MAXIMAL RESIDUE DIFFERENCE SETS MODULO $p$

DUNCAN A. BUELL AND KENNETH S. WILLIAMS

Abstract. Let $p \equiv 1 \pmod{4}$ be a prime. A residue difference set modulo $p$ is a set $S = \{a_i\}$ of integers $a_i$ such that $(\frac{a_i}{p}) = +1$ and $(\frac{a_i - a_j}{p}) = +1$ for all $i$ and $j$ with $i \neq j$, where $(\frac{n}{p})$ is the Legendre symbol modulo $p$. Let $m_p$ be the cardinality of a maximal such set $S$. The authors estimate the size of $m_p$.

1. Introduction. Let $p \equiv 1 \pmod{4}$ be a prime. A residue difference set modulo $p$ is a set of integers $\{a_1, \ldots, a_k\}$, with $1 \leq a_i \leq p - 1$, such that
   
   (i) \( (\frac{a_i}{p}) = +1, \ 1 \leq i \leq k, \)
   
   (ii) \( (\frac{a_i - a_j}{p}) = +1, \ 1 \leq i, j \leq k, i \neq j, \)

where $(\frac{n}{p})$ is the Legendre symbol modulo $p$. The maximal cardinality of a residue difference set modulo $p$ is denoted by $m_p$. The problem of estimating $m_p$ was posed at the West Coast Number Theory Conference in La Jolla, California in December 1976. We obtain the following estimates.

Theorem. (i) $m_p > \frac{1}{2} \log p$ for all $p$,
   
   (ii) $m_p < p^{1/2} \log p$ for all $p$,

   (iii) $m_p < (1 + \varepsilon) p^{1/2} \log p / 4 \log 2$ for all $p > C$, where $C \equiv C(\varepsilon)$ is a constant depending only on $\varepsilon$.

Any residue difference set can be transformed into a set containing 1 (by multiplication by any $a_i^{-1} \pmod{p}$), so we need only consider residue difference sets of the form

$$S = \{a_1, a_2, \ldots, a_k\},$$

where $1 = a_1 < a_2 < \cdots < a_k$. Let $N_p(k)$ be the number of such sets. The value of $N_p(2)$ is exactly $(p - 5)/4$; we shall, in proving the theorem, obtain a lower bound for $N_p(k)$ for $k \geq 3$.

The proof of the theorem requires the following lemma, which we state here and prove in §3.

Lemma. For any integer $k \geq 1$, let $a_0, a_1, \ldots, a_{k-1}$ be $k$ integers such that

\begin{footnotesize}
Received by the editors March 14, 1977.


Research supported under National Research Council of Canada Grant No. A-7233.
\end{footnotesize}
\[ a_0 = 0, \ a_1 = 1, \ 1 < a_i < p \ (i = 2, 3, \ldots, k - 1), \ a_i \neq a_j \text{ for } i \neq j. \] Set

\[ S(a_0, \ldots, a_{k-1}) = \sum_{x \neq a_0, \ldots, a_{k-1}}^{p-1} \left\{ \prod_{j=0}^{k-1} \left( 1 + \left( \frac{x - a_j}{p} \right) \right) \right\}. \]

Then \[ |S(a_0, \ldots, a_{k-1}) - p| \leq p^{1/2}((k - 2)2^{k-1} + 1) + k2^{k-1}, \text{ and if } p \geq k^2 \text{ the expression on the right-hand side of this inequality is at most } p^{1/2}k2^{k-1}. \]

Use will also be made of the following simple and easily-proved inequality: if \( b_1, \ldots, b_n \) are \( n \ (\geq 1) \) numbers such that \( p \geq b_1 > b_2 > \cdots > b_n > 0 \) then

\[ (p - b_1) \cdots (p - b_n) > p^n - p^{n-1}(b_1 + \cdots + b_n). \]

2. Proof of the theorem. As \( m_5 = 1, \ m_{13} = m_{17} = 2, \ m_{29} = m_{37} = 3, \ m_{41} = m_{53} = 4 \), part (i) of the theorem is easily verified for \( p < 53 \). Thus we can assume \( p > 61 \), so that \( \frac{1}{2} \log p > 2 \). In order to complete the proof we must show that \( N_p(k) > 0 \) for \( 2 < k < \frac{1}{2} \log p \). To do this, we use the following expression for \( N_p(k) \):

\[
N_p(k) = \frac{1}{2^{(k-1)(k+2)/2}} \sum_{a_2, \ldots, a_k} \left\{ \sum_{1 < a_2 < \cdots < a_k < p} \left( 1 + \left( \frac{a_2}{p} \right) \right) \cdots \left( 1 + \left( \frac{a_k}{p} \right) \right) \right\}
\]

\[
\cdot \left\{ \prod_{2 < j < i < k} \left( 1 + \left( \frac{a_i - a_j}{p} \right) \right) \right\}
\]

\[
= \frac{1}{2^{(k-1)(k+2)/2}(k-1)!} \sum_{1 < a_2 < p} \sum_{a_i \neq a_j, i \neq j} (p) \cdots \sum_{1 < a_k < p} \left( 1 + \left( \frac{a_2}{p} \right) \right) \cdots \left( 1 + \left( \frac{a_k}{p} \right) \right)
\]

\[
\cdot \left\{ \frac{1}{2^{(k-1)(k+2)/2}(k-1)!} \sum_{1 < a_2 < p} \left( 1 + \left( \frac{a_2}{p} \right) \right) \left( 1 + \left( \frac{a_2 - 1}{p} \right) \right) \right\}
\]

\[
\cdot \left\{ \prod_{2 < j < i < k} \left( 1 + \left( \frac{a_i - a_j}{p} \right) \right) \right\}
\]

\[
= \frac{1}{2^{(k-1)(k-2)/2}(k-1)!} \sum_{1 < a_2 < p} \left( 1 + \left( \frac{a_2}{p} \right) \right) \left( 1 + \left( \frac{a_2 - 1}{p} \right) \right)
\]

\[
\cdots \sum_{1 < a_k < p} \left( 1 + \left( \frac{a_k}{p} \right) \right) \left( 1 + \left( \frac{a_k - 1}{p} \right) \right)
\]

\[
\cdot \left\{ \prod_{j=2}^{k-2} \left( 1 + \left( \frac{a_k - a_j}{p} \right) \right) \right\} S(a_0, \ldots, a_{k-1}).
\]
Since \( p > (\frac{1}{2} \log p)^2 \) (for all \( p \)) and as all the summands in the above expression for \( N_p(k) \) are nonnegative, we can apply the lemma successively to obtain

\[
N_p(k) \geq \frac{1}{2^{(k-1)(k+2)/2}(k-1)!} \left( p - 2 \cdot 2^{\frac{k}{2}} \right) \cdots \left( p - k \cdot 2^{k-1}p^{\frac{k}{2}} \right).
\]

Since for all integers \( k \geq 2 \) we have \( \log(k - 1) + k \log 2 < k \), and as \( k \leq \frac{1}{2} \log p \), we obtain

\[
2^{1/2} > (k - 1)2^k > k2^{k-1}.
\]

Applying (1.1) we obtain

\[
N_p(k) \geq \frac{1}{2^{(k-1)(k+2)/2}(k-1)!} \left\{ p^{k-1} - p^{k-3/2} \left( 2 \cdot 2 + \cdots + k \cdot 2^{k-1} \right) \right\}
\]

\[
= \frac{1}{2^{(k-1)(k+2)/2}(k-1)!} \left\{ p^{k-1} - (k - 1)2^k p^{k-3/2} \right\},
\]

and \( N_p(k) > 0 \) follows from (2.1). Thus \( m_p > \frac{1}{2} \log p \) for all primes \( p \).

We now turn to the proofs of parts (ii) and (iii) of the theorem. The set of possible values of \( a_2 \) so that \( \{1, a_2\} \) is a residue difference set modulo \( p \) is

\[
A_2 = \left\{ b \left| \left( \frac{b}{p} \right) = \left( \frac{b - 1}{p} \right) = +1 \right\} \right. .
\]

Fixing a value of \( a_2 \in A_2 \), the set of possible values of \( a_3 \) so that \( \{1, a_2, a_3\} \) is a residue difference set modulo \( p \) is

\[
A_3 = \left\{ b \left| b \in A_2, \left( \frac{b - a_2}{p} \right) = +1 \right\} \right. .
\]

Continuing in this way, one obtains for any residue difference set \( S = \{1, a_2, \ldots, a_{k-1}\} \), a set \( A_k \) of possible values of \( a_k \) so that \( \{1, a_2, \ldots, a_k\} \) is a residue difference set. If \( \alpha_k \) denotes the number of elements of \( A_k \), then the residue difference set of maximal length that contains \( S \) as a subset certainly has at most \( k - 1 + \alpha_k \) elements, where

\[
\alpha_k = \frac{1}{2^k} \sum_{a_{k-1} < a_k < p} \left\{ 1 + \left( \frac{a_k}{p} \right) \right\} \left\{ 1 + \left( \frac{a_k - 1}{p} \right) \right\} \left\{ 1 + \left( \frac{a_k - a_2}{p} \right) \right\} \cdots \left\{ 1 + \left( \frac{a_k - a_{k-1}}{p} \right) \right\}
\]

\[
< \frac{1}{2^k} \sum_{a \neq a_0, a_1, \ldots, a_{k-1}} a = 0 \rightarrow k - 1 \prod_{i = 0}^{p-1} \left( 1 + \left( \frac{a - a_i}{p} \right) \right) = \frac{1}{2^k} S(a_0, \ldots, a_{k-1}).
\]

Thus, if \( m_p > k - 1 \), there exists a set \( S = \{1, a_2, \ldots, a_{k-1}\} \) which is a subset of a residue difference set of \( m_p \) elements, and

\[
m_p < k - 1 + \frac{1}{2^k} S(a_0, \ldots, a_{k-1}).
\]
Hence from the lemma we have
\[ m_p \leq k - 1 + \frac{1}{2^k} \{ p + p^{1/2}((k - 2)2^{k-1} + 1) + k2^{k-1} \} \]
\[ \leq \frac{3k}{2} - 1 + \frac{p}{2^k} + \frac{(k - 1)}{2} p^{1/2}. \]
If we now choose \( k = 1 + \lceil \log p / 2 \log 2 \rceil \), we see that \( m_p \geq \lceil \log p / 2 \log 2 \rceil \) implies
\[ m_p \leq \frac{3}{4 \log 2} \log p + \frac{1}{2} + p^{1/2} + \frac{p^{1/2} \log p}{4 \log 2}. \]
Now for \( p \geq 37 \) we have
\[ m_p \leq \left( \frac{3}{4 \sqrt{37} \log 2} + \frac{1}{2 \sqrt{37} \log 37} + \frac{1}{\log 37} + \frac{1}{4 \log 2} \right) p^{1/2} \log p \]
\[ < (0.18 + 0.03 + 0.28 + 0.37) p^{1/2} \log p \]
\[ = 0.86 p^{1/2} \log p \]
\[ < p^{1/2} \log p. \]
As the inequality \( m_p < p^{1/2} \log p \) is easy to check for \( p = 5, 13, 17 \) and 29, this completes the proof of (ii).
Part (iii) follows by choosing \( p \geq C(\epsilon) \) so that
\[ \frac{3}{4 \log 2} \log p + \frac{1}{2} + p^{1/2} \leq \epsilon \frac{p^{1/2} \log p}{4 \log 2}. \]

3. Proof of Lemma. Let \( f(x) = (x - c_1) \cdots (x - c_t) \), where the \( c_i \) are \( t \) \((\geq 1)\) integers which are congruent modulo an odd prime \( p \). Then the following estimate is a consequence of a deep result of A. Weil (see for example [1], [2]):

\[ \left| \sum_{x=0}^{p-1} \left( \frac{f(x)}{p} \right) \right| \leq (t - 1) p^{1/2}. \]

The term corresponding to the product of the 1's in \( S(a_0, \ldots, a_{k-1}) \) is
\[ \sum_{x=0}^{p-1} 1 = p - k. \]
A typical term amongst the remaining \( 2^k - 1 \) terms is
\[ \sum_{x=0}^{p-1} \left( \frac{(x - a_{i_1}) \cdots (x - a_{i_r})}{p} \right) \]
where \( k \geq r \geq 1, 0 \leq i_1 < \cdots < i_r \leq k - 1 \). By (3.1) this sum is bounded in absolute value by \( (r - 1) p^{1/2} + k - r \). We thus have
\[ |S(a_0, \ldots, a_{k-1}) - (p - k)| \leq \sum_{r=1}^{k} \left\{ (r - 1) p^{1/2} + (k - r) \right\} \binom{k}{r} \]

\[ = (p^{1/2} - 1) \sum_{r=1}^{k} r \binom{k}{r} - (p^{1/2} - k) \sum_{r=1}^{k} \binom{k}{r} \]

\[ = (p^{1/2} - 1) k 2^{k-1} - (p^{1/2} - k)(2^k - 1) \]

\[ = p^{1/2} \{(k - 2)2^{k-1} + 1\} + \{k2^{k-1} - k\}, \]

so that

\[ |S(a_0, \ldots, a_{k-1}) - p| \leq p^{1/2} \{(k - 2)2^{k-1} + 1\} + k2^{k-1}. \]

If \( p > k^2 \) then the right-hand side of the above is

\[ < p^{1/2} \{(k - 2)2^{k-1} + 1 + 2^{k-1}\} \]

\[ < p^{1/2}k2^{k-1}. \]

4. Remarks. We note that the above arguments can be slightly refined to obtain marginal improvements in the constants appearing in the theorem. However, it appears to be a difficult problem to obtain the true order of magnitude of \( m_p \). We have computed \( N_p(k) \) and \( m_p \) for all primes \( p \leq 617 \) and observed that for \( p \) in the range \( 401 < p < 617 \), \( m_p/\log p \) varies between 1.27 and 1.72. One might expect, therefore, that \( m_p \sim c \log p \) for some constant \( c \) with \( 1 < c < 2 \). However, our arguments, unless significantly modified, would not seem to yield a result of the type \( m_p \geq \log p \).

The residue difference sets modulo \( p \) form a tree with the nodes of the second level corresponding to the elements of \( \mathbb{A}_2 \), the nodes of the third level corresponding to the elements of all sets \( \mathbb{A}_3 \), etc. The computation of \( N_p(k) \) was done by a depth-first search through this tree on the Xerox Data Systems Sigma 9 computer at Carleton University. As an indication of the number of nodes involved we note that for \( p = 617 \) there were 1,374,659 nodes.

REFERENCES


DEPARTMENT OF MATHEMATICS, CARLETON UNIVERSITY, OTTAWA (K1S 5B6), ONTARIO, CANADA (Current address of K. S. Williams)

Current address (D. A. Buell): Department of Computer Science, Bowling Green State University, Bowling Green, Ohio 43403