ACTA ARITHMETICA
127.1 (2007)

Equality in Pollard's theorem on set addition of congruence classes

by

E. Nazarewicz, M. O'Brien, M. O'Neill
and C. Staples (Claremont, CA)

1. Introduction. Let $p \in \mathbb{Z}$ be a prime, and let A and B be non-empty subsets of $\mathbb{Z}_p$. Let $|S|$ denote the cardinality of $S \subset \mathbb{Z}_p$ and let $\overline{S} \subset \mathbb{Z}_p$ denote the complement of S. Cauchy [2], and independently Davenport [3], showed that

$$|A+B| \geq \min\{p, |A|+|B|-1\} \tag{1}$$

where $A+B = \{a+b : a \in A, b \in B\}$. Davenport reports on his later discovery of Cauchy's priority in the historical note [4].

The problem of characterizing the pairs of sets (A,B) for which equality holds in the Cauchy–Davenport inequality was solved by Vosper in [12]. He obtained the following

Theorem 1. *Let p be a prime number, and let A and B be non-empty subsets of $\mathbb{Z}_p$. Then*

$$|A+B| = \min\{|A|+|B|-1, p\}$$

if and only if at least one of the following conditions holds:

(i) $\min\{|A|,|B|\} = 1$,
(ii) $|A|+|B| \geq p+1$,
(iii) $B = c - \overline{A}$ *for some* $c \in \mathbb{Z}_p$,
(iv) *A and B are arithmetic progressions with the same common difference.*

The following generalization of (1) was obtained in [9] by Pollard.

Theorem 2. *Let p be a prime number, and let A and B be non-empty subsets of $\mathbb{Z}_p$. For $t = 1, 2, \ldots, \min\{|A|,|B|\}$, let $N_t = N_t(A,B)$ denote the number of congruence classes in $\mathbb{Z}_p$ that have at least t representations in the form $a+b$, where $a \in A$ and $b \in B$. Then*

$$N_1 + \cdots + N_t \geq t \min\{p, |A|+|B|-t\}. \tag{2}$$

2000 *Mathematics Subject Classification*: 11A07, 11B13.

Our main theorem characterizes the pairs (A, B) for which equality holds in (2) for a given $t \geq 2$. We will call such pairs *t-critical*. Since (2) reduces to (1) when $t = 1$, our theorem complements Vosper's theorem. To be precise, we have

THEOREM 3. *Let p be a prime number. For $2 \leq t \leq \min\{|A|, |B|\}$, the pair (A, B) of non-empty subsets of $\mathbb{Z}_p$ is t-critical if and only if at least one of the following conditions holds*:

(i) $\min\{|A|, |B|\} = t$,
(ii) $|A| + |B| \geq p + t$,
(iii) $|A| = |B| = t + 1$ *and* $B = g - A$ *for some* $g \in \mathbb{Z}_p$,
(iv) A *and* B *are arithmetic progressions with the same common difference.*

An important tool in the proof of the theorem is the *e-transform* of a pair of sets. For the reader's convenience, we recall the definition and a few properties.

Given a pair (A, B) of subsets of $\mathbb{Z}_p$ and an element $e \in \mathbb{Z}_p$, let $A(e) = A \cup (B + e)$ and $B(e) = B \cap (A - e)$. Then

(1) $A(e) + B(e) \subseteq A + B$,
(2) $A(e) \setminus A = e + (B \setminus B(e))$,
(3) $|A(e)| + |B(e)| = |A| + |B|$.

The book [8] is an excellent reference for the background briefly described above. We restate and prove our main theorem in Section 3. In Section 2 we prove several lemmas needed for the proof of the theorem. The final section contains a few ideas for further work.

The authors are indebted to the anonymous referee for suggesting many improvements. In particular, arguments used in Lemmas 1, 5 and 8 were suggested by the referee and replace our less digestible original presentation.

2. Preliminaries. For $x \in \mathbb{Z}_p$, let $r_{A,B}(x)$ denote the number of solutions of $x = a + b$ with $a \in A$ and $b \in B$. Let $S(A, B, t) = N_1 + \cdots + N_t$. Then we have

$$S(A, B, t) = \sum_{x \in \mathbb{Z}_p} \min\{t, r_{A,B}(x)\}.$$

It is clear that $r_{A,B}(x) \leq \min\{|A|, |B|\}$ for each $x \in \mathbb{Z}_p$. Therefore, when $t = \min\{|A|, |B|\}$, we have

$$S(A, B, t) = \sum_{x \in \mathbb{Z}_p} \min\{t, r_{A,B}(x)\} = \sum_{x \in \mathbb{Z}_p} r_{A,B}(x) = |A|\,|B|.$$

For $|A|+|B| \geq p+t$, we have $r_{A,B}(g) \geq t$ for all $g \in \mathbb{Z}_p$. So in this case,

$$S(A,B,t) = \sum_{x\in\mathbb{Z}_p} \min\{t, r_{A,B}(x)\} = \sum_{x\in\mathbb{Z}_p} t = tp.$$

Accordingly, for the remainder of this section, we shall assume $|A|+|B| < p+t$ and $1 < t < \min\{|A|,|B|\}$.

LEMMA 1. *Let (A,B) be a t-critical pair of subsets of $\mathbb{Z}_p$, and suppose that $1 < t < \min\{|A|,|B|\}$ and $|A|+|B| < p+t$. If A is an arithmetic progression then B is an arithmetic progression with the same common difference.*

Proof. Without loss of generality, we may assume that $|A| \geq |B|$, that $0 \in B$ and that $A = \{0,1,\ldots,k-1\} + p\mathbb{Z}$, where $k = |A|$. For brevity, write $\ell = |B|$. We shall complete the proof of the lemma by showing that $B = \{b, b+1, \ldots, b+\ell-1\}$ for some $b \in B$. Choose integers

$$0 = r_0 < r_1 < \cdots < r_{\ell-1} < p$$

so that, with $b_j = r_j + p\mathbb{Z}$, we may write

$$B = \{b_0, b_1, \ldots, b_{\ell-1}\}.$$

For $0 \leq i \leq t$ let $r_{\ell+i} = p + r_i$. The set $A+B$ is the union of the intervals $[r_j, r_j+k-1]+p\mathbb{Z}$, and for any $x \in \mathbb{Z}_p$ we have

$$\min\{t, r_{A,B}(x)\} = |\{j \in [0,\ell-1] : r_j \leq x \leq r_j + \min\{k-1, r_{j+t}-r_j-1\}\}|.$$

Writing $s_{j,t} = \min\{k, r_{j+t}-r_j\}$, we then obtain

$$\begin{aligned} S(A,B,t) &= \sum_{x\in\mathbb{Z}_p} \min\{t, r_{A,B}(x)\} = \sum_{x\in\mathbb{Z}_p} |\{j \in [0,\ell-1] : r_j \leq x \leq r_j + s_{j,t} - 1\}| \\ &= \sum_{j=0}^{\ell-1} \sum_{x=r_j}^{r_j+s_{j,t}-1} 1 = \sum_{j=0}^{\ell-1} \min\{k, r_{j+t}-r_j\}. \end{aligned}$$

From the above equality, we have

$$\begin{aligned} t(k+\ell-t) &= \sum_{j=0}^{\ell-1} \min\{k, r_{j+t}-r_j\} \\ &= \sum_{j=0}^{\ell-1} (r_{j+t} - r_j - \max\{0, r_{j+t}-r_j-k\}) \\ &= tp - \sum_{j=0}^{\ell-1} \max\{0, r_{j+t}-r_j-k\}, \end{aligned} \tag{3}$$

from which it follows that

$$t(p+t-k-\ell)=\sum_{j=0}^{\ell-1}\max\{0,r_{j+t}-r_j-k\}. \tag{4}$$

Set

$$J_0=\{j\in[0,\ell-1]: r_{j+t}-r_j>k\},\qquad J_1=\{j\in[0,\ell-1]: r_{j+t}-r_j\le k\}.$$

For each $j\in[0,\ell-1]$ we have $\{r_{j+t+1},\dots,r_{j+\ell-1}\}\subset[r_{j+t}+1,r_j+p-1]$ and therefore $r_{j+t}-r_j\le p+t-\ell$ for all such j. This fact and (4) now give

$$t(p+t-k-\ell)=\sum_{j\in J_0}(r_{j+t}-r_j-k)\le(p+t-\ell-k)|J_0|,$$

implying $|J_0|\ge t$. Next, from (3) we get

$$\begin{aligned}t(k+\ell-t)&=k|J_0|+\sum_{j\in J_1}(r_{j+t}-r_j)\ge k|J_0|+t(\ell-|J_0|)=t\ell+(k-t)|J_0|\\&\ge t(\ell+k-t).\end{aligned}$$

It follows that $|J_0|=t$, $|J_1|=\ell-t$, and moreover, $r_{j+t}-r_j=t$ for each $j\in J_1$. Observe that the latter equality implies that for each $j\in J_1$ we have $r_j+i+p\mathbb{Z}\in B$ for each $i\in\{0,\dots,t\}$.

To complete the proof, write $B_1=\{r_j+p\mathbb{Z}: j\in J_1\}$ and represent B as a disjoint union $B=\bigcup_{i=1}^s P_i$, where each P_i is an arithmetic progression in $\mathbb{Z}_p$ with difference 1, and any two of these progressions are separated by at least one element from $\bar{B}$. We then have $|P_i\cap B_1|\le|P_i|-\min\{|P_i|,t\}$ for each $i\in[1,s]$ and hence

$$\ell-t=|B_1|\le\sum_{i=1}^s(|P_i|-\min\{|P_i|,t\})=\ell-\sum_{i=1}^s\min\{|P_i|,t\},$$

implying

$$\sum_{i=1}^s\min\{|P_i|,t\}\le t.$$

Finally, from $\sum_{i=1}^s|P_i|=\ell>t$ we conclude that $s=1$, as desired. ∎

Lemma 1 provides an important tool in the induction argument that we use to prove the main theorem. The next four lemmas dispose of various endpoint cases.

Lemma 2. *If $|A|=t+1<p$ then $(A,g-A)$ is a t-critical pair for any $g\in\mathbb{Z}_p$.*

Proof. We have $N_1(A, g-A) + \cdots + N_{|A|}(A, g-A) = |A|^2$ and $|A| < p$. By Pollard's inequality we have

$$\begin{aligned} |A|^2 - N_{|A|}(A, g-A) &= N_1(A, g-A) + \cdots + N_{|A|-1}(A, g-A) \\ &\geq t \min\{2|A| - t, p\} = |A|^2 - 1. \end{aligned}$$

Therefore $N_{|A|}(A, g-A) \leq 1$, and since g has exactly $|A|$ distinct representations in $A+(g-A)$, equality holds. It follows that $(A, g-A)$ is t-critical. ∎

LEMMA 3. *Suppose that (A, B) is a t-critical pair of subsets of $\mathbb{Z}_p$. If $|A| = |B| = t+1$ and $|A| + |B| < p + t$, then $B = g - A$ for some $g \in \mathbb{Z}_p$.*

Proof. Since $|A| + |B| < p + t$ we have

$$\begin{aligned} N_1 + \cdots + N_{t+1} &= |A|\,|B| = t^2 + 2t + 1, \\ N_1 + \cdots + N_t &= t(|A| + |B| - t) = t^2 + 2t. \end{aligned}$$

It follows that $N_{|A|} = 1$, which implies that B is of the form $g - A$ for some $g \in \mathbb{Z}_p$. ∎

LEMMA 4. *Suppose that (A, B) is a t-critical pair of subsets of $\mathbb{Z}_p$. If $|A| > |B| = t+1$ and $|A| + |B| < p + t$ then A and B are arithmetic progressions with the same common difference.*

Proof. We have

$$\begin{aligned} N_1 + \cdots + N_t &= t(|A| + |B| - t) = t(|A| + 1), \\ N_1 + \cdots + N_{t+1} &= (t+1)|A| \end{aligned}$$

so that $N_{t+1} = |A| - t$.

Let $C \subset A + B$ be the set of elements $c \in \mathbb{Z}_p$ which have $t+1$ distinct representations as $c = a + b$ with $a \in A$ and $b \in B$. Then $C - B \subseteq A$, so $|C - B| \leq |A|$. Since $|A| + |B| < p + t$, we have $|A| < p - 1$, so by (1),

$$|C - B| \geq \min\{p, |C| + |B| - 1\} = |C| + |B| - 1 = |A| - t + |B| - 1 = |A|.$$

It follows that $(C, -B)$ is a critical pair for (1). Since $|C - B| = |A| < p - 1$ and $|A| > |B| > 1$, Vosper's theorem implies that C and $-B$ are arithmetic progressions with the same common difference. By Lemma 1 then, A is an arithmetic progression with the same common difference. ∎

Observe that Lemmas 3 and 4 have the following corollary.

COROLLARY 1. *If (A, B) is t-critical and $\min\{|A|, |B|\} = t+1$ while $|A| + |B| < p + t$, then either $B = g - A$ for some $g \in \mathbb{Z}_p$ or A and B are arithmetic progressions with the same common difference.*

LEMMA 5. *Suppose that (A, B) is a t-critical pair of subsets of $\mathbb{Z}_p$. If $1 < t < \min\{|A|, |B|\}$ and $|A| + |B| = p + t - 1$, then A and B are arithmetic progressions with the same common difference.*

Proof. Note that the bound $\max\{|A|,|B|\} \le p-2$ follows from our standing assumption that $1 < t < \min\{|A|,|B|\}$. From the remark at the beginning of this section, it follows that $r_{A,B}(g) \ge t-1$ for all $g \in \mathbb{Z}_p$. Therefore we have

$$N_1 = \cdots = N_{t-1} = p.$$

Since (A,B) is t-critical, we also have

$$N_1 + \cdots + N_t = t(|A| + |B| - t) = t(p-1).$$

It follows that $N_{t-1} - N_t = t$.

Let C be the set of all those elements of $A+B$ with exactly $t-1$ representations as a sum of an element of A and an element of B, so that $|C| = t$. Then for any $c \in C$ the number of representations of c as a sum of an element of $\overline{A}$ and an element of B is $|B| - (t-1)$, and consequently the number of representations as a sum of an element of $\overline{A}$ and an element of $\overline{B}$ is $|\overline{A}| - (|\overline{B}| - (t-1)) = p - |A| - |B| + (t-1) = 0$. It follows that C is disjoint from $\overline{A} + \overline{B}$ and therefore by the Cauchy–Davenport theorem that

$$p - t = |\overline{C}| \ge |\overline{A} + \overline{B}| \ge |\overline{A}| + |\overline{B}| - 1 = p - t.$$

We notice that the Cauchy–Davenport theorem applies here since $|\overline{A}| = p - |A| = |B| - t + 1 \ge 2$, and similarly $|\overline{B}| \ge 2$. Vosper's theorem now shows that $\overline{A}$ and $\overline{B}$ are arithmetic progressions with the same common difference and therefore so are A and B. ∎

LEMMA 6. *Suppose that* $3 \le |A| \le p-3$ *and that* (A,B) *is a* 2*-critical pair of subsets of* $\mathbb{Z}_p$*. If* $B = g - \overline{A}$ *for some* $g \in \mathbb{Z}_p$ *then* A *and* B *are arithmetic progressions with the same common difference.*

Proof. From $B = g - \overline{A}$ we easily get $N_1 = p-1$, and since (A,B) is 2-critical we have $N_1 + N_2 = 2\min\{|A| + |B| - 2, p\} = 2p-4$. Thus $N_2 = p-3$, and $N_1 - N_2 = 2$. Let c be one of the two elements of $\mathbb{Z}_p$ which have a unique representation as $c = a + b$ with $a \in A$ and $b \in B$. Then $(c - B) \cap A = \{a\}$ so that $(c - (g - \overline{A})) \cap A = \{a\}$. It follows that $(c - g + A) \cap A = A \setminus \{a\}$, which implies that A is an arithmetic progression with common difference $c - g$. By Lemma 1 then B is an arithmetic progression with the same common difference. ∎

The next lemma provides one more necessary analytical tool. Recall that according to our notation, $r_{A(e),B(e)}(c)$ denotes the number of solutions of $c = a + b$ with $a \in A(e)$ and $b \in B(e)$, and that $N_t(A(e),B(e))$ denotes the number of those $c \in \mathbb{Z}_p$ with $r_{A(e),B(e)}(c) \ge t$.

LEMMA 7. *For any* $e \in \mathbb{Z}_p$ *the e-transform* $(A(e),B(e))$ *of the pair* (A,B) *satisfies*

$$N_t(A(e),B(e)) \le N_t$$

for all t.

Proof. It is sufficient to prove that for all c in the sum set $A(e)+B(e)$,

$$r_{A(e),B(e)}(c) \le r_{A,B}(c).$$

But it is easily checked (see for example [8, pp. 46–47]) that

$$r_{A,B}(c) = r_{A(e),B(e)}(c) + r_{A\setminus(B+e),B\setminus(A-e)}(c). \blacksquare$$

A final technical lemma is required at the end of the proof of Theorem 3.

LEMMA 8. *Suppose that* $2 \le s \le |B| \le p-2$ *for some* $B \subset \mathbb{Z}_p$ *and that*

$$|(x_1+B)\cap\cdots\cap(x_s+B)| \ge |B|-(s-1)$$

for some set $X=\{x_1,\dots,x_s\}\subset\mathbb{Z}_p$. *If* $|B|=s$ *then* $X=g-B$ *for some* $g\in\mathbb{Z}_p$ *and if* $s<|B|$ *then* B *is an arithmetic progression.*

Proof. We have

$$p-(|B|-(s-1)) \ge |(x_1+\overline{B})\cup\cdots\cup(x_s+\overline{B})| = |X+\overline{B}| \ge s+(p-|B|)-1$$

by the Cauchy–Davenport theorem. The conclusions of the lemma then follow from Vosper's theorem. ∎

3. A Vosper type theorem for Pollard's inequality. We can now prove Theorem 3. We restate it here for the reader's convenience.

THEOREM 3. *Let* p *be a prime number. For* $2 \le t \le \min\{|A|,|B|\}$, *the pair* (A,B) *of non-empty subsets of* $\mathbb{Z}_p$ *is* t*-critical if and only if at least one of the following conditions holds*:

(i) $\min\{|A|,|B|\}=t$,
(ii) $|A|+|B|\ge p+t$,
(iii) $|A|=|B|=t+1$ *and* $B=g-A$ *for some* $g\in\mathbb{Z}_p$,
(iv) A *and* B *are arithmetic progressions with the same common difference.*

Proof. The sufficiency of the first two conditions has been discussed at the beginning of Section 2. The sufficiency of condition (iii) follows from Lemma 2 and the sufficiency of condition (iv) is an easy exercise. To prove the necessity, suppose that $t \ge 2$ is the smallest positive integer for which the theorem fails. Then by Lemmas 1 and 2, and Corollary 1, there exists a t-critical pair (A,B) such that $|A| \ge |B| > t+1$ and such that neither A nor B is an arithmetic progression. Since A and B are not arithmetic progressions, we have $\max\{|A|,|B|\} \le p-2$, and then by Lemma 5, we have $|A|+|B| \le p+t-2$.

We may choose the pair (A,B) so that

(i) $|A+B|$ is minimal,
(ii) $|A|+|B|$ is maximal subject to (i),
(iii) $|B|$ is minimal subject to (i) and (ii).

In broad outline, the proof will now proceed in the following four steps:

(1) Rule out the existence of $e \in \mathbb{Z}_p$ such that $t < |B(e)| < |B|$.
(2) Rule out the existence of $e \in \mathbb{Z}_p$ such that $0 < |B(e)| < t$.
(3) Show the existence of many $e \in \mathbb{Z}_p$ such that $|B(e)| = t$.
(4) Obtain a contradiction with the form of the pair (A, B) assumed at the beginning of the proof.

With an eye toward step (1) then, we continue the proof as follows.

If $t > 2$ then, since $|A| + |B| \le p + t - 2$ and $|B| > t + 1$, by minimality of t we have

$$N_1 + \cdots + N_{t-1} > (t-1)(|A| + |B| - (t-1)).$$

Indeed, Vosper's theorem shows that this estimate also holds for $t = 2$, as A and B are not arithmetic progressions and Lemma 6 shows that B is not of the form $g - \overline{A}$ with $g \in \mathbb{Z}_p$.

Suppose that there is some $e \in \mathbb{Z}_p$ such that $t < |B(e)| < |B|$ and, for brevity, set $N_i' = N_i(A(e), B(e))$. By Pollard's inequality we find

$$N_1' + \cdots + N_t' \ge t(|A(e)| + |B(e)| - t) = t(|A| + |B| - t) = N_1 + \cdots + N_t.$$

By Lemma 7, it follows that $N_i' = N_i$ for $1 \le i \le t$, and therefore

$$N_1' + \cdots + N_{t-1}' > (t-1)(|A| + |B| - (t-1)) = (t-1)(|A(e)| + |B(e)| - (t-1)).$$

By Lemma 1, since $(A(e), B(e))$ is a t-critical pair, if one of $(A(e), B(e))$ is an arithmetic progression, then so is the other and the two progressions have the same common difference. Therefore, the previous inequality implies that neither of $(A(e), B(e))$ is an arithmetic progression. By the properties of the e-transform recorded in the introduction, it follows that $(A(e), B(e))$ is also a t-critical pair that violates the conclusion of the theorem. This is in contradiction to the minimality of $|B|$. Therefore there can be no $e \in \mathbb{Z}_p$ such that $t < |B(e)| < |B|$.

Suppose now that there is an $e \in \mathbb{Z}_p$ such that $0 < |B(e)| < t$. Suppose further that $|B(e)|$ is minimized over such choices of e. Observing that

$$N_i(A(e), B(e)) = N_i((A - e) \cup B, B(e))$$

and $N_i(A - e, B) = N_i(A, B)$ for each i, we consider the pair

$$(U, I) = ((A - e) \cup B, (A - e) \cap B)$$

with $|I| < t$. Write $A_e = (A - e)$ and let $A_e' = (A - e) \setminus B$, $B' = B \setminus (A - e)$ and $t' = t - |I|$. For each $x \in \mathbb{Z}_p$ we have

$$r_{A_e,B}(x) = r_{U,I}(x) + r_{A_e',B'}(x),$$

so that

$$\begin{aligned} \min\{t, r_{A_e,B}(x)\} &\ge \min\{|I|, r_{U,I}(x)\} + \min\{t', r_{A_e',B'}(x)\} \\ &= r_{U,I}(x) + \min\{t', r_{A_e',B'}(x)\}. \end{aligned}$$

We also have

$$1 \le t' = t - |I| < |B| - |I| = |B'|$$

and

$$|A'_e| + |B'| - t' = (|A_e| - |I|) + (|B| - |I|) - (t - |I|) = |U| - t < |U| \le p.$$

Therefore

$$\begin{aligned} t(|A_e| + |B| - t) &= \sum_{x \in \mathbb{Z}_p} \min\{t, r_{A_e,B}(x)\} \ge \sum_{x \in \mathbb{Z}_p} r_{U,I}(x) + \sum_{x \in \mathbb{Z}_p} \min\{t', r_{A'_e,B'}(x)\} \\ &\ge |U|\,|I| + t'(|A'_e| + |B'| - t') = |U|\,|I| + (t - |I|)(|U| - t) \\ &= t(|A_e| + |B| - t). \end{aligned}$$

It follows that

$$\sum_{x \in \mathbb{Z}_p} \min\{t', r_{A'_e,B'}(x)\} = t'(|A'| + |B'| - t')$$

and therefore that (A'_e, B') is a t'-critical pair. Since $|A'_e| \ge |B'| > t' + 1$ and $|A'_e| + |B'| \le p + t' - 2 - |I|$, we conclude, by the minimality of t, that A'_e and B' are arithmetic progressions with the same common difference.

We have found that if $0 < |B(e)| < t$ for some $e \in \mathbb{Z}_p$ then the sets $B \setminus (A - e)$ and $(A - e) \setminus B$ must be arithmetic progressions with the same common difference. We will now show that, together with our assumptions on the pair (A, B), this implies that A and B are arithmetic progressions with the same common difference. Let d be the common difference of the progressions $B \setminus (A - e)$ and $(A - e) \setminus B$. Dividing by $d \pmod p$ we may assume that $d = 1$. By a translation we may assume that

$$(A - e) \setminus B = \{0, 1, \dots, m\}, \qquad B \setminus (A - e) = \{m + j, \dots, m + j + k\}.$$

Let $X = B(e) \cap (m, m + j)$ and $Y = B(e) \cap (m + j + k, p)$ and assume first that both X and Y are non-empty. Write $X = \{x_1, \dots, x_s\}$ and $Y = \{y_1, \dots, y_r\}$, where $m < x_1 < \cdots < x_s < m + j$ and $m + j + k < y_1 < \cdots < y_r < p$ and congruence classes are identified by their representatives in $[0, p - 1]$. Put

$$d' = \min\{x_2 - x_1, \dots, x_s - x_{s-1}, y_2 - y_1, \dots, y_r - y_{r-1}\},$$

so that $(B + d') \cap (A - e)$ is non-empty, and therefore

$$(*) \qquad |(B + d') \cap (A - e)| \ge |B(e)|$$

by minimality of $|B(e)|$. We have $|(X + d') \cap X| \le |X| - 1$, and equality holds if and only if X is an arithmetic progression with difference d'. A similar observation applies to Y. From $(*)$, it now follows that X and Y are arithmetic progressions with difference d', and moreover, that $y_1 = m + j + k + d'$ and $y_r = p - d'$. In the same way, considering $(B - d') \cap (A - e)$ instead of

$(B + d') \cap (A - e)$, we deduce that $x_1 = m + d'$ and $x_s = m + j - d'$. We now claim that, in fact, $d' = 1$ holds: for otherwise

$$0 < |(B + d' + 1) \cap (A - e)| < |B(e)|,$$

contradicting the minimality of $|B(e)|$. But this implies that A and B are arithmetic progressions with the same common difference.

To remove the assumption that both X and Y must be non-empty, notice that at least one of them must be non-empty. If X is empty, let

$$c = \min\{j, y_2 - y_1, \ldots, y_r - y_{r-1}\}.$$

Then, arguing as above, we see that $0 < |(B-c)\cap(A-e)| < |B(e)|$ unless Y is an arithmetic progression with difference 1 for which $y_1 = m+j+k+1$ and $y_r = p - 1$. A similar argument applies if Y is empty and X is non-empty.

Since A and B are not arithmetic progressions, we have shown that there can be no $e \in \mathbb{Z}_p$ such that $0 < |B(e)| < t$. The only possible values for $|(A - e) \cap B|$ are thus 0, t and $|B|$.

We are in a position to proceed with step (3) of our initial outline of the proof. We will show that for any $b \in B$ there are many values of $e \in \mathbb{Z}_p$ such that $b \in B(e) \subsetneq B$ and therefore such that $|B(e)| = t$.

It is easily seen that the set of those e with the property in question is precisely the set $E = \{e \in A - b : B + e \not\subseteq A\}$. Let $E' = (A - b) \setminus E$. If $|E'| \geq 2$, then taking into account that $B + E' \subseteq A$ and that B is not an arithmetic progression, by Vosper's theorem we get

$$|A| \geq |B + E'| \geq |B| + |E'| = |B| + (|A| - |E|)$$

whence $|E| \geq |B|$.

If E' is empty then we have $|E| = |A| \geq |B|$ and if $|E'| = 1$ then $|E| = |A| - 1 \geq |B| - 1$. Notice that $|E'| = |\{e \in A - b : B + e \subseteq A\}|$ is independent of $b \in B$ and therefore $|E|$ is also independent of $b \in B$.

We now assume that $|E| \geq |B|$. The case where $|E'| = 1$ and $|E| = |A| - 1 = |B| - 1$ will be treated separately at the end of the proof by making a modification to the following argument.

By making translations, we may assume that $0 \in A \cap B$ and that $|A \cap B| = t$. Let $U = A \cup B$ and $I = A \cap B$. Then since $|U|\,|I| = t(|A| + |B| - t)$ we have

$$S(U, I, t) = S(A, B, t)$$

and $r_{A,B}(x) = r_{U,I}(x) + r_{A',B'}(x)$ for each x (with the same notation as earlier). Since (A, B) is a t-critical pair, it follows that $A' + B'$ is contained in the subset of elements of $U + I$ which have t distinct representations as $u + i$ with $u \in U$ and $i \in I$. That is,

$$A' + B' \subseteq \bigcap_{b \in I} (b + U) \subseteq \{x : r_{A,B}(x) \geq t\}. \tag{5}$$

As before, let $E(b) = \{e \in A - b : B + e \not\subseteq A\}$. Then, by assumption, $|E(b)| \geq |B|$ for each $b \in I$ and for any $e \in E(b)$ we have $b \in (A - e) \cap B$. Fix $b^* \in I$ and $e \in E(b^*)$ and keep in mind that $b^* \in (A - e) \cap B$. We have, as for the case of (U, I),

$$S(A(e), B(e), t) = S(A, B, t)$$

and $r_{A,B}(x) = r_{A(e),B(e)}(x) + r_{A\setminus(B+e),B\setminus(A-e)}(x)$ for each x. It follows, in the same way as for the pair (U, I), that

$$A \setminus (B+e) + B \setminus (A-e) \subseteq \bigcap_{b \in (A-e)\cap B} (b + (A \cup (B+e))) \subseteq \{x : r_{A,B}(x) \geq t\}.$$

In fact, since (A, B) is t-critical we have

$$r_{U,I}(x) = r_{A(e),B(e)}(x) = r_{A,B}(x)$$

whenever $r_{A,B}(x) < t$ and we have

$$\{x : r_{A,B}(x) \geq t\} = \{x : r_{U,I}(x) = t\} = \{x : r_{A(e),B(e)}(x) = t\}.$$

Consequently,

$$A' + B' \subseteq \bigcap_{b \in (A-e)\cap B} (b + (A \cup (B + e))).$$

Since this inclusion holds for each $e \in E(b^*)$ we obtain

$$A' + B' \subseteq \bigcap_{e \in E(b^*)} (b^* + (A \cup (B + e))).$$

It follows that

$$(A' + B') \setminus (b^* + A) \subseteq b^* + \bigcap_{e \in E(b^*)} (B + e).$$

If $|B| < |E(b^*)|$, then by Lemma 8, we have $\bigcap_{e \in E(b^*)}(B + e) = \emptyset$ so that $A' + B' \subseteq b^* + A$. If $|B| = |E(b^*)|$ then Lemma 8 shows that $E(b^*) = g' - B$ for some $g' \in \mathbb{Z}_p$ with $\{g'\} = \bigcap_{e \in E(b^*)}(B + e)$. In this case, since $b^* + E(b^*) \subseteq A$, we have $g' = b^* + (g' - b^*) \in A$ and it still follows that $A' + B' \subseteq b^* + A$.

Since b^* was an arbitrary point of I, we have

$$A' + B' \subseteq \bigcap_{b \in I} (b + A)$$

so that, by the Cauchy–Davenport inequality,

$$\Big|\bigcap_{b \in I}(b + A)\Big| \geq |A' + B'| \geq |A| + |B| - 2t - 1 \geq |A| - t + 1 \tag{6}$$

since $|B| > t + 1$. By Lemma 8, this can only occur if A is an arithmetic progression. This is a contradiction, so the theorem is proved except for the case when $|E'| = 1$ and $|E| = |A| - 1 = |B| - 1$.

When $|E'| = 1$ and $|A| = |B|$ we have $A = g + B$ for some $g \in \mathbb{Z}_p$. We shall assume further, for the moment, that $|B| < (p+1)/2$. We require two observations. The first is that in this case we must have $|B| > 2t$.

Indeed, we have, using our assumptions on B and the Cauchy–Davenport inequality,

$$\begin{aligned}(7)\quad |B|^2 = |A|\,|B| &= \sum_{e\in\mathbb{Z}_p} |(A-e)\cap B| \\ &= |\{e \in \mathbb{Z}_p : |(A-e)\cap B| = t\}| \cdot t + |B| = (|B-B|-1)\cdot t + |B| \\ &> (2|B|-2)\cdot t + |B|,\end{aligned}$$

where the strict inequality holds by Vosper's theorem, because B is not an arithmetic progression and $|B| < (p+1)/2$. We cannot have $A = g - \overline{B}$ for some $g \in \mathbb{Z}_p$ since $|A| = |B|$ and p is prime. To get the second equality above, write $|(A-e)\cap B| = \sum_{x\in\mathbb{Z}_p} \chi_A(x+e)\chi_B(x)$ where χ denotes the indicator function of the set. Then change the order of summation. It follows that $|B| > 2t$.

Next, we observe that in the present case, A' and B' cannot be arithmetic progressions with the same common difference. If they were, we would have $|A'| = |B'| > t$ because $|B| > 2t$. The only possible sizes of $(A-e)\cap B$ are 0, t and $|B|$. So if we choose $g' \neq 0$ such that $g' + B' = A'$, then $g' + B = A$ and $g' + (A \cap B) = A \setminus (g' + B') = A \cap B$. This can only happen if $A \cap B = \mathbb{Z}_p$.

Returning to the argument in the proof which produced (6), we have, for each $b^* \in I$,

$$(A' + B') \setminus (b^* + A) \subseteq b^* + \bigcap_{e\in E(b^*)} (B + e).$$

Lemma 8 implies that

$$\Big|\bigcap_{e\in E(b^*)} (B+e)\Big| \le 1$$

but gives no information on the structure of the set $E(b^*)$ when

$$\Big|\bigcap_{e\in E(b^*)} (B+e)\Big| = 1$$

since $|E(b^*)| = |B| - 1$. However, we do have

$$A' + B' \subseteq \Big(\bigcap_{b\in I}(b + A)\Big) \cup F$$

where $|F| \le t$.

Then using our previous two observations we obtain

$$\begin{aligned}\Big|\bigcap_{b\in I}(b+A)\Big| &\ge |A' + B'| - |F| \ge |A'| + |B'| - t \ge |A| + (|B| - 2t) - t \\ &\ge |A| - t + 1,\end{aligned}$$

which by Lemma 8 implies that A is an arithmetic progression as before.

In the remaining case we have $|E'| = 1$, $A = g + B$ for some $g \in \mathbb{Z}_p$ and $|B| \geq (p+1)/2$. Since $|B| \geq (p+1)/2$, we have $(B+e) \cap B \neq \emptyset$ for all $e \in \mathbb{Z}_p$, $e \neq 0$. In the present situation, this means that $|(B+e) \cap B| = t$ for all $e \in \mathbb{Z}_p$, $e \neq 0$, and therefore

$$|(B+e) \cap \overline{B}| = |B| - t \tag{8}$$

for all $e \in \mathbb{Z}_p$, $e \neq 0$. Looking back at (5), we notice that we may assume that $|A' + B'| \leq |A'| + |B'|$. For if $|A' + B'| \geq |A'| + |B'| + 1$ then Lemma 8 implies that U and I are arithmetic progressions with the same common difference. From this we quickly deduce the same for A and B. From the calculation in (7) and Vosper's theorem we have

$$|B|^2 = (p-1)t + |B|$$

and using $|B| \geq (p+1)/2$ we obtain $|B| \leq 2t$.

We shall appeal to the following theorem of Hamidoune and Rødseth from [5] which characterizes pairs of sets (X, Y) in $\mathbb{Z}_p$ for which $|X + Y| = |X| + |Y|$. Call an arithmetic progression with difference d from which one term has been removed an *almost-progression* with difference d. Note that an arithmetic progression with fewer than p elements is also an almost progression.

THEOREM 4. *Suppose that* $|X|, |Y| \geq 3$, *and that*

$$7 \leq |X + Y| = |X| + |Y| \leq p - 4.$$

Then X *and* Y *are almost-progressions with the same difference.*

Represent B as a disjoint union $B = \bigcup_{i=1}^{s} P_i$ where each P_i is an arithmetic progression in $\mathbb{Z}_p$ with difference 1 and any two of these progressions are separated by at least one element from $\overline{B}$. Considering that $|(B+1) \cap \overline{B}| = |(B-1) \cap \overline{B}| = |B| - t$ we see that $s = |B| - t$. The set $\overline{B}$ is also a disjoint union of progressions with difference 1. Since $|(B+d) \cap \overline{B}| = t$ for each $d \neq 0$, we may divide by an appropriate d and assume that one of the progressions with difference 1 which makes up $\overline{B}$ has at least two elements.

If $|A'| + |B'| = |A' + B'| < 7$ then we have $2|B| - 2t < 7$, so that $|B| \leq t + 3$. We must then have $|B| = t + 2$ or $|B| = t + 3$. Considering congruence classes as points on a circle it is easily checked that if either $B = P_1 \cup P_2$ or $B = P_1 \cup P_2 \cup P_3$ then $|(B+e) \cap \overline{B}|$ is not a constant function of $e \neq 0$.

Moreover, since $|B| \leq 2t$, we have $|A'| + |B'| = 2|B| - 2t \leq |B|$. If $|B| = p-3$, then $|\overline{B}| = 3$ and it is once again simple to verify that $|(B+e) \cap \overline{B}|$ cannot be a constant function of $e \neq 0$. We may then assume that

$$7 \leq |A'| + |B'| \leq p - 4.$$

As before, write $B = \bigcup_{i=1}^{s} P_i$ with $s = |B| - t$ and assume, as we may, that $\overline{B}$ contains an arithmetic progression with difference 1 having at least two elements. By translating, we may take $A = B$ and $B' = (B+1) \cap \overline{B}$ and $A' = B \setminus (B+1)$. If $|A'+B'| = |A'|+|B'|-1$ we use Vosper's theorem and if $|A'+B'| = |A'|+|B'|$ we use the Hamidoune–Rødseth theorem to see that we must have $|P_i| = |P_j|$ for $1 \le i, j \le s$. Considering the set B as a subset of the circle, it is clear that $|(B+1) \cap \overline{B}| \neq |(B+2) \cap \overline{B}|$. We have ruled out (8) for all possible sizes of $|B|$ and this completes the proof. ∎

4. Remarks. We briefly indicate two possible directions for further investigation.

(1) In [10], Pollard extended his theorem to the case of sums of h subsets in cyclic groups of composite order. It might be interesting to try to extend the result in this paper to those cases or to the case of general finite abelian groups along the lines of Kneser's theorem (see [8, Chapter 4]).

(2) The Riesz rearrangement inequality (from [11]) states that

$$\iint f(y)g(x-y)h(x)\,dy\,dx \le \iint f^*(y)g^*(x-y)h^*(x)\,dy\,dx$$

where f^*, g^*, and h^* are the spherically decreasing rearrangements of the functions f, g, and h on $\mathbb{R}^n$. (See also [6, Chapter 10].)

A connection with Pollard's inequality arises because

$$|A|\,|B| - S(A,B,t) = \Big(\sum_{s\in\mathbb{Z}_p} \sum_{x\in\mathbb{Z}_p} \chi_B(x)\chi_A(s-x)\chi_{C_t}(s) \Big) - t|C_t| \tag{9}$$

where $C_t = \{s \in \mathbb{Z}_p : r_{A,B}(s) \ge t\}$.

The cases of equality in the Riesz inequality were determined by Burchard in [1]. Our theorem characterizes pairs which maximize the quantity in (9). We do not know if Burchard's theorem implies ours. The connection between Pollard's theorem and various other results in additive number theory and the rearrangement theory treated in [6] is explored in the paper [7].

References

[1] A. Burchard, *Cases of equality in the Riesz rearrangement inequality*, Ann. of Math. 143 (1996), 499–527.

[2] A.-L. Cauchy, *Recherches sur les nombres*, J. École Polytechnique 9 (1813), 99–123.

[3] H. Davenport, *On the addition of residue classes*, J. London Math. Soc. 10 (1935), 30–32.

[4] —, *A historical note*, ibid. 22 (1947), 100–101.

[5] Y. O. Hamidoune and Ø. J. Rødseth, *An inverse theorem mod p*, Acta Arith. 92 (2000), 251–262.

[6] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge Univ. Press, 1959.
[7] V. F. Lev, *Linear equations over $\mathbb{F}_p$ and moments of exponential sums*, Duke Math. J. 107 (2001), 239–263.
[8] M. B. Nathanson, *Additive Number Theory. Inverse Problems and the Geometry of Sumsets*, Grad. Texts in Math. 165, Springer, 1996.
[9] J. M. Pollard, *A generalisation of the theorem of Cauchy and Davenport*, J. London Math. Soc. (2) 8 (1974), 460–462.
[10] —, *Addition properties of residue classes*, ibid. (2) 11 (1975), 147–152.
[11] F. Riesz, *Sur une in égalité intégrale*, ibid. 5 (1930), 162–168.
[12] A. G. Vosper, *The critical pairs of subsets of a group of prime order*, ibid. 31 (1956), 200–205; addendum, ibid. 31 (1956), 280–282.

Claremont McKenna College
Claremont, CA 91711, U.S.A.
E-mail: moneill@mckenna.edu

Received on 24.6.2005
and in revised form on 22.11.2006 (5016)