1. Introduction

It is a well-known problem to give an estimate for the largest clique of the Paley-graph, i.e., to give an estimate for $|A|$ if $A \subset F_p$ ($p \equiv 1 \pmod{4}$) is such that $A - A = \{a - a' \mid a, a' \in A\}$ avoids the set of quadratic nonresidues. In this paper we will study a much simpler problem namely when $A - A$ is substituted by the set $FS(A) = \{\sum \varepsilon_a a \mid \varepsilon_a = 0 \text{ or } 1 \text{ and } \sum \varepsilon_a > 0\}$. In other words we will estimate the maximal cardinality of $A \subset F_p$ if $FS(A)$ avoids the set of quadratic nonresidues. We will show that this problem is strongly related to the problem of the estimation of the least quadratic nonresidue. If $n(p)$ denotes the least quadratic nonresidue then the set $\{1, 2, \ldots, [n(p)/2]\}$ satisfies the conditions, this already gives a lower bound for the maximal value of $|A|$. Later we will prove that the maximal value of $|A|$ is $\Omega(\log \log p)$. On the other hand we will prove that $|A| = O(n(p) \log^3 p)$. The proof is based on the fact that if $t$ is a quadratic nonresidue then $FS(A) \cap t \cdot FS(A) = \emptyset$ or $\{0\}$ where by definition $t \cdot B = \{tb \mid b \in B\}$. We will show that if $t$ is small than $|FS(A)|$ is much greater than $|A|$. In the next section we will study the case when $t = n(p) = 2$. In the third part we will prove the upper bound $|A| = O(n(p) \log^3 p)$. In the last part we will show that the maximal value of $|A|$ is $\Omega(\log \log p)$.

2. The case $n(p) = 2$

In this part we will study the case $n(p) = 2$. In this case $FS(A) \cap 2 \cdot FS(A) = \emptyset$ or $\{0\}$. At first we consider the case $FS(A) \cap 2 \cdot FS(A) = \emptyset$.

**Theorem 2.1.** If $FS(A) \cap 2 \cdot FS(A) = \emptyset$ then $|FS(A)| = 2^{|A|}$.

**Proof** We have to show that if $FS(A) \cap 2 \cdot FS(A) = \emptyset$ then all the subset sums are different. Indeed, if there were two different sums with the same value then omitting the intersection we got that $s = a_{i_1} + a_{i_2} + \cdots + a_{i_t} = a_{j_1} + \cdots + a_{j_m} (i_u \neq j_v)$. In this case $s$ and $2s = a_{i_1} + a_{i_2} + \cdots + a_{i_t} + a_{j_1} + \cdots + a_{j_m}$ would be the elements of $FS(A)$, which contradicts the condition.

A trivial consequence of Theorem 1 is

**Corollary 2.2.** If $n(p) = 2$ (i.e. $(\frac{2}{p}) = -1$) and every element of $FS(A)$ is a quadratic residue then $|A| \leq \frac{\log p}{\log 2}$.

\[2000 \text{ Mathematics Subject Classification. Primary: 11B75.} \]
\[\text{Key words and phrases. Subset sums, quadratic residues.}\]
Theorem 2.3. Assume that $0 \notin A$. If $FS(A) \cap 2 \cdot FS(A) = \emptyset$ or $\{0\}$ then $|A| \leq \frac{2}{\log 2} \log p$.

Remark 1. Assuming that $0 \notin A$ is just a simplifying condition, if we leave out the 0 from $A$ then $FS(A)$ will not change and the cardinality of $A$ will only decrease by 1.

Proof. We will say that $\sum_{i \in I} a_i = a$ is an irreducible $a$-sum if there is no $\emptyset \neq J \subset I$ for which $\sum_{i \in J} a_i = 0$. Two irreducible $a$-sums have to be disjoint because if $\sum_{i \in I_1} a_i = \sum_{j \in I_2} a_j$ then $\sum_{i \in I_1 \cup I_2} a_i = s \neq 0$ and $s, 2s \in FS(A)$ contradicts the condition. On the other hand in case $a \neq 0$ there cannot be two disjoint irreducible $a$-sums. Thus we only get an $a$-sum as the sum of "the" irreducible $a$-sum and a 0-sum. We only get a 0-sum as the sum of irreducible 0-sums so the number of the 0-sums is at most $\frac{2^{|A|}}{2}$ since every irreducible 0-sum has at least two elements (here we have used the simplifying condition that 0 is not in $A$). Hence $p \cdot 2^{|A|/2} \geq 2^{|A|}$ whence $|A| \leq \frac{2}{\log 2} \log p$. □

Corollary 2.4. If $n(p) = 2$ and every element of $FS(A)$ is a square then $|A| \leq \frac{2}{\log 2} \log p$.

Corollary 2.5. If $A \subset \{1, 2, \ldots, N\}$ and every element of $FS(A)$ is a perfect square then $|A| = O(\log \log N)$.

Proof. We will use Gallagher’s larger sieve. Let $y = 20 \log N \log \log N$ and let $S = \{p \leq y| p \text{ prime } p \equiv 3 \text{ or } 5 \pmod{8}\}$. By Corollary 2, $\nu(p) \leq \frac{2}{\log 2} \log p$ for these primes $p$. By the larger sieve

$$|A| \leq \sum_{p \in S} \frac{\Lambda(p)}{\nu(p)} - \log N$$

if the denominator is positive. We have

$$\log y \leq 2 \log \log N$$

if $N$ is large enough. Furthermore

$$\sum_{p \in S} \Lambda(p) = \frac{1}{2} y + o(y)$$

and

$$\sum_{p \in S} \frac{\Lambda(p)}{\nu(p)} \geq \frac{y}{4 \log y} + o\left(\frac{y}{\log y}\right) \geq \frac{y}{5 \log y}$$

if $y$, thus also $N$ is large enough. Hence for large $N$,

$$\sum_{p \in S} \frac{\Lambda(p)}{\nu(p)} \geq \frac{20 \log N \log \log N}{10 \log \log N} = 2 \log N.$$ 

Thus $|A| \leq 20 \log \log N$. □
3. Upper bound

At first we will prove a theorem on Abelian groups from which the upper bound follows.

**Theorem 3.1.** Let $A \subset G$ where $G$ is a finite Abelian group. Assume that $|A| \geq 20000t \log^3 |G|$. Then there exists a $d \neq 0$ for which \{d, 2d, \ldots, td\} $\subset FS(A)$.

**Proof.** We prove by contradiction. Assume that there exists a set $A$ for which $|A| = n > 20000t \log^3 |G|$ such that $FS(A)$ does not contain a set \{d, 2d, \ldots, td\} where $d \neq 0$. We can also assume that $0 \notin A$. Let $r$ be a fixed positive integer which we will choose later. We will use the Erdős-Rado theorem on $\Delta$-systems.

**Lemma 3.2.** (Erdős-Rado) Assume that the $r$-uniform hypergraph has more than $r!(t-1)^r$ edges, then it contains a $\Delta$-system with more than $t$ elements, i.e., a set system $A_1, A_2, \ldots, A_l$ such that $A_k \cap A_l = \cap_{j=1}^t A_j$ for all $1 \leq k < l \leq t$.

Again at first we will give an upper bound for the number of irreducible sums. (We recall that a $\sum_{a \in J} a$ sum is irreducible if there is no $J \subset I$ nonempty set such that $\sum_{a \in J} a = 0$, and we call a sum irreducible $a$-sum if it is irreducible and its value is $a$). We estimate the number of $r$-term irreducible $a$-sums. If $a \neq 0$ then there exist at most $r!(t-1)^r$ $r$-term irreducible $a$-sums, indeed, otherwise these sums as sets contain a $\Delta$-system with $t$ elements by the lemma. If we leave out the intersection of these sets we get $t$ disjoint sums having the same nonzero value since these were irreducible sums. Let $d$ be the value of these sums then adding together some of these disjoint sums we get that for this $d \neq 0$ we have \{d, 2d, \ldots, td\} $\subset FS(A)$ contradicting the indirect assumption. This argument cannot be applied for $a = 0$ immediately since it may occur that $t$ disjoint irreducible $r$-term sums form the $\Delta$-system. Although we can easily solve this problem, now we can say that there are at most $n(r-1)!(t-1)^{r-1}$ irreducible 0-sums since if there are more irreducible 0-sums then there is an element $a \in A$ which is contained in more than $(r-1)!(t-1)^{r-1}$ irreducible sums as a summand. Omitting $a$ from these sums we get the previous case with $(r-1)$-term sums instead of $r$, since these new sums have $-a$ value which is not 0 by $0 \notin A$ and irreducible since a subsum of an irreducible sum is still irreducible.

Now we give an upper bound for the number of $r$-term $a$-sums. Every $a$-sum is a sum of an irreducible $a$-sum and some irreducible 0-sums (this is, of course, not unique, but it is not a problem since we only give an upper bound). Let us consider those representations where the irreducible $r$-term $a$-sum has $k_1$ terms and the irreducible 0-sums have $k_2, \ldots, k_m$ terms, respectively. According to the previous argument the number of these sums is at most

$$k_1!(t-1)^{k_1}n(k_2-1)!(t-1)^{k_2-1} \ldots n(k_m-1)!(t-1)^{k_m-1} \leq$$

$$\leq \prod_{i=1}^m (n(k_i-1)!(t-1)^{k_i-1}) = n^m \prod_{i=1}^m (k_i-1)! (t-1)^{r-m}$$

where $\prod_{i=1}^m (k_i-1)!$ is the product of the factorials of all the terms up to $m$. This bound is not tight, but it is sufficient for our purposes.
since \( \sum_{i=1}^{m} k_{i} = r \) and we will choose \( r \) later so that \( k_{1}(t - 1) \leq r(t - 1) \leq n \).

We will show that

\[
n^{m}(\prod_{i=1}^{m} (k_{i} - 1)!)(t - 1)^{r - m} \leq r^{r/2}n^{r/2+1}(t - 1)^{r/2}.
\]

Indeed, since every irreducible 0-sum has at least two elements (again we use the fact that \( 0 \notin A \)) \( m - 1 \leq r/2 \) and \( n^{r/2+1-m} \geq (r(t - 1))^{r/2+1-m} \). Hence

\[
r^{r/2}n^{r/2+1}(t - 1)^{r/2} \geq r^{r/2}n^{m}(r(t - 1))^{r/2+1-m}(t - 1)^{r/2} \geq n^{m}r^{r-m}(t - 1)^{r-m} \geq n^{m}(\prod_{i=1}^{m} (k_{i} - 1)!)(t - 1)^{r-m}
\]

since \( \prod_{i=1}^{m} (k_{i} - 1)! \leq (r - m)! \leq r^{r-m} \). We can decompose \( r \) into positive integers in \( p(r) \) ways where \( p(r) \) denotes the number of partitions of \( r \). Thus every \( a \in G \) can be represented as a sum of \( r \) elements of \( A \) in at most \( p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2} \) ways. Since there are \( \binom{n}{r} \) \( r \)-term sums we have

\[
\binom{n}{r} \leq |G| \cdot p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2}.
\]

We will choose \( r \) so that

\[
\frac{\binom{n}{r}}{p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2}}
\]

is nearly maximal. For two consecutive \( r \)'s consider the fraction

\[
\frac{\binom{n}{r}}{p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2}} : \frac{\binom{n}{r+1}}{p(r+1)(r+1)(r+1)^{r/2}n^{r+1}/2(t - 1)^{(r+1)/2}} = \frac{r + 1}{n - r} p(r+1) \left(1 + \frac{1}{r}\right)^{r/2}(n+1)(t - 1)^{1/2}.
\]

For the best choice of \( r \) this must be approximately 1. Let us choose \( r = [n^{1/3} : e(t - 1)^{1/3}] \), up to a constant factor this is the best choice. Now we can use the elementary estimates \( m(n/e)^{m} > m! > (n/e)^{m} \) which is valid for \( m \geq 6 \):

\[
|G| \geq \frac{\binom{n}{r}}{p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2}} \geq \frac{\binom{n}{r}}{r(n - r)(\sqrt{r})^{r}(n-r)p(r)r^{r/2}n^{r/2+1}(t - 1)^{r/2}} = \frac{1}{nr(n - r)p(r)} \left(\frac{n}{n - r}\right)^{n-r} \left(\frac{n^{1/2}}{r^{3/2}(t - 1)^{1/2}}\right)^{r} \geq \frac{1}{|G|^{3}p(r)^{3/2}}.
\]

Here we used the classical fact \( p(r) < \exp\left(\frac{2r}{\sqrt{7r}}\right) < \exp\left(\frac{1}{4r}\right) \). It follows that \( |G|^4 > e^r \) so 4 \( \log |G| \geq r \). Thus 4 \( \log^{3} |G| \geq r^3 > n^{3} \) whence 2000 \( (t - 1) \log^{3} |G| > n \), which contradicts the indirect assumption. \( \square \)

**Remark 2.** The basic idea of this proof can be found in an article of Erdős and Sárközy [3]. In this article the authors study what can be said about the length of an arithmetic progression contained in the set of the subset sums of a subset of \( \{1, 2, \ldots, N\} \).

The statement of the theorem is nearly sharp since the set

\( A = \{t, t + 1, \ldots, [\sqrt{2}t]\} \subset Z_{n} \) with \( t^{3} < n \) shows that there are no two
elements of $FS(A)$ whose quotient is $t$, and $|A| = \Omega(t)$. On the other hand a basis of $\mathbb{Z}_3^d$ shows that the set of subset sums does not contain two elements having the quotient 2, and we have $|A| = \Omega(\log |Z_3|^n)$. Other much trickier examples can be found in the above mentioned article.

**Corollary 3.3.** Let $A \subset F_p$. Assume that $FS(A)$ avoids the quadratic nonresidues. Then $|A| = O(n(p)\log^2 p)$, where $n(p)$ denotes the least quadratic nonresidue.

**Proof.** One can apply Theorem 3. with $t = n(p)$ and get that there exists a $d \neq 0$ for which $d$ and $n(p)d$ are both quadratic residues, which is a contradiction. □

**Remark 3.** If we also assume the condition $0 \notin FS(A)$, i.e., every element of $FS(A)$ is a quadratic residue then $|A| = O(n(p)\log^2(p))$, so that we can win a factor $\log p$ since we need not to estimate the number of irreducible sums, we can apply the Erdős-Rado theorem immediately. On the other hand obviously one can substitute the set of quadratic nonresidues by the set of quadratic residues since one can multiply each element of $A$ with the same quadratic nonresidue and by the construction no element of the subset sums of the new set is a quadratic residue.

**Remark 4.** Since $n(p) = O_\varepsilon(p^{1/2-\varepsilon} \log^{1+c} p)$ [1] thus we get this upper bound also for the maximal value of $|A|$. According to a result of Burgess and Elliot [2], if $g(p)$ denotes the least primitive root modulo $p$ then

$$\frac{1}{\pi(x)} \sum_{p \leq x} g(p) \leq C \log^2 x \log \log^4 x$$

Since $n(p) \leq g(p)$ this shows that in average the maximal value of $|A|$ cannot be greater than $\log^6 p$.

4. **LOWER BOUND**

In this section we will show that the maximal value of $|A|$ is at least $\Omega(\log \log p)$. The proof is based on Weil’s estimation of character sums.

**Theorem 4.1.** There exists an $A \subset F_p$ such that $|A| = \Omega(\log \log p)$ and $FS(A)$ avoids the set of quadratic nonresidues.

First we prove a lemma.

**Lemma 4.2.** Let $Q$ be the set of quadratic residues. Assume that for some set $B$ we have $Q + B = F_p$. Then $|B| \geq \frac{1}{4} \log p$.

**Proof.** Let $B = \{b_1, \ldots, b_k\}$ and $Q_i = Q + b_i$. Then

$$|F_p - \bigcup_{i=1}^k Q_i| = |F_p| - \sum |Q_i| + \sum |Q_i \cap Q_j| - \ldots$$

by the inclusion-exclusion formula.

$$|Q_{i_1} \cap \cdots \cap Q_{i_t}| = \sum_{a} \frac{1}{2^t} \left( 1 + \left( \frac{a - b_{i_1}}{p} \right) \right) \cdots \left( 1 + \left( \frac{a - b_{i_t}}{p} \right) \right) + m(i_1, \ldots, i_t)$$
where $|m(i_1, \ldots, i_l)| \leq \frac{1}{2}$ since it may occur that $a - b_i = 0$. By Weil’s theorem [4]

$$\left| \sum_{n=1}^{p} \left( \frac{f(n)}{p} \right) \right| \leq (t - 1)\sqrt{p}$$

where $f(x) = \prod_{i=1}^{l} (x - a_i)$ and $a_1, \ldots, a_l$ are distinct elements of $F_p$. Multiplying out the product we see that

$$\left(1 + \left( \frac{a - b_i}{p} \right) \right) \cdots \left(1 + \left( \frac{a - b_i}{p} \right) \right) = 1 + \sum \left( \frac{f(a)}{p} \right)$$

where $f$ runs through $2^l - 1$ polynomials of the type considered above. Hence

$$|Q_{i_1} \cap \cdots \cap Q_{i_l}| = \frac{p}{2^l} + m'(i_1, \ldots, i_l)$$

where $|m'(i_1, \ldots, i_l)| \leq \frac{1}{2^l}(2^l - 1)(l - 1)\sqrt{p} + \frac{l}{2}$. Since $l \leq k \leq \sqrt{p}$ (we can assume this inequality, if $k \geq \sqrt{p}$ then we are done), thus $|m'(i_1, \ldots, i_l)| \leq k\sqrt{p}$. It follows that

$$0 = |F_p - \cup_{i=1}^{k} Q_i| = p - \sum_{i=1}^{k} \left( \frac{p}{2} + m'(i) \right) + \sum \left( \frac{p}{4} + m'(i, j) \right) \cdots = p \left( 1 - \frac{1}{2} \right)^{k} + M$$

where $|M| \leq 2^k k\sqrt{p}$. Hence $\frac{p}{2^l} = |M| \leq 2^k k\sqrt{p}$, thus $\sqrt{p} < k4^k < e^{2k}$ so that $k \geq \frac{1}{4} \log p$. □

**Remark 5.** Clearly the same statement holds for the set of quadratic nonresidues $R$.

**Theorem 4.1** There exists a set $A \subset F_p$ for which $|A| = \Omega(\log \log p)$ and $FS(A)$ avoids the set of quadratic nonresidues.

**Proof.** Let us take a maximal set $A$ for which $FS(A)$ avoids the quadratic nonresidues. We will show that $|A| \geq \frac{1}{\log^2} \log \log p - 2$. Let us assume that $|A| \leq \frac{1}{\log^2} \log \log p - 2$. Then $|FS(A)| \leq 2|A| \leq \frac{1}{3} \log p$, thus $R - FS(A) \neq F_p$ so there exists an $s \in F_p$ for which $s \notin R - (a_{i_1} + \cdots + a_{i_l})$ for any $a_{i_1}, \ldots, a_{i_l} \in A$. In this case one can add the element $s$ to $A$, which contradicts the maximality of $A$. Hence $|A| \geq \frac{1}{\log^2} \log \log p - 2$. □

**Remark 6.** There exists a set $B$ for which $|B| = [10 \log p]$ and $Q + B = F_p$.

Let us choose the elements of $B$ in random way with probability $P(b \in B) = \frac{c \log p}{p}$ independently. Then

$$P(x \notin Q + B) = \prod_{i=1}^{(p-1)/2} P(x - i^2 \notin B) = \left( 1 - \frac{c \log p}{p} \right)^{p-1/2}$$

since we have chosen the elements independently. Hence

$$P(Q + B \neq F_p) \leq \sum_{x=0}^{p-1} P(x \notin Q + B) = p \left( 1 - \frac{c \log p}{p} \right)^{p-1/2} \leq pe^{-\frac{1}{2}c \log p}$$
On the other hand, by the Chernoff-inequality [5] we have
\[ P(|B| - c \log p \geq \lambda \sigma) \leq 2 \max(e^{-\lambda^2/4}, e^{-\lambda \sigma/2}) \]
where \( \frac{1}{2} c \log p \leq \sigma^2 = p \frac{\log p}{p} (1 - \frac{\log p}{p}) \leq c \log p \). Choosing \( c = 4 \) and \( \lambda = \sqrt{8 \log p} \) we get that
\[ P(|B| - 4 \log p \geq 4 \sqrt{2 \log p}) \leq 2 e^{-2 \log p} = \frac{2}{p^2}. \]
We have \( p e^{-\frac{1}{2} \log p} = p^{-3/4} \). Since \( \frac{2}{p^2} + \frac{1}{p^{3/4}} < 1 \) for \( p \geq 3 \) thus with positive probability \( |B| \leq 10 \log p \) and \( Q + B = F_p \).
We have shown that in case \( \left( \frac{2}{p} \right) = -1 \) we have \( |FS(A)| = 2^{|A|} \). Thus in general probably one cannot say better than \( \Omega(\log \log p) \), since after the selection of \( |A| - 1 \) elements the set of subset sums has \( 2^{|A| - 1} \) elements and it must not be the additive complement of \( -R \), while the sets with more than \( 10 \log p \) elements are additive complements with high probability.

Acknowledgement. I profited much from discussions with A. Sárközy and K. Gyarmati.

References

Eötvös Loránd University, Department of Algebra and Number Theory, H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
E-mail address: csiki@cs.elte.hu