'}, \end{aligned}$$

where ϕ is Euler's function. □

This lemma implies

$$S_D^{1/2} \leqslant |A|H_D^{1/2} + |A|^{1/2}H_D^{1/2} \sum_{D'\subseteq D} H_{D'}^{1/2}p_{D'}^{1/2}.$$

Substituting this estimate into (2.2), we obtain

$$|A|\sigma \leqslant |A|^{1/2}T_1 + T_2, \tag{2.3}$$

where

$$T_1 = \sum_{D\subseteq[n]}{}' H_D^{1/2},$$

$$T_2 = \sum_{D\subseteq[n]}{}' H_D^{1/2} \sum_{D'\subseteq D} H_{D'}^{1/2}p_{D'}^{1/2}.$$

Our further argument is roughly the following. By the induction hypothesis we have $H_D \ll_n (mp_D^{-1})^{1/2}$ (as is usual, the notation $B \ll_n C$ with positive B, C means that $B \leqslant f(n)C$ for an appropriate function f); then

$$T_1 \ll_n m^{1/4} \sum_D p_D^{-1/4} \ll m^{1/4} \sum_D 1 \ll_n m^{1/4}$$

and, similarly,

$$T_2 \ll_n m^{1/2} \sum_D p_D^{-1/4} \sum_{D' \subseteq D} p_{D'}^{1/4} \ll m^{1/2} \sum_D \sum_{D' \subseteq D} 1 \ll_n m^{1/2}.$$

Hence, (2.3) gives us

$$|A| \ll_n |A|^{1/2} m^{1/4} + m^{1/2},$$

and we get a contradiction if $|A| \gg_n m^{1/2}$. So we easily have a bound $|A| \ll_n m^{1/2}$ with some explicit dependence of the constant on n, and this is enough to prove the theorem in the case where $\omega_3(m)$ is even. If $\omega_3(m)$ is odd, we have a better bound $H_D \ll_n (mp_D^{-1}q^{-1})^{1/2}$ if $\omega_3(p_D)$ is even, and the similar argument gives $T_1 \ll_n (mq^{-1})^{1/4}$; the only problem for getting immediately the bound $T_2 \ll_n (mq^{-1})^{1/2}$ (which would imply the theorem in the same way up to explicit dependence on n) is that we have only "trivial" bound $H_{D'} \ll_n (mp_{D'}^{-1})^{1/2}$ if $\omega_3(p_{D'})$ is odd. However, it turns out to be just a technical difficulty and we are still able to proceed as above. Note that these crude bounds for T_1 and T_2 imply the theorem in the form $|A| \leqslant m^{1/2}q^{-1/2}f(n)$ with $f(n) = \exp(O(n^2))$, whereas we are aiming for a better dependence on n.

We turn to the details. For $D \subseteq [n]$, $D \neq \varnothing$, let q_D be the least prime divisor of mp_D^{-1} of the form $4k+3$ if $\omega_3(mp_D^{-1})$ is odd, and $q_D = 1$ otherwise. Since D is non-empty, we have $\omega(mp_D^{-1}) < n$ and we can apply the induction hypothesis, which gives us

$$H_D \leqslant (mp_D^{-1})^{1/2} q_D^{-1/2} (10n)^{2(n-|D|)}. \tag{2.4}$$

Recall that the summation in the sums T_1 and T_2 is taken over D with even $\omega_3(p_D)$ (let us call these D *proper*). Set $q = q_\varnothing$; if $\omega_3(m)$ is even, then $q_D = q = 1$ for any proper D. If $\omega_3(m)$ is odd, then $q_D \geqslant q$ for any proper D. Hence, in both cases we have

$$H_D \leqslant (mp_D^{-1})^{1/2} q^{-1/2} (10n)^{2(n-|D|)}. \tag{2.5}$$

for any proper D.

Now we estimate the sums T_1 and T_2. We begin with a bound for T_1. Using (2.5) and extending the summation to all D, we have

$$(mq^{-1})^{-1/4} T_1 \leqslant {\sum_D}' p_D^{-1/4} (10n)^{n-|D|} \leqslant (10n)^n \sum_{d=1}^n \sum_{|D|=d} p_D^{-1/4} (10n)^{-d} \leqslant \tag{2.6}$$
$$(10n)^n \sum_{d=1}^n (10n)^{-d} \binom{n}{d} \leqslant (10n)^n \sum_{d=1}^n (10n)^{-d} \frac{n^d}{d!} \leqslant$$
$$(10n)^n \sum_{d=1}^\infty \frac{10^{-d}}{d!} \leqslant 0.12(10n)^n.$$

Finally, we estimate T_2. We have

$$T_2 = T_2' + T_2'', \tag{2.7}$$

where

$$T_2' = {\sum_D}' H_D^{1/2} {\sum_{D'\subseteq D}}' H_{D'}^{1/2} p_{D'}^{1/2}$$

(the inner summation is over proper D'), and

$$T_2'' = {\sum_D}' H_D^{1/2} {\sum_{D'\subseteq D}}'' H_{D'}^{1/2} p_{D'}^{1/2}$$

(the inner summation is over non-proper D'). We first work with T_2'. Using the bound (2.5), we find

$$H_{D'} p_{D'} \leqslant (m p_{D'}^{-1})^{1/2} q^{-1/2} (10n)^{2(n-|D'|)} p_{D'} = m^{1/2} p_{D'}^{1/2} q^{-1/2} (10n)^{2(n-|D'|)},$$

and hence

$$T_2' \leqslant (mq^{-1})^{1/2} {\sum_D}' (10n)^{n-|D|} p_D^{-1/4} {\sum_{D'\subseteq D}}' (10n)^{n-|D'|} p_{D'}^{1/4}. \tag{2.8}$$

Now we estimate T_2''. Fix some proper D and non-proper $D' \subseteq D$. Since $\omega_3(p_D)$ is even and $\omega_3(p_{D'})$ is odd, D' can be represented in the form $D' = D_1 \backslash \{p_j\}$ with proper $D_1 \subseteq D$ such that $\omega_3(p_{D_1}) \geqslant 2$ and some $p_j \equiv 3 \pmod 4$; note that $p_j \geqslant q$. Thus, using the bounds (2.4) and $q_{D'} \geqslant 1$, we have

$$H_{D'} p_{D'} \leqslant (m p_{D_1}^{-1} p_j)^{1/2} (10n)^{2(n-|D_1|+1)} p_{D_1} p_j^{-1} \leqslant m^{1/2} p_{D_1}^{1/2} q^{-1/2} (10n)^{2(n-|D_1|+1)}.$$

Then we may rewrite the inner summation in T_2'' over non-proper $D' \subseteq D$ as the summation over proper D_1 with $\omega_3(p_{D_1}) \geqslant 2$; any bound of the above type occurs at most $|D_1|$ times, and hence by (2.5) we have

$$T_2'' \leqslant (mq^{-1})^{1/2} {\sum_D}' (10n)^{n-|D|} p_D^{-1/4} {\sum_{\substack{D_1\subseteq D\\ \omega_3(p_{D_1})\geqslant 2}}}' |D_1| (10n)^{n-|D_1|+1} p_{D_1}^{1/4}. \tag{2.9}$$

Combining (2.8) and (2.9) with (2.7), we obtain

$$\begin{aligned}
&(mq^{-1})^{-1/2}(10n)^{-2n} T_2 \leqslant \\
&\quad {\sum_D}' (10n)^{-|D|} {\sum_{D_1\subseteq D}}' (|D_1|+1)(10n)^{-(|D_1|-1)} (p_{D_1}/p_D)^{1/4} \leqslant \\
&\quad \sum_D (10n)^{-|D|} \sum_{D_1\subseteq D} (|D_1|+1) \leqslant \sum_D (10n)^{-|D|} (|D|+1) 2^{|D|} \leqslant \\
&\qquad\qquad \sum_{d=1}^{n} (d+1) 2^d (10n)^{-d} \frac{n^d}{d!} \leqslant \sum_{d=1}^{\infty} (d+1) \frac{5^{-d}}{d!} \leqslant 0.47.
\end{aligned}$$

In light of this and (2.6), we see from (2.3) that

$$L := |A|^{1/2} \left(|A|^{1/2} \sigma - 0.12 (mq^{-1})^{1/4} (10n)^n \right) \leqslant 0.47 (mq^{-1})^{1/2} (10n)^{2n} =: R.$$

Assume that

$$|A| > (mq^{-1})^{1/2} (10n)^{2n}.$$

But $n \geqslant 2$; hence, $|A| > 100$ and $\sigma = 1 - |A|^{-1} > 0.99$. Therefore

$$L > (0.99 - 0.12)(mq^{-1})^{1/2} (10n)^{2n} > R,$$

a contradiction. This completes the proof.

3. Proof of Theorem 1.2

We begin with the following simple lemma.

Lemma 3.1. *Let $m = m_1m_2$, where $(m_1, m_2) = 1$, and assume that we have a bound $|A_2| \leqslant g(m_2)$ for all $A_2 \subset \mathbb{Z}_{m_2}$ with* (1.1)*. Then for any $A \subset \mathbb{Z}_m$ with* (1.1) *we have*

$$|A| \leqslant m_1 g(m_2).$$

Proof. For any $a_1 \in \mathbb{Z}_{m_1}$ set

$$A(a_1) = \{a \in A : a \equiv a_1 \pmod{m_1}\}.$$

Then $A(a_1) \pmod{m_2}$ obeys (1.1) and, hence, we have $|A(a_1)| \leqslant g(m_2)$. Summing over a_1 completes the proof. □

The idea of the proof of Theorem 1.2 is the following. First, we restrict our attention to those $\sqrt{x} < m \leqslant x$ for which

(i) its powerful[1] part $P(m) = \prod_{p^\alpha || m,\, \alpha \geqslant 2} p^\alpha$ is at most $\log x$;
(ii) $\omega(m) \leqslant 2 \log\log x$.

Almost all integers m satisfy (i) and (ii). Indeed, the number of powerful integers $\leqslant y$ is $O(\sqrt{y})$ and hence by partial summation the number of $m \leqslant x$ failing (i) is at most

$$\sum_{\substack{d > \log x \\ d \text{ is powerful}}} \frac{x}{d} \ll \frac{x}{\sqrt{\log x}}.$$

The number of $m \leqslant x$ failing (ii) is $O(x/(\log x)^{\log 4 - 1})$ by the Hardy-Ramanujan [HR] (see also Theorem 1.5 of [N] or Theorem 3 of [F]) estimate

$$\#\{m \leqslant x : \omega(m) = k\} \ll \frac{x}{\log x} \cdot \frac{(\log\log x + O(1))^{k-1}}{(k-1)!},$$

after summing over $k > 2\log\log x$ (the first summand dominates) and using Stirling's formula. Thus,

$$\#\{m \leqslant x : m \text{ fails (i) or fails (ii)}\} \ll \frac{x}{(\log x)^{1/3}}. \tag{3.1}$$

For m obeying (i) and (ii), let $q_1 > q_2 > \dots$ be prime divisors of $m/P(m)$ of the form $4k + 3$. Take some q_{2j-1} ($j \geqslant 1$) and set $d_{2j-1} = \prod_{i > 2j-1} q_i$. Denote $m_1 = P(m) d_{2j-1}$ and $m_2 = m/m_1$; then m_2 is square-free, $\omega_3(m_2)$ is odd, and the least prime divisor of m_2 of the form $4k+3$ is q_{2j-1}. Now suppose that $A \subset \mathbb{Z}_m$ obeys (1.1). Applying Lemma 3.1 and Theorem 1.1, we get by (i) and (ii)

$$\begin{aligned} |A| &\leqslant P(m) d_{2j-1} (m_2 q_{2j-1}^{-1})^{1/2} (10\omega(m_2))^{2\omega(m_2)} \\ &= m^{1/2} \left(\frac{d_{2j-1}}{q_{2j-1}}\right)^{1/2} P(m)^{1/2} (10\omega(m_2))^{2\omega(m_2)} \\ &= m^{1/2} \left(\frac{d_{2j-1}}{q_{2j-1}}\right)^{1/2} \exp(O(\log_2 x \log_3 x)), \end{aligned} \tag{3.2}$$

[1] A number n is *powerful* if every prime in its prime factorization occurs with exponent at least 2.

where we use iterated logarithm notation $\log_2 x = \log\log x$, $\log_3 x = \log\log\log x$, etc. We see that our bound is good if q_{2j-1} is significantly larger than d_{2j-1}. If

(iii) there exists $q_{2j-1} > x^{\varepsilon}$ such that $q_{2j-1} > d_{2j-1}^2$,

then we have from (3.2)

$$|A| \leqslant m^{1/2-\varepsilon/4} \exp(O(\log_2 x \log_3 x)) \leqslant m^{1/2-\varepsilon/5}$$

for large enough x, since $m > \sqrt{x}$ and $\varepsilon \geqslant (\log x)^{-1/2}$. Thus to prove Theorem 1.2 it suffices to prove that (iii) holds for all but $O(c(\varepsilon)x)$ numbers $m \leqslant x$. Note that if $\varepsilon > 0$ is fixed then (iii) fails for a positive proportion of all m, e.g. those which are m^{ε}-smooth.

For a positive integer m and $y > 0$, we set

$$D(m,y) = \prod_{\substack{q<y,\ q^{\alpha}||m \\ q\equiv 3 \pmod 4}} q^{\alpha}.$$

Theorem 1.2 will evidently follow from the next lemma.

Lemma 3.2 (the condition (iii))**.** *Let* $\varepsilon \in [(\log x)^{-1/2}, 1]$ *and* $c(\varepsilon) = \exp(-(\log \varepsilon^{-1})^{1/10})$. *Then all but* $O(c(\varepsilon)x)$ *integers* $m \leqslant x$ *have a prime divisor* $q_{2j-1} > x^{\varepsilon}$ *with* $q_{2j-1} > D(m, q_{2j-1})^2$.

The same proof gives a similar statement with the exponent 2 replaced by any fixed constant, but we do not need this here.

We note that in the work [Bo] (see also [HT], Chapter 1, and the work [E]) the following question (very close to (iii)) was studied. For $m \leqslant x$ and $y > 0$ define

$$d(m,y) := \max\{d|m : P^{+}(d) < y\},$$

where $P^{+}(d)$ is the largest prime divisor of d. It was shown in [Bo] that for any $u > 0$, almost all m have about $\beta(u)\log\log m$ prime divisors $p|m$ with $p^u > d(m,p)$, where $\beta\colon [0,+\infty) \to [0,1]$ is a continuous increasing function with $\beta(0) = 0$ and $\lim_{u\to\infty} \beta(u) = 1$. Nevertheless, we have several extra requirements in (iii) and thus we cannot use this result directly.

4. Proof of Lemma 3.2

For an integer m and set T of primes, let $\omega(m,T)$ be the number of distinct primes from T which divide m. The proof relies on the fact that if $T_1, \dots, T_r$ are disjoint subset of primes below $y = x^{o(1)}$ as $x \to \infty$, then $\omega(m,T_j)$ behave like independent Poisson random variables with parameters $H_1(T_j) = \sum_{p\in T_j} \frac{1}{p}$. Define the Total Variation Distance $d_{TV}(X,Y)$ between two random variables X and Y taking values in a discrete space Ω by

$$d_{TV}(X,Y) := \max_{A\subset\Omega} |\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)| = \frac{1}{2}\sum_{\omega\in\Omega} |\mathbb{P}(X=\omega) - \mathbb{P}(Y=\omega)|.$$

We cite [F, Theorem 1]. Here

$$H_2(T) := \sum_{p\in T} \frac{1}{p^2}.$$

Lemma 4.1 (Ford [F]). *Let $2 \leqslant y \leqslant x$ and suppose that $T_1, \dots, T_r$ are disjoint nonempty sets of primes in $[2, y]$. Then*

$$d_{TV}\Big((\omega(n, T_1), \dots, \omega(n, T_r)), (Z(T_1), \dots, Z(T_r))\Big) \ll \sum_{j=1}^{r} \frac{H_2(T_j)}{1 + H_1(T_j)} + u^{-u}, \quad u = \frac{\log x}{\log y},$$

where, for any set T of primes, $Z(T)$ is a Poisson random variable with parameter $H_1(T)$, and $Z(T_1), \dots, Z(T_r)$ are independent.

Now we are ready to prove Lemma 3.2. We may suppose that ε is small enough since otherwise the claim follows by taking the implied constant large enough. Set

$$\theta = C(\log \varepsilon^{-1})^{1/10},$$

where C is a fixed, large constant. We consider the primes of the form $4k + 3$ in the interval $(x^{\varepsilon}, x^{\sqrt{\varepsilon}}]$. Set $y_0 = x^{\sqrt{\varepsilon}}$ and $y_j = y_0^{1/\theta^j}$ for $j = 1, \dots, J$, where

$$J = \max\{j : \theta^{-j} \geqslant \varepsilon^{1/2}\} = \left\lfloor \frac{\log \varepsilon^{-1}}{2 \log \theta} \right\rfloor \asymp \frac{\log(\varepsilon^{-1})}{\log \log(\varepsilon^{-1})}. \tag{4.1}$$

Further, define

$$T_j = \{q \equiv 3 \pmod 4 \text{ prime } : q \in (y_j, y_{j-1}]\} \qquad (1 \leqslant j \leqslant J).$$

Then $T_1, \dots, T_J$, are disjoint subsets of primes in $(x^{\varepsilon}, x^{\sqrt{\varepsilon}}]$ and by the Mertens theorem for arithmetic progressions (see [W]),

$$\lambda_j := H_1(T_j) = \frac{1}{2} \log \frac{\log y_{j-1}}{\log y_j} + O(1) = \frac{1}{2} \log \theta + O(1) \in (\frac{1}{3} \log \theta, \log \theta), \tag{4.2}$$

since ε is small enough. For a randomly chosen $m \in [1, x]$, let $\omega_j = \omega(m, T_j)$ and $Z_j = Z(T_j)$ for each j, and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_J)$ and $\mathbf{Z} = (Z_1, \dots, Z_J)$. Applying Lemma 4.1, we obtain

$$d_{TV}(\boldsymbol{\omega}, \mathbf{Z}) \ll \exp(-\varepsilon^{-1/2}) + (\log \theta)^{-1} \sum_{q > x^{\varepsilon}} q^{-2} \ll \exp(-\varepsilon^{-1/2}).$$

Here we used that $\varepsilon \geqslant (\log x)^{-1/2}$. In particular, for any event $E \subset \mathbb{N}_0^J$ we have

$$|\mathbb{P}(\boldsymbol{\omega} \in E) - \mathbb{P}(\mathbf{Z} \in E)| \ll \exp(-\varepsilon^{-1/2}). \tag{4.3}$$

Our main idea is to show that the event

$$E = \{(e_1, \dots, e_J) \in \mathbb{N}_0^J : \exists\, j \leqslant J - 3 \text{ with } e_{j+3} = 0, e_{j+2} = 1, e_{j+1} = 0 \text{ and } e_j = 1\}$$

is very likely. This corresponds to $\omega_{j+3} = 0$, $\omega_{j+2} = 1$, $\omega_{j+1} = 0$ and $\omega_j = 1$ for some j. For such m, it is then very likely that q', the unique prime divisor of m in T_{j+2}, satisfies $q' > D(m, q')^2 = D(m, y_{j+3})^2$ and that q'', the unique prime divisor of m in T_j, satisfies $q'' > D(m, q'')^2 = D(m, y_{j+1})^2$. Furthermore, one of the primes q', q'' has an odd index, that is, equals q_{2h-1} for some h.

For any $k \leqslant (J - 3)/4$, we have by (4.2)

$$\begin{aligned}\mathbb{P}\big(Z_{4k+3} = 0, Z_{4k+2} = 1, Z_{4k+1} = 0, Z_{4k} = 1\big)& \\ = \lambda_{4k+2}\lambda_{4k} \mathrm{e}^{-\lambda_{4k+3} - \lambda_{4k+2} - \lambda_{4k+1} - \lambda_{4k}} &\geqslant \theta^{-4}.\end{aligned}$$

The 4-tuples $(Z_{4k+3}, Z_{4k+2}, Z_{4k+1}, Z_{4k})$ are independent for different k. Therefore,

$$\mathbb{P}(\mathbf{Z} \notin E) \leqslant (1 - \theta^{-4})^{(J/4)-2} \leqslant \mathrm{e}^{-\theta^{-4}(J/4-2)}.$$

By (4.3), it follows that

$$\mathbb{P}(\boldsymbol{\omega} \notin E) \ll \mathrm{e}^{-\theta^{-4}(J/4-2)} + \mathrm{e}^{-\varepsilon^{-1/2}} \ll \mathrm{e}^{-(\log \varepsilon^{-1})^{1/2}} \ll c(\varepsilon) \tag{4.4}$$

using (4.1).

We do not want m to have a big smooth part. Consider the condition

(iv) For every $0 \leqslant j \leqslant J$, $D(m, y_j) \leqslant y_j^{\theta/2}$.

Using Theorem 07 of [HT], it follows for some absolute constant $c_0 > 0$ that the number of $m \leqslant x$ failing (iv) is at most

$$\ll \sum_{j=0}^{J} x\mathrm{e}^{-c_0\theta/2} \ll xc(\varepsilon)$$

if C is taken large enough in the definition of θ. Therefore, by (4.4), the number of integers $m \leqslant x$ which fail (iv) or have $\boldsymbol{\omega} \notin E$ is $O(c(\varepsilon)x)$. By (3.1), the number of $m \leqslant x$ failing (i) or failing (ii) is likewise $O(c(\varepsilon)x)$.

Consider now an integer $m \leqslant x$ satisfying (i), (ii), (iv) and $\boldsymbol{\omega} \in E$. In particular, by (i) the primes $q|m$ with $q > x^{\varepsilon}$ divide m to the first power only. By $\boldsymbol{\omega} \in E$, there is at least one $j \leqslant J - 3$ such that

$$\omega_{j+3} = 0, \omega_{j+2} = 1, \omega_{j+1} = 0, \omega_j = 1.$$

Let q' be the unique prime divisor of m in T_{j+2} and q'' be the unique prime divisor of m in T_j. By (iv),

$$D(m, q') = D(m, y_{j+3}) \leqslant y_{j+3}^{\theta/2} = y_{j+2}^{1/2} \leqslant (q')^{1/2}.$$

and likewise

$$D(m, q'') = D(m, y_{j+1}) \leqslant y_{j+1}^{\theta/2} = y_j^{1/2} \leqslant (q'')^{1/2}.$$

Furthermore, one of the primes q', q'' has an odd index, that is, equals q_{2h-1} for some h. This completes the proof of Lemma 3.2.

5. Proofs of Theorems 1.4 and 1.5

Proof of Theorem 1.4. We may assume that $\varepsilon \in (0, 1/2)$. We claim that M_ε contains every number m that has a prime factor $q \geqslant m^{1-\varepsilon}$ with $q \equiv 3 \pmod 4$. To see this, suppose that $A \subset \mathbb{Z}_m$ obeys $(A - A) \cap R_m = \{0\}$. Then by Lemma 3.1 (with $m_1 = mq^{-1}$, $m_2 = q$ and $g(m_2) = 1$) we see that $|A| \leqslant mq^{-1} \leqslant m^{\varepsilon}$. As each number m has at most one such prime factor q, the number of such $m \leqslant x$ is at least

$$\sum_{x^{1-\varepsilon} \leqslant q \leqslant x} \#\{m \in (1, x] : q|m\} = \frac{x}{2} \log\left(\frac{1}{1-\varepsilon}\right) + O\left(\frac{x}{\log x}\right),$$

by the Mertens theorem for arithmetic progressions. □

To prove Theorem 1.5, we first need the following two results.

Lemma 5.1 (Cohen [C]). *For any prime $p \equiv 1 \pmod 4$ there exists $A \subset \mathbb{Z}_p$ obeying* (1.1) *with $|A| \geqslant \frac{1}{2\log 2}\log p$.*

For sake of completeness, we provide a proof here. We follow the short argument from [G].

Proof. Consider the complete graph $G = (V, E)$ with $V = \mathbb{Z}_p$ and the partition $E = E_1 \bigsqcup E_2$, where $E_1 = \{(x, y) : x - y \text{ is a residue}\}$ and $E_2 = E \backslash E_1$. Then, by Ramsey's theorem for two colours (see, for instance, [TV], Theorem 6.9), one can find a complete monochromatic subgraph $G' = (V', E')$ of our graph G with $|V'| = n$ whenever $|V| = p \geqslant \binom{2n-2}{n-1}$. We thus can find such a subgraph of size $n \geqslant \log p/\log 4$. If $E \subseteq E_2$, then the set V' of its vertices gives an example we need; if $E \subseteq E_1$, then for any non-residue $\xi \in \mathbb{Z}_p$ we get such an example in the form $\xi V'$. The claim follows. □

Lemma 5.2. *Let* $q_1 > q_2$ *be primes* 3 (mod 4). *Then there exists* $A \subset \mathbb{Z}_{q_1 q_2}$ *obeying* (1.1) *with* $|A| \geqslant \frac{1}{\log 2} \log q_2$.

Proof. For $j = 1, 2$, let $V_j = (\mathbb{Z}_{q_j}, E_{q_j})$ be the tournament of quadratic residues modulo q_j, that is, the directed graph with the set of vertices $\mathbb{Z}_{q_j}$ and $\{a, b\} \in E_{q_j}$ iff $a - b \in R_{q_j} \backslash \{0\}$. It is well-known ([St]; see also [EM]) that any tournament on n vertices contains a transitive subtournament of size $\lfloor \log n/\log 2 \rfloor + 1$ (it follows from the fact that any tournament on 2^n vertices contains a transitive subtournament of size $n + 1$, which can be proved by an easy induction). Applying this to V_j, we find two sets $A_1 = \{a_1^{(1)}, \ldots, a_k^{(1)}\} \subset \mathbb{Z}_{q_1}$ and $A_2 = \{a_1^{(2)}, \ldots, a_k^{(2)}\} \subset \mathbb{Z}_{q_2}$ of size $k \geqslant \log q_2/\log 2$ such that $a_s^{(j)} - a_t^{(j)} \in R_{q_j}$ for $1 \leqslant s < t \leqslant k$ and $j = 1, 2$. Then the set $A = \{(a_s^{(1)}, a_{k+1-s}^{(2)})\}_{s=1}^k \subset \mathbb{Z}_{q_1 q_2}$ is what we need, since $\left(\frac{-1}{q_j}\right) = -1$ for $j = 1, 2$. □

Note that Graham and Ringrose [GR] showed that the least quadratic non-residue $n(p)$ satisfies $n(p) \gg \log p \log\log\log p$ infinitely often, and one can expect that the same bound holds infinitely often for primes $p \equiv 1 \pmod 4$ and $q \equiv 3 \pmod 4$ separately. Then, as was mentioned in the introduction, we can use the sets $\xi \cdot \{1, \ldots, n(p)\}$, where $\xi \in \mathbb{Z}_p \backslash R_p$ is any non-residue, instead of those we constructed in the proof of Lemma 5.1, and the sets $\{(s, n + 1 - s)\}_{s=1}^n$, where $n = \min\{n(q_1), n(q_2)\}$ instead of those from Lemma 5.2. As this bound on $n(p)$ applies only to a very sparse set of primes p, using it would not affect our lower bound in Theorem 1.5.

Proof of Theorem 1.5. Let $\omega_j(m, t) = \#\{p \leqslant t : p | m, p \equiv j \pmod 4\}$, $j \in \{1, 3\}$. Consider the set of $m \leqslant x$ such that

(a) $p^2 | m$ implies that $p > \log x$;
(b) $|\omega_j(m, t) - 0.5 \log\log t| < (\log\log x)^{2/3} \quad (3 \leqslant t \leqslant m,\ j \in \{1, 3\})$.

The number of m failing (a) is $O(x/\log x)$. Almost all integers satisfy (b), and this may be derived from Theorem 7.2 in Kubilius [K], upon taking f to be the strongly additive function with $f(p) = 1$ if $p \equiv j \pmod 4$ and $f(p) = 0$ otherwise. Now suppose that $m \in (\sqrt{x}, x]$ obeys (a) and (b). For $p | m$ and $p > \log x$, p divides m to the first power. By Lemma 5.1, for such $p \equiv 1 \pmod 4$ there is a set $A_p \subset \mathbb{Z}_p$ with (1.1) of size $\gg \log p$. Let $q_1 > q_2 > \ldots$ be the primes 3 (mod 4) dividing m, greater than $\log x$; each of them also divides m to the first power. By Lemma 5.2, there is a set $A_{q_{2j-1} q_{2j}} \subset \mathbb{Z}_{q_{2j-1} q_{2j}}$ with (1.1) of size $\gg \log q_{2j}$. Further, it is easy to see that the set

$$A=\prod_{\substack{p|m\\ p\equiv 1 \pmod 4\\ p>\log x}} A_p \times \prod_{\substack{q_{2j}|m\\ q_{2j}>\log x}} A_{q_{2j-1}q_{2j}} \subset \prod_{p|m}\mathbb{Z}_p \subseteq \mathbb{Z}_m$$

obeys (1.1). It remains to estimate $|A|$. By (b) with $t=m$, we get

$$\begin{aligned}\log|A| &\geqslant \sum_{\substack{p|m, p>\log x\\ p\equiv 1 \pmod 4}} (\log\log p - O(1)) + \sum_{\substack{q_{2j}|m,\\ q_{2j}>\log x}} (\log\log q_{2j} - O(1))\\ &= \sum_{\substack{p|m, p>\log x\\ p\equiv 1 \pmod 4}} \log\log p + 0.5 \sum_{\substack{q|m, q>\log x\\ q\equiv 3 \pmod 4}} \log\log q - O(\log\log x). \end{aligned} \tag{5.1}$$

Using Abel's summation technique, we find by (b) that

$$\begin{aligned}&\sum_{\substack{p|m, \log x<p\leqslant\sqrt{x}\\ p\equiv 1 \pmod 4}} \log\log p\\ &\qquad \geqslant \omega_1(m,\sqrt{x})\log\log\sqrt{x} - \omega_1(m,\log x)\log\log\log x - \int_{\log x}^{\sqrt{x}} \frac{\omega_1(m,u)du}{u\log u}\\ &\qquad = 0.25(\log\log x)^2 + O((\log\log x)^{5/3}).\end{aligned}$$

Analogously,

$$\sum_{\substack{q|m, \log x<q\leqslant\sqrt{x}\\ q\equiv 3 \pmod 4}} \log\log q \geqslant 0.25(\log\log x)^2 + O((\log\log x)^{5/3}).$$

The claim follows from (5.1). □

Acknowledgment

The authors thank Sergei Konyagin for suggesting Lemma 5.2.

References

[BPPS] A. Balog, J. Pelikán, J. Pintz, and E. Szemerédi, *Difference sets without κth powers*, Acta Math. Hungar. **65** (1994), no. 2, 165–187, DOI 10.1007/BF01874311. MR1278767

[Bo] J. D. Bovey, *On the size of prime factors of integers*, Acta Arith. **33** (1977), no. 1, 65–80, DOI 10.4064/aa-33-1-65-80. MR437476

[Bu] D. A. Burgess, *The distribution of quadratic residues and non-residues*, Mathematika **4** (1957), 106–112, DOI 10.1112/S0025579300001157. MR93504

[C] Stephen D. Cohen, *Clique numbers of Paley graphs*, Quaestiones Math. **11** (1988), no. 2, 225–231. MR949363

[E] P. Erdős, *On some properties of prime factors of integers*, Nagoya Math. J. **27** (1966), 617–623. MR204378

[EM] P. Erdős and L. Moser, *On the representation of directed graphs as unions of orderings* (English, with Russian summary), Magyar Tud. Akad. Mat. Kutató Int. Közl. **9** (1964), 125–132. MR168494

[F] K. Ford, *Joint Poisson distribution of prime factors in sets*, preprint. `arXiv:2006.12650`, Math. Proc. Cambridge Philos. Soc., to appear.

[G] M. R. Gabdullin, *On subsets of $\mathbb{Z}_m$ whose difference does not contain squares* (Russian, with Russian summary), Mat. Sb. **209** (2018), no. 11, 60–68, DOI 10.4213/sm8992; English transl., Sb. Math. **209** (2018), no. 11, 1603–1610. MR3871552

[GR] S. W. Graham and C. J. Ringrose, *Lower bounds for least quadratic nonresidues*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math., vol. 85, Birkhäuser Boston, Boston, MA, 1990, pp. 269–309. MR1084186

[HT] Richard R. Hall and Gérald Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge University Press, Cambridge, 1988, DOI 10.1017/CBO9780511566004. MR964687

[HP] B. Hanson and G. Pertidis, *Refined estimates concerning sumsets contained in the roots of unity*, Proc. Lond. Math. Soc. (3), **121** (2020), no. 2, 287–292.

[HR] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Math. Oxford **48**, (1917), 76–92.

[K] J. Kubilius, *Probabilistic methods in the theory of numbers*, Translations of Mathematical Monographs, Vol. 11, American Mathematical Society, Providence, R.I., 1964. MR0160745

[La] Edmund Landau, *Handbuch der Lehre von der Verteilung der Primzahlen. 2 Bände* (German), Chelsea Publishing Co., New York, 1953. 2d ed; With an appendix by Paul T. Bateman. MR0068565

[Le] Mark Lewko, *An improved lower bound related to the Furstenberg-Sárközy theorem*, Electron. J. Combin. **22** (2015), no. 1, Paper 1.32, 6. MR3315474

[Li] U. V. Linnik, *A remark on the least quadratic non-residue*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **36** (1942), 119–120. MR0007758

[MR] M. Matolcsi, I. Z. Ruzsa, *Difference sets and positive exponential sums II: Quadratic and cubic residues in cyclic groups*, preprint.

[N] Karl K. Norton, *On the number of restricted prime factors of an integer. I*, Illinois J. Math. **20** (1976), no. 4, 681–705. MR419382

[PSS] János Pintz, W. L. Steiger, and Endre Szemerédi, *On sets of natural numbers whose difference set contains no squares*, J. London Math. Soc. (2) **37** (1988), no. 2, 219–231, DOI 10.1112/jlms/s2-37.2.219. MR928519

[R] I. Z. Ruzsa, *Difference sets without squares*, Period. Math. Hungar. **15** (1984), no. 3, 205–209, DOI 10.1007/BF02454169. MR756185

[Sa] A. Sárkőzy, *On difference sets of sequences of integers. I*, Acta Math. Acad. Sci. Hungar. **31** (1978), no. 1-2, 125–149, DOI 10.1007/BF01896079. MR466059

[St] Richard Stearns, *The voting problem*, Amer. Math. Monthly **66** (1959), 761–763, DOI 10.2307/2310461. MR109087

[TV] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006, DOI 10.1017/CBO9780511755149. MR2289012

[V] I. M. Vinogradov, *On the distribution of quadratic residues and non-residues*, (in Russian), Journal of the Physico-Mathematical Society of Perm University (1919), no. 2, 1–16.

[W] Kenneth S. Williams, *Mertens' theorem for arithmetic progressions*, J. Number Theory **6** (1974), 353–359, DOI 10.1016/0022-314X(74)90032-8. MR364137

[Y] K. Younis, *Lower bounds in the polynomial Szemerédi theorem*, available at `arXiv:1908.06058`.

Department of Mathematics, University of Illinois at Urbana-Champaign,1409 West Green Street, Urbana, Illinois 61801

Email address: `ford@math.uiuc.edu`

Steklov Mathematical Institute, Gubkina str. 8, 119991 Moscow, Russia

Email address: `gabdullin.mikhail@yandex.ru, gabdullin@mi-ras.ru`