POPULAR DIFFERENCES AND GENERALIZED SIDON SETS

MAX WENQIANG XU

Abstract. For a subset \( A \subseteq [N] \), we define the representation function \( r_{A-A}(d) := \# \{(a, a') \in A \times A : d = a - a'\} \) and define \( M_D(A) := \max_{1 \leq d < D} r_{A-A}(d) \) for \( D > 1 \). We study the smallest possible value of \( M_D(A) \) as \( A \) ranges over all possible subsets of \([N]\) with a given size. We give explicit asymptotic expressions with constant coefficients determined for a large range of \( D \). We shall also see how this problem connects to a well-known problem about generalized Sidon sets.

1. Introduction

For a positive integer \( N \), denote by \([N]\) the discrete interval \( \{1, 2, \cdots, N\}\). For a subset \( A \subseteq [N] \), we study the difference set \( A - A := \{a - a' : a, a' \in A\} \).

We introduce the representation function \( r_{A-A} \), defined by
\[
r_{A-A}(d) := \# \{(a, a') \in A \times A : d = a - a'\}
\]
The representation function has been investigated a lot in the literature. See, for example, [2], [3]. In this paper, we are interested in the maximum size of \( r_{A-A}(d) \), as \( d \) ranges over all positive integers smaller than \( D \) for different thresholds \( D \). More precisely, let
\[
M_D(A) := \max_{1 \leq d < D} r_{A-A}(d)
\]
and \( f_D(N, \alpha) \) be the smallest possible value of \( M_D(A) \), as \( A \) ranges over all subsets \( A \subseteq [N] \) with \( |A| \geq \alpha N \), i.e.,
\[
f_D(N, \alpha) := \min_{A \subseteq [N], |A| \geq \alpha N} M_D(A).
\]

Theorem 1.1. Let \( 2 \leq D \leq N \) be positive integers and let \( \alpha \in (0, 1) \). Write \( D = (1 + \gamma)\alpha^{-1} \) for some \(-1 < \gamma \), then we have the following:

(1) If \( \gamma \leq 0 \), then \( f_D(N, \alpha) = 0 \).
(2) If \( 0 < \gamma \leq 1 - \delta \) for some positive number \( \delta \), \( N \gg \delta \alpha^{-2} \) and \( D \) is sufficiently large in terms of \( \delta^{-1} \), then \( f_D(N, \alpha) = (1 + o(1)) \frac{2\alpha}{1+\gamma} \alpha^2 N \).
(3) If \( \alpha^2 N \rightarrow \infty \), \( \gamma \rightarrow \infty \) and \( D \alpha \rightarrow \infty \) as \( N \rightarrow \infty \), then \( f_D(N, \alpha) = (1 + o(1))\alpha^2 N \).

Here \( \gamma = \gamma(N), \alpha = \alpha(N), D = D(N) \). They are all regarded as functions of \( N \).

M.W.Xu is supported by a London Mathematics Society Undergraduate Research Bursary and the Mathematical Institute at University of Oxford.
Remark 1.2. 1) When $\alpha \geq 1/2$, the above conclusions can be easily checked. So we focus on $\alpha \in (0, 1/2)$ from now on.
2) A more precise statement of Theorem 1.1 (2) is Theorem 2.1.

We ask the following question related to (2) of Theorem 1.1.

**Question:** Suppose that there is a set $A$ such that
\[
MD(A) = (1 + o(1)) \frac{2\gamma}{(1 + \gamma)^2} \alpha^2 N,
\]
i.e., it takes the extremal value. Then, what can you say about the structure of $A$?

This is a vague inverse problem and one purpose is to investigate if the construction we give later in the Proposition 2.3 is essentially unique or not. We believe the set we constructed is unique up to some permutations.

1.1. Connections with Sidon sets. We shall see that there are some connections between problems about Sidon sets and the problem we are investigating in the case $D = N$.

**Definition 1.1.** (Sidon sets and $g$-Sidon sets) A set of natural numbers $S$ is called a Sidon set if the equation $a + b = c + d$ has only the trivial solution $\{a, b\} = \{c, d\}$, where $a, b, c, d$ are elements of the set $S$. A set $A$ is called a $g$-Sidon set if for any integer $x$ we have:
\[
r_{A_A}(x) := \#\{(a, b) \in A \times A : a + b = x\} \leq g.
\]

Cilleruelo, Ruzsa and Vinuesa (2010) [1] proved the following result.

**Theorem 1.3.** Define $\beta_g(N) := \max\{|A|\}$ where $A \subseteq [N]$ and $A$ is a $g$-Sidon set. Then,
\[
\sigma_1(g) \sqrt{gN(1 - o(1))} \leq \beta_g(N) \leq \sigma_2(g) \sqrt{gN(1 + o(1))}.
\]
Moreover,
\[
\lim_{g \to \infty} \sigma_1(g) = \lim_{g \to \infty} \sigma_2(g) = \sigma, \text{ for some constant } \sigma.
\]

We can translate our result in Theorem 1.1 into their language.

**Corollary 1.4.** Define $\alpha_g(N) := \max\{|A|\}$ where $A \subseteq [N]$, $r_{A-A}(x) \leq g$ for all non-zero $x$. Then,
\[
\sigma_1(g) \sqrt{gN(1 - o(1))} \leq \alpha_g(N) \leq \sigma_2(g) \sqrt{gN(1 + o(1))}.
\]
Moreover,
\[
\lim_{g \to \infty} \sigma_1(g) = \lim_{g \to \infty} \sigma_2(g) = 1.
\]

1.2. Plan for the paper. We first prove Theorem 1.1 (1) in Section 2 by constructing an explicit example. Theorem 1.1 (2) is proved in Section 2 by using the principle of inclusion and exclusion to prove a lower bound for $f_D(N, \alpha)$ and constructing an explicit example to attain the upper bound. To prove the lower bound in Theorem 1.1 (3) in Section 3, we use a deterministic method by first proving results in the case of cyclic groups and then applying the results to the case of integers.
Acknowledgement. The author would like to express his sincere gratitude to Andrew Granville and Fernando Xuancheng Shao, for carefully reading early drafts of the paper and providing many useful comments, and also to Oleksiy Klurman who inspired the author to study this problem. Thanks to the anonymous referee for many helpful comments.

2. Proof of Theorem 1.1 (1) and (2): the case $D$ is small

We first prove Theorem 1.1 (1): $f_D(N, \alpha) = 0$ when $D \leq \alpha^{-1}$.

Proof of Theorem 2.1 (1). We prove this fact by constructing a set $A$ with $|A| \geq \alpha N$ and $M_D(A) = 0$. Let $a := \lceil \alpha^{-1} \rceil$ and $l$ is chosen to satisfy $la \leq N - 1 < (l + 1)a$.

$$A := \{1, a + 1, 2a + 1, 3a + 1, \ldots, la + 1\}.$$ 

The density of $A$ is,

$$\frac{|A|}{N} = \frac{l + 1}{N} \geq \frac{1}{a} \geq \alpha.$$ 

Since $D \leq \alpha^{-1}$, $D \leq a$ also holds. Since all non-zero positive differences here are at least $a$, we get $M_D(A) = 0$. This gives us $f_D(N, \alpha) = 0$. □

Now we prove the following theorem, which gives the result for the case when $D$ is relatively small.

Theorem 2.1. Suppose that $D, N$ are positive integers and $\alpha \in (0, 1)$. Let $D, \alpha, N$ satisfy the following relations.

1. $\alpha^{-1} < D \leq (2 - \delta)\alpha^{-1}$, for some positive $\delta < 1$ such that $D\delta > 1$.
2. $N \geq 2D^3/(D\delta - 1)$.

Then

$$f_D(N, \alpha) = \frac{2(D\alpha - 1)}{D(D - 1)} N + O(1).$$

Notice that this is just a more accurate statement of Theorem 1.1 (2).

2.1. Lower bound for $f_D(N, \alpha)$. In this section we are going to use the basic technique, the principle of inclusion and exclusion, to find the lower bound.

Proposition 2.2. For any $2 \leq D \leq N$ and any $\alpha \in (0, 1)$, we have

$$f_D(N, \alpha) \geq \frac{2(D\alpha - 1)}{D(D - 1)} N - \frac{2}{D}.$$ 

Proof. Let $A \subset [N]$ be any subset with $|A| \geq \alpha N$. We need to show that

$$M_D(A) \geq \frac{2(D\alpha - 1)}{D(D - 1)} N - \frac{2}{D}.$$ 

For any $d \in \mathbb{Z}$, define the translate $A + d$ to be the set

$$A + d := \{a + d : a \in A\}.$$
By the principle of inclusion and exclusion, we estimate the size of the union of $A + d$ as $d$ ranges over $0 \leq d < D$ as following:

$$\left| \bigcup_{0 \leq d < D} (A + d) \right| \geq \sum_{0 \leq d < D} |A + d| - \sum_{0 \leq d < d' < D} |(A + d) \cap (A + d')|. \tag{2.1}$$

Since $A + d \subseteq [N + D - 1]$ for $0 \leq d < D$, the left hand side above is at most $N + D - 1$. On the other hand, the first sum on the right hand side is $D|A|$. Thus we obtain

$$\sum_{0 \leq d < d' < D} |(A + d) \cap (A + d')| \geq D|A| - (N + D - 1) \geq (D\alpha - 1)N - (D - 1).$$

Then by the principle of inclusion and exclusion, we estimate the size of the union $\bigcup_{0 \leq d < D} (A + d)$. The conclusion follows by noting that $|(A + d) \cap (A + d')|$ is precisely $r_{A-A}(d' - d)$, and $1 \leq d' - d < D$.

## 2.2. Upper bound for $f_D(N, \alpha)$.

**Proposition 2.3.** Let $2 \leq D \leq N$ and $\alpha \in (0, 1)$ be parameters satisfying

$$D(2 - D\alpha) > 1 \tag{2.2}$$

and

$$N \geq \frac{D(D - 1)(2D + 1)}{D(2 - D\alpha) - 1}. \tag{2.3}$$

Then

$$f_D(N, \alpha) \leq \frac{2(D\alpha - 1)}{D(D - 1)}N + 4. \tag{2.4}$$

**Remark 2.4.** For example, if $\alpha = o(1)$ then these assumptions hold when $D \leq (2 - c)/\alpha$ for any absolute constant $c > 0$ such that $Dc > 1$ and $N$ is much larger than $\alpha^{-2}$.

**Proof.** We construct one subset $A$ with particular structure such that it has density at least $\alpha$ and $M_D(A)$ attains the upper bound in (2.4).

We define a set $B \subseteq \{1, 2, \cdots, D^2\}$ as following:

$$B := \{D, 2D, 3D, \cdots, D^2\} \cup \{1, D + 2, 2D + 3, \cdots, (D - 2)D + D - 1\}. \tag{2.5}$$

so that $|B| = 2D - 1$. Let $M = \lfloor N/D \rfloor$ and let $k = \lfloor (\alpha N - M)/(D - 1) \rfloor$ so that

$$k \leq \frac{\alpha N - M}{D - 1} + 1 \leq \frac{\alpha N - N/D}{D - 1} + 2 = \frac{D\alpha - 1}{D(D - 1)}N + 2.$$

By the lower bound on $N$, we have

$$kD \leq \frac{D\alpha - 1}{D - 1}N + 2D \leq \frac{N}{D} - 1 \leq M.$$

Thus we may define $A_1, A_2 \subseteq [N]$ by

$$A_1 := \{x + iD^2 : x \in B, i = 0, 1, 2, \cdots, k-1\}, \quad A_2 := \{jD : kD < j \leq M\}. \tag{2.6}$$
and then take $A = A_1 \cup A_2$. Hence
\[ |A| = k|B| + (M - kD) = kD + M - k \geq \alpha N, \]
where the last inequality follows by our choice of $k$.

We only need to show that $r_{A-A}(d) \leq 2k$ holds for any $1 \leq d < D$ to complete our proof. For any $1 \leq d < D$, if $d = a - a'$ for some $a, a' \in A$, then we must have $a, a' \in A_1$ since any two distinct elements must have difference at least $D$ if at least one of them is in $A_2$. Writing $a = x + iD^2$ and $a' = x' + i'D^2$ with $x, x' \in B$ and $0 \leq i, i' < k$, we first observe that the possible values of $|i - i'|$ can only be 0 or 1.

If $|i - i'| = 2$, then by triangle inequality we have
\[ |(x + iD^2) - (x' + i'D^2)| \leq |x - x'| = D^2 + 1 > D. \]

If $|i - i'| = 1$, by considering the structure of $B$ we have if $\{x, x'\} \neq \{D^2, 1\}$, then
\[ |(x + iD^2) - (x' + i'D^2)| \geq D. \]

So one must have $\{x, x'\} = \{D^2, 1\}$ and $d = 1$ if $|i - i'| = 1$. As for $d = 1$, if $i = i'$, then one must have $x = (D - 2)D + D - 1$ and $x' = (D - 1)D$. So $r_{A-A}(1) = k + (k - 1) < 2k$. As for $d \geq 2$, we showed that $i$ and $i'$ must satisfy $i = i'$ which gives $d = x - x'$. And the number of possible choices for $x, x'$ in $B$ is not larger than 2 in this case. So $r_{A-A}(d) \leq 2k$ for all $d \neq 1$.

In conclusion, the result holds for all $1 \leq d < D$. \hfill \Box

At the end of this subsection, we give an explanation for why the principle of inclusion and exclusion works. Since for any $y \in (A + i) \cap (A + j) \cap (A + k)$, where $i, j, k \in \{0, 1, 2, \cdots, D - 1\}$, there exists $x \in A$, such that $x + i, x + j, x + k$ are all in $A$, i.e., they lie in an interval with length $D - 1$. But this is impossible in our constructed example, where any interval with length $D - 1$ has at most 2 elements instead of 3.

There is no point in the intersection of the 3 sets, so it is even impossible for 4 or more sets. Hence the values of all the moments with order greater than two must be zero.

2.3. Conclusion for $f_D(N, \alpha)$ with $\alpha^{-1} < D \leq (2 - \delta)\alpha^{-1}$. In this part, we combine the two propositions before to prove Theorem 2.1.

Proof of Theorem 2.1. The lower bound in Proposition 2.2 and the upper bound in Proposition 2.3 together give us the result in Theorem 2.1. We only need to check that the assumptions (2.2) and (2.3) are satisfied. Since there exists $\delta > 0$ such that $D \leq (2 - \delta)\alpha^{-1}$, the assumption (1) in Theorem 2.1 holds by the fact that
\[ D(2 - D\alpha) \geq D\delta > 1. \]

Now we check that assumption (2.3) holds. Notice that
\[ \frac{D(D - 1)(2D + 1)}{D(2 - D\alpha) - 1} \leq \frac{2D^3}{D\delta - 1}. \]

From assumption (2) in Theorem 2.1, we know (2.3) holds. \hfill \Box
Remark 2.5. From the whole pattern, we shall see that $D = \lfloor \alpha^{-1} \rfloor + 1$ is a crucial case in the problem. We state this result separately here, which has been covered in Theorem 2.1. To make the form nicer, we state the result in a special case that $\alpha^{-1}$ is an integer.

$$f_D(N, \alpha) = \frac{2\alpha^3}{1 + \alpha}N + O(1).$$

This result can also compare to [4, Lemma 2.3]. We state their result here.

**Lemma 2.6.** Suppose $S \subset [H, H + N]$ and $|S| = \alpha N$. Here $[H, H + N]$ is an integer interval. Then there exists a positive integer $d$, such that $d \leq 1/\alpha$ and

$$r_{S - S}(d) \geq \frac{\alpha^3}{2}N.$$

In Theorem 1.1, there is one interesting phenomenon when $\gamma$ is relatively small. Precisely, if $\gamma = O(\alpha)$ then we have $f \asymp \alpha^3 N$, e.g., the case in remark 2.5.

### 3. Proof of the lower bound in Theorem 1.1 (3)

In this section, we use Fourier analytic tools to prove Theorem 3.1 which gives the lower bound of $f_D(N, \alpha)$ for a wide range of $D$. The argument is similar to the proof of [3, Theorem 1.1].

**Theorem 3.1.** Let $2 \leq D \leq N$, then

$$f_D(N, \alpha) \geq \frac{\alpha^2 N^2}{N + D} - \frac{\alpha N + 1}{D}. \quad (3.1)$$

In particular, if $D \alpha \to \infty$, we have

$$f_D(N, \alpha) \geq (1 + o(1))\alpha^2 N. \quad (3.2)$$

**Proof.** Let $A$ be any subset of $[N]$ with density at least $\alpha$. We regard $A$ as a subset of $\mathbb{Z}/(N + D)\mathbb{Z}$ in the natural way. To fix our notation, we define the convolution of $f, g$ where $f, g : G \to \mathbb{C}$ are two functions on the abelian group $G$.

$$f * g(x) := \sum_{y \in G} f(y)g(x - y).$$

And we use our traditional notation to get that

$$1_A * 1^o_A(x) = \# \{(a, a') \in A \times A : a - a' = x\},$$

where $f^o(x) := f(-x)$.

We use $M_D(A)$ to denote the quantity the same as in the case of integers before. It is not true that for general $x \in \mathbb{Z}/(N + D)\mathbb{Z}$, $1_A * 1^o_A(x) \leq M_D(A)$ in modular version. However, for any $x$ with $0 < |x| \leq D$, in the modular version we still have

$$1_A * 1^o_A(x) \leq M_D(A).$$

Let $I$ be the characteristic function of the set $\{1, 2, \cdots, D\}$. Then we estimate the following quantity.

$$E := \sum_{x \in \mathbb{Z}/(N + D)\mathbb{Z}} 1_A * 1^o_A(x)I * I^o(x).$$
We first notice that each summand is zero when \(|x| > D\). Combining this and what we discussed above, we have
\[
E \leq |A|D + \sum_{0 < |x| \leq D} 1_A \ast 1^o_A(x)I \ast I^o(x).
\]
The right hand side can be further upper bounded by
\[
|A|D + M_D(A)D^2.
\]
On the other hand, we use Plancherel’s identity to get
\[
E = \left( N + D \right) \sum_r |\hat{1}_A(r)|^2 |\hat{I}(r)|^2 \geq \frac{|A|^2D^2}{N + D}.
\]
The above two inequalities give us the bound for \(M_D(A)\) for any subset \(A\) with \(|A| \geq \alpha N\). For any \(A\) such that can make \(M_D(A) = f_D(N, \alpha)\), we must have \(|A| = \lceil \alpha N \rceil\). By substituting this into our expression, we get the lower bound. Notice that if \(D\alpha \to \infty\), the lower bound in (3.1) has main term \(\alpha^2N\). This main term is a decreasing function of \(D\). Therefore, under the assumption that \(D\alpha \to \infty\), it is better to use the lower bound in (3.1) with \(D = o(N)\) and it gives (3.2). Since function \(f_D(N, \alpha)\) is an increasing function of \(D\), with the assumption that \(D\alpha \to \infty\), (3.2) holds no matter \(D = o(N)\) holds or not. \(\square\)

4. Proof of the upper bound in Theorem 1.1 (3)

In this section we study the upper bound of \(f_D(N, \alpha)\) with given density \(\alpha\) and \(N\). We consider the case \(D = N\), i.e., we can regard it as restriction free on the range of differences of any two distinct elements in \(A\).

Let \(f_N(\alpha)\) simplify the notation \(f_D(N, \alpha)\) we defined before in the case \(D = N\).

**Theorem 4.1.** Let \(\alpha \in (0, 1)\). If \(\alpha^2N \to \infty\) as \(N \to \infty\), then
\[
f_N(\alpha) = (1 + o(1))\alpha^2N.
\]

In the above condition, we need to prove the result for \(\alpha < 1 - \epsilon\) for any positive \(\epsilon\) smaller than one. Aiming for a simple expression and without loss of generality, we will only prove the case \(\alpha < 1/2\) for the upper bound. The upper bound here can be proved by the deterministic method. It also can be proved by the probabilistic method, however, a shortcoming of probabilistic method here is we can not always avoid the non-essential needed restriction on the range of density \(\alpha\) that it be much larger than \(N^{-1/2}(\log N)^{-1/2}\); at least the author does not know how to avoid this. So in this paper, we use the deterministic approach to prove asymptotic formulas for the function \(f\) with loose restriction on \(\alpha\). Before we prove the upper bound in the case of integers, we first prove the following theorem which gives us the results in cyclic groups and would imply the results in \(\mathbb{Z}\).

**Theorem 4.2.** Let \(N = p^2 s\) for some positive integer \(s\) and prime number \(p\). Let \(\mathbb{Z}/N\mathbb{Z}\) be a cyclic group and \(\alpha \in (0, 1)\). We define
\[
f_N(\alpha) := \min_{A \subseteq \mathbb{Z}/N\mathbb{Z}, |A| \geq \alpha N} \max_{d \neq 0} r_{A \setminus A}(d).
\]
If $p, s, p\alpha \to \infty$, then
\[ f_N(\alpha) = (1 + o(1))\alpha^2 N. \]

On the other side, there is a similar result proved in generalized Sidon sets. See \[1, \text{Theorem 1.6}\]. We state it by using our notation.

**Theorem 4.3.** Let $\mathbb{Z}/N\mathbb{Z}$ be a cyclic group and $\alpha \in (0, 1)$. We define
\[ g_N(\alpha) := \min_{A \subseteq \mathbb{Z}/N\mathbb{Z}, |A| \geq \alpha N} \max_d r_{A+A}(d). \]
Then as $N$ tends to infinity,
\[ \liminf_{N \to \infty} \frac{g_N(\alpha)}{\alpha^2 N} = 1, \]
Here
\[ r_{A+A}(d) = \{(a, a') \in A \times A : d = a + a'\}. \]
Similarly, we have the following in the case of difference sets.
\[ \liminf_{N \to \infty} \frac{f_N(\alpha)}{\alpha^2 N} = 1. \quad (4.1) \]

**Proof of (4.1).** The lower bound of $f_N(\alpha)$ can be regarded as a corollary of Theorem 3.1. We prove it by a simple counting argument here. We still use $M_N(A)$ to denote the maximum value of $r_{A-A}(x)$ for all possible non-zero $x$ in the cyclic group context. There are $|A|^2$ number of pairs $(a, a') \in A \times A$ but $|A|$ of them satisfy $a = a'$. So we have the following
\[ |A|^2 - |A| \leq (N - 1)M_N(A). \]
By noting the fact that $|A| \geq \alpha N$, and letting $N$ tend to infinity, we have the following when $N$ is sufficiently large.
\[ f_N(\alpha) \geq \alpha^2 N - 1. \]
This gives the lower bound of $f_N(\alpha)$. Combining this with Theorem 4.2, we complete the proof. \hfill \Box

To prove Theorem 4.2, we first do the construction in $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ where $p$ is a prime and then move on to $\mathbb{Z}/q\mathbb{Z}$ with a special type of $q$. The argument here is similar to [1, Section 3, 4]. Finally, the results in the cyclic group will give us the results in $\mathbb{Z}$ that we are aiming for.

### 4.1. Construction in $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$

We prove that there is a set $A$ such that $r_{A-A}$ has a relatively small upper bound and $|A|$ is reasonably large in $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

**Proposition 4.4.** Given a positive integer $k$, for any odd prime number $p \geq 2k$ and any $(a, b) \not\equiv (0, 0) (\text{mod } p)$, there is a set $A \subseteq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ with $|A| = kp - k + 1$ such that $r_{A-A}(a, b) \leq [k^2 + 7k^{7/4}]$.

In the context $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we do our construction by pasting several disjoint sets together. The main techniques will be used are properties of quadratic equations. First we define the set which plays an important role in this section. For $u \not\equiv 0 (\text{mod } p)$, define
\[ A_u := \{(x, x^2/u) : x \in \mathbb{Z}/p\mathbb{Z}\}. \quad (4.2) \]
For any \((a, b) \in \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}\), we use the following notation to denote the representation function. For \(uv \not\equiv 0 \pmod{p}\), define
\[
 r_{A_u - A_v}(a, b) := \# \{(x, y) : a \equiv x - y, b \equiv x^2/u - y^2/v \pmod{p}\}.
\]
This is exactly the number of solutions to the equations
\[
\begin{align*}
 a &\equiv x - y \pmod{p} \\
 b &\equiv x^2/u - y^2/v \pmod{p}.
\end{align*}
\]

In the next lemma we study the possible values of \(r_{A_u - A_v}(a, b)\), where \((a, b) \not\equiv (0, 0) \pmod{p}\).

**Lemma 4.5.** Suppose \(r_{A_u - A_v}(a, b)\) is defined as above, \(p\) is an odd prime and \((a, b) \not\equiv (0, 0) \pmod{p}\).

1. If \(u \equiv v \pmod{p}\), then \(r_{A_u - A_v}(a, b) \leq 1\).
2. If \(u - v \equiv u' - v' \pmod{p}\) and \((uvu') \not\equiv 0 \pmod{p}\), then \(r_{A_u - A_v}(a, b) + r_{A_{u'} - A_{v'}}(a, b) = 2\) for all \((a, b) \not\equiv (0, 0) \pmod{p}\).

The symbol \((-)\) we use in (2) is Legendre symbol.

**Proof.** 1) In the case \(u \equiv v \pmod{p}\), it is clear that if \(a \equiv 0 \pmod{p}\) then there is no solution since we required that \((a, b) \not\equiv (0, 0) \pmod{p}\). Otherwise we have \(x + y \equiv uba^{-1} \pmod{p}\) which leads to a unique solution with the fact that \(x - y \equiv a \pmod{p}\).

2) In the case that \(u - v \equiv u' - v' \pmod{p}\) and \((uvu') \not\equiv 0 \pmod{p}\), we have \(u \not\equiv v \pmod{p}\) and \(u' \not\equiv v' \pmod{p}\). It means that we are studying proper quadratic equations in this case. Substituting \(y \equiv x - a \pmod{p}\) into \(b \equiv x^2/u - y^2/v \pmod{p}\), we have the quadratic equation
\[
(u - v)x^2 - 2aux + ua^2 + bwv \equiv 0 \pmod{p}.
\]
and the discriminant is
\[
\Delta \equiv 4uv(a^2 - (u - v)b) \pmod{p}.
\]
The number of solutions can be written as
\[
r_{A_u - A_v}(a, b) = \begin{cases} 
 1 & \text{if } (\Delta_p) = 0. \\
 2 & \text{if } (\Delta_p) = 1. \\
 0 & \text{if } (\Delta_p) = -1.
\end{cases}
\]
This can be simplified as
\[
r_{A_u - A_v}(a, b) = 1 + (\Delta_p).\]

To calculate the sum of \(r_{A_u - A_v}(a, b)\) and \(r_{A_{u'} - A_{v'}}(a, b)\), we need to calculate the corresponding sum \((\Delta_p) + (\Delta'_p)\). Notice that the product of these two is
\[
\frac{\Delta}{p} \cdot \frac{\Delta'}{p} = \left(16uvu'(u - v)b - a^2)((u' - v')b - a^2) \right) = -\left(\frac{(u - v)b - a^2}{p}\right)^2.
\]
In the last step we used the fact that \(u - v \equiv u' - v' \pmod{p}\) and \((uvu') \not\equiv 0 \pmod{p}\). Now there are two possibilities. If \((u - v)b - a^2 \equiv 0\), then both \((\Delta_p)\) and
equal to zero. If \((u - v)b - a^2 \neq 0\), then the product of \(\left(\frac{\Delta}{p}\right)\) and \(\left(\frac{\Delta'}{p}\right)\) equals to \(-1\). These two cases give us the same result

\[ \left(\frac{\Delta}{p}\right) + \left(\frac{\Delta'}{p}\right) = 0, \]
which means that \(r_{A_n - A_n'}(a, b) + r_{A_n' - A_n}(a, b) = 2\).

In the next lemma, the main technique we use is Weil’s Theorem.

**Lemma 4.6.** Suppose \(0 \leq n \leq k - 1\), \(1 \leq |l| \leq k - 1\), \(1 \leq i, j \leq k\) and \(p\) is an odd prime number. Let \(S_n\) define the sum

\[ S_n := \sum_{1 \leq |l| \leq k - 1} \left| \sum_{i - j = l} \left(\frac{(n + i)(n + j)}{p}\right) \right|. \]

Then we have the following upper bound for the sum of \(S_n\)

\[ \sum_{0 \leq n \leq p - 1} S_n \leq \sqrt{2p^2 k^2 (2k - 1) + 8p^{3/2} k^4}. \]

To estimate the sum, we first use the Cauchy inequality to rewrite the sum which roughly gives us a function with power 4 in the numerator of Legendre symbol, and then apply the Weil’s Theorem.

**Proof.** First we use Cauchy inequality to get the following estimation.

\[ \sum_{0 \leq n \leq p - 1} S_n = \sum_{n,l} \left| \sum_{i - j = l} \left(\frac{(n + i)(n + j)}{p}\right) \right| \leq \sqrt{2kp \sum_{n,l} \left( \sum_{i - j = l} \left(\frac{(n + i)(n + j)}{p}\right) \right)^2}. \]

The right hand side above can be written as

\[ \sqrt{2kp \sum_{i - j = l'} \sum_{i' - j' = l} \left(\frac{(n + i)(n + j)(n + i')(n + j')}{p}\right)}. \]

Let \(f(n) = (n + i)(n + j)(n + i')(n + j')\). We estimate the quantity above by using Weil’s Theorem.

1) If \(f(n)\) is a constant multiple of a square, then we bound the value \(\left(\frac{f(n)}{p}\right)\) trivially by 1. The number of quadruples \((i, j, i', j')\) that can make \(f(n)\) a constant multiple of a square is bounded by \(k(k - 1) + k^2\). The reasons are the following. First notice that if the restriction on the quadruple is \(i + j' = i' + j\), then we have two possible cases, either \(i = i'\) and \(j = j'\), or \(i = j\) and \(i' = j'\). These two cases both give \(k^2\) choices but there is an overlap that all of them are equal, i.e., we need to subtract \(k\) choices. In total, we have the following bound

\[ \sum_{f \text{ is square}} \sum_n \left(\frac{f(n)}{p}\right) \leq pk(2k - 1). \]

2) If \(f(n)\) is not a constant multiple of a square, we can use Weil’s Theorem. For each quadruple \((i, j, i', j')\), we have the following,

\[ \sum_{0 \leq n \leq p - 1} \left(\frac{f(n)}{p}\right) \leq (\deg f)\sqrt{p} = 4\sqrt{p}. \]
The total number of quadruples under the restriction $i-j = i'-j'$ is bounded by $k^3$. We also use this bound as our bound for the number of quadruples which make $f$ satisfy Weil’s Theorem. Then we have the following

$$\sum_{f \text{ isn’t a square}} \sum_n \left( \frac{f(n)}{p} \right) \leq 4k^3 \sqrt{p}.$$  

Combining case 1) and case 2), we have

$$\sum_{0 \leq n \leq k-1} S_n \leq \sqrt{2pk(pk(2k-1) + 4k^3 \sqrt{p})} = \sqrt{2p^2k^2(2k-1) + 8pk^3/2k^4}.$$  

□

Now we will use the above two lemmas to prove Proposition 4.4.

**Proof of Proposition 4.4.** First we point out how the set $A$ looks. Let $A_u$ be the same as what we defined in (4.2), and $A$ be defined as follows:

$$A := \bigcup_{n+1 \leq u \leq n+k} A_u.$$  

The crucial point is to find a suitable $n$ which makes $A$ be the set we are searching for.

First notice that the size of $A$ is independent of $n$. And the size of $A$ is $k(p - 1) + 1$ since $A_u$ and $A_v$ only intersect at point $(0, 0)$ for any $u \neq v$.

Next we study the representation function $r_{A-A}$. From the definition of $A$, we can have a trivial upper bound. For any $(a, b)$ non-zero, we have

$$\sum_{n+1 \leq u, v \leq n+k} r_{A_u-A_v}(a, b) \leq k.$$  

To study the right hand side above, we first parametrize the variables above. Write $u = n + i$, $v = n + j$, then $i, j \in \{1, 2, 3, \ldots, k\}$. Let $l = i - j = u - v$, we have the following,

$$|l| \leq k - 1.$$  

Fixing the value of $l$, we have $k - |l|$ pairs of $(i, j)$, i.e., $k - |l|$ pairs of $(u, v)$. Now we are going to use Lemma 4.5.

1) If $l = 0$, from the first part of the lemma, we have the bound

$$\sum_{u=v} r_{A_u-A_v}(a, b) \leq k.$$  

2) If $|l| \geq 1$, we consider the value of $\left( \frac{uv}{p} \right)$ for each pair, $(u, v)$. Let $\alpha(l) := \#\{(u, v) : u - v = l, \left( \frac{uv}{p} \right) = 1\}$, $\beta(l) := \#\{(u, v) : u - v = l, \left( \frac{uv}{p} \right) = -1\}$. Then we have the following fact

$$\alpha(l) + \beta(l) = k - |l|, \quad \alpha(l) - \beta(l) = \sum_{u=v=l} \left( \frac{uv}{p} \right).$$  

Our strategy is to match pairs into pairs according to their values under Legendre symbol. Precisely, if two pairs have different signs, we can match them and then use the second part in Lemma 4.5 to estimate the sum. Since the values of $\alpha(l)$ and $\beta(l)$ are not necessarily equal, for the rest of the pairs
which have not been matched with other pairs, we trivially bound their function values by 2. Hence, we have

\[
\sum_{1 \leq |l| \leq k} r_{A_u-A_v}(a, b) \leq 2 \min\{\alpha(l), \beta(l)\} + 2(\max\{\alpha(l), \beta(l)\} - \min\{\alpha(l), \beta(l)\}).
\]

The above actually gives us the result \(2 \max\{\alpha(l), \beta(l)\}\), i.e., \(\alpha(l) + \beta(l) + |\alpha(l) - \beta(l)|\), which can be further written as

\[
k - |l| + \left| \sum_{u-v=l} \left( \frac{uv}{p} \right) \right|.
\]

Summing the above expression over \(l\) where \(1 \leq |l| \leq k-1\) and considering the obvious bound that we have showed for \(l = 0\) in case 1), we have the following

\[
\sum_{|l| \leq k-1} r_{A_u-A_v}(a, b) = k + \sum_{1 \leq |l| \leq k-1} r_{A_u-A_v}(a, b) = k^2 + \sum_{1 \leq |l| \leq k-1} \left| \sum_{u-v=l} \left( \frac{uv}{p} \right) \right|.
\]

Notice that the second part on the right hand side above is exactly what we defined in Lemma 4.6, i.e., the upper bound now is

\[
k^2 + S_n. \tag{4.3}
\]

Since \(uv \not\equiv 0 \pmod{p}\), we have the following upper bound for \(n\),

\[
n + k \leq p - 1 \iff n \leq p - k - 1.
\]

We define \(n^*\) to be the smallest integer index among \(\{0, 1, 2, \cdots, p-k-1\}\) such that

\[S_{n^*} = \min\{S_n : 0 \leq n \leq p-k-1\}\]

Then we can see that \(n^*\) is uniquely defined. By using Lemma 4.6, we have

\[
S_{n^*} \leq \sum_{0 \leq n \leq p-k-1} \frac{S_n}{p-k} \leq \sum_{0 \leq n \leq p-1} \frac{S_n}{p-k} = \frac{\sqrt{2p^2k^2(2k-1) + 8p^{3/2}k^4}}{p-k}.
\]

This bound is smaller than

\[
\frac{2\sqrt{p^2k^3 + 2p^{3/2}k^4}}{p-k}.
\]

Write \(p = \lambda k\), where \(\lambda \geq 2\), then the above expression is a decreasing function with respect to \(\lambda\). So we can further get a bound

\[
S_{n^*} < 7k^{7/4}. \tag{4.4}
\]

By substituting (4.4) into (4.3), we get the upper bound

\[
k^2 + 7k^{7/4}.
\]

Hence, we have found a set \(A\) with \(kp-k+1\) elements and \(r_{A-A}\) is upper bounded by \([k^2 + 7k^{7/4}]\). The set \(A\) can be represented as following

\[
A = \bigcup_{n^*+1 \leq u \leq n^*+k} A_u.
\]
4.2. Construction in $\mathbb{Z}/q\mathbb{Z}$ with $q = p^2s$. We construct a large set $A'$ in a certain cyclic group. The strategy is to project the set $A$ constructed in $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ onto $\mathbb{Z}/q\mathbb{Z}$.

**Lemma 4.7.** Suppose that $A$ is a subset of $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ with size $m$, such that $r_{A-A}$ is upper bounded by $h$. Then there exists a subset $A'$ in $\mathbb{Z}/p^2s\mathbb{Z}$ with size $|A'| = ms$ such that $r_{A-A'}$ is upper bounded by $h(s+1)$.

**Proof.** Notice that any element in $A$ has two coordinates and we choose the natural way to write them as integers in $[0, p - 1]$. Now we use set $A$ to define set $A'$ as following

$$A' = \{x = a + cp + bsp : (a, b) \in A, 0 \leq c \leq s - 1\}.$$  

From the definition, we have $|A'| = |A| \times s = ms$. We show that $r_{A'-A'}$ is bounded by $h(s + 1)$. In other words, we need to study the representation function $r_{A-A}(x)$ for non-zero $x \in \mathbb{Z}/p^2s\mathbb{Z}$. Given $x = a + cp + bsp$, then $a, b, c$ are fixed. Write $x$ as the difference of two elements in $A'$,

$$a + cp + bsp \equiv (a_1 + c_1p + b_1sp) - (a_2 + c_2p + b_2sp) \pmod{p^2s}. \quad (4.5)$$

From the above we can get

$$a \equiv a_1 - a_2 \pmod{p}.$$  

There are two possible values of the difference $a_1 - a_2$. Either $a_1 - a_2 = a$ or $a_1 - a_2 = a - p$. For convenience, we write $a_1 - a_2 = a - \delta p$ where $\delta = 0, 1$. This also tells us that $\delta = (a + a_2 - a_1)/p$ is fixed once $a_1, a_2$ are given.

Now we substitute this expression into (4.5). We can derive that

$$c + bs \equiv -\delta + c_1 - c_2 + (b_1 - b_2)s \pmod{ps}. \quad (4.6)$$

Again, we have the following from (4.6)

$$c \equiv -\delta + c_1 - c_2 \pmod{s}. \quad (4.7)$$

Thus we have $c = -\delta + c_1 - c_2$ or $c = -\delta + c_1 - c_2 + s$. We write this as:

$$c - \eta s = -\delta + c_1 - c_2 \text{ where } \eta = 0, 1. \quad (4.8)$$

We substitute this into (4.6) and we get

$$b \equiv -\eta + b_1 - b_2 \pmod{p}.$$  

Now we have the expression of the element $(a, b)$ in $A$,

$$(a, b) = (a_1, b_1) - (a_2, b_2) + (\delta, -\eta).$$

We also have the following form

$$(a, b) \equiv (a_1, b_1) - (a_2, b_2) + (0, -\eta) \pmod{p}.$$  

We have that $r_{A-A}$ is upper bounded by $h$. Once we have fixed $\eta$, the number of choices for $a_1, a_2, b_1, b_2$ is no more than $h$, and $\delta$ is also determined. So now we only need to consider the number of choices for $c_1, c_2$ once $a_1, a_2, b_1, b_2, \delta$ are fixed. We come back to Equation (4.8) and discuss the possible cases in terms of the values of $\eta$.

1) If $\eta = 0$, we have $c_1 = c_2 + c + \delta \geq c$, which leads to at most $s - c$ possibilities.

2) If $\eta = 1$, we have $c_1 = c - s + \delta + c_2 \leq c - s + 1 + s - 1 = c$, which leads to at most $c + 1$ possibilities.
In total, the above will give us at most \( h(s - c + c + 1) = h(s + 1) \) possibilities. \( \square \)

4.3. **Conclusion for cyclic group.** First we combine Proposition 4.4 and Lemma 4.7 to get the following.

**Proposition 4.8.** For any positive integers \( k, s \), and any positive odd prime integer \( p \geq 2k \), there is a set \( A \subseteq \mathbb{Z}/p^2 \mathbb{Z} \) with size \((kp - k + 1)s\) such that \( r_{A-A} \) is upper bounded by \( (k^2 + 7k^{7/4})(s + 1) \).

Now we can prove Theorem 4.2.

**Proof of Theorem 4.2.** The lower bound is proved in the proof of (4.1). Now we prove the upper bound. For any positive \( \alpha < 1/2 \), choose \( k \) to be the integer such that \( k = \lceil p\alpha \rceil + 1 \). Since \( p\alpha \) tends to infinity,

\[
\frac{k}{p} = (1 + o(1))\alpha.
\]

From Proposition 4.8, we know there exists a subset \( A \subseteq \mathbb{Z}/\mathbb{Z} \) with size \((kp - k + 1)s\) such that the representation function \( r_{A-A} \) is upper bounded by \( (k^2 + 7k^{7/4})(s + 1) \). First we consider the density of \( A \),

\[
\frac{(kp - k + 1)s}{p^2} \geq \frac{kp - k + 1}{p^2} \geq \frac{k - 1}{p} \geq \alpha.
\]

Then we consider the expression of the upper bound for \( r_{A-A} \),

\[
(k^2 + 7k^{7/4})(s + 1) = (1 + o(1))(k^2/p^2)N = (1 + o(1))\alpha^2N.
\]

The expression is valid since we let \( s \) and \( k \) tend to infinity and Equation (4.9) holds. So the above has showed that for a special type of integers, \( N \), \( f_N(\alpha) \) is upper bounded by \((1 + o(1))\alpha^2N\). This completes the proof. \( \square \)

Now, we can easily prove Theorem 4.1

**Proof of Theorem 4.1.** For any \( N \) in the form \( N = p^2 s \) for some suitable prime \( p \) and integer \( s \), let set \( A \subseteq \mathbb{Z}/\mathbb{Z} \) be the set in Theorem 4.2. We can also view it as a subset of \([N]\) in the natural way. To distinguish, we use \( r'_{A-A} \) to denote the representation function in \( \mathbb{Z}/\mathbb{Z} \). Then a natural inequality is

\[
r_{A-A}(x) \leq r'_{A-A}(x) \leq (1 + o(1))\alpha^2N.
\]

For \( N \) not in the form \( N = p^2s \), we choose \( p = o(\sqrt{N}) \) to be a prime number in the interval \([P, 2P]\) for some integer \( P \). This is valid once \( P \) is sufficiently large. Also, we require that \( p\alpha \) tends to infinity. This is valid by our assumption that \( N\alpha^2 \) tends to infinity as \( N \) tends to infinity. Choose \( s = \left\lfloor \frac{N}{p} \right\rfloor \) and notice that \( k = O(p\alpha) \), then both \( k \) and \( s \) tend to infinity as \( N \) tends to infinity. We have \( p^2s < N < p^2(s + 1) \) from the choice of \( s \). From the conclusion for \( p^2s \) in Theorem 4.2, we know for any \( \alpha < 1/2 \), there exists a set \( A \) such that \( |A| \geq \alpha p^2s \) and \( r_{A-A} \leq (1 + o(1))\alpha^2 p^2s \). Regarding \( A \) as a subset of \([N]\) in the natural way, we have \( |A| \geq \alpha \times \frac{N}{p^2} \) and \( r_{A-A} \leq (1 + o(1))\alpha^2N \). Write \( \alpha' = \alpha \times \frac{s}{p^2}, \) then the result can be restated as

\[
|A| \geq \alpha'N, \quad r_{A-A} \leq (1 + o(1)) \left( \frac{s + 1}{s} \right)^2 \alpha^2N.
\]
The right hand side can be written as $(1 + o(1))\alpha^2 N$ once we let $s$ tend to infinity. The above fact gives us the upper bound in Theorem 4.1. Hence, we complete the proof.

□

References

[1] J. Cilleruelo, I. Ruzsa, and C. Vinuesa. Generalized Sidon sets. *Adv. Math.*, 225(5):2786–2807, 2010.
[2] P. Erdős and P. Turán. On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.*, 16:212–215, 1941.
[3] B. Green. The number of squares and $B_h[g]$ sets. *Acta Arith.*, 100(4):365–390, 2001.
[4] Oleksiy Klurman and Marc Munsch. Distribution of factorials modulo $p$. *J. Théor. Nombres Bordeaux*, 29(1):169–177, 2017.

Department of Mathematics, University College London, Gower Street, London, WC1E 6BT, United Kingdom

*E-mail address: wenqiang.xu@ucl.ac.uk*