Sums of Sets

Vladimir Barzov

Abstract

This article is based on a problem from number theory. We present several generalizations and corollaries. The paper studies residues modulo a prime, sets of such residues, and a special type of summing sets. The examples are taken from past mathematical competitions.

1 Background

During the preparation of the US team for the 2001 IMO, the following problem was proposed:

Problem 1 Let d and n be positive integers, and let p be a prime. Prove that there exist integers $x_1, x_2, \ldots, x_d$ such that

$$x_1^d + x_2^d + \cdots + x_d^d \equiv n \pmod{p}.$$

This problem can be solved combinatorially by using a bijection between sets of remainders, or algebraically, by using a minimal polynomial. However, both approaches may not work in a different setting. We will define several new concepts and show that they have interesting properties. Some of these properties can be formulated as separate problems. In addition, the article also gives a refinement of the conclusion of the problem above.

2 Sum of sets

Fix a prime p. For any two nonempty sets A and B of elements of $\mathbb{Z}_p$, define their sum as follows:

$$A + B = \{a + b \mid a \in A, b \in B\}.$$

We can also define translation and multiplication by elements from $\mathbb{Z}_p$ in the following way:

$$t + A = \{t + a \mid a \in A\} \qquad cA = \{ca \mid a \in A\}, \quad c \neq 0.$$

For convenience, define $-A$ to be the set $\{-a : a \in A\}$. A nonempty set $A \subseteq \mathbb{Z}_p$ is called an arithmetic series if there exist elements $a, d \in \mathbb{Z}_p, d \neq 0$ such that

$$A = \{a, a+d, \ldots, a+(|A|-1)d\}.$$

Also, for any two $A, B \subseteq \mathbb{Z}_p$ we will write $A \sim B$ if and only if there exist elements $a, b, d \in \mathbb{Z}_p$, such that

$$A = \{a, a+d, \ldots, a+(|A|-1)d\} \quad \text{and} \quad B = \{b, b+d, \ldots, b+(|B|-1)d\}.$$

Note that the statement $A \sim B$ implies that both A and B are arithmetic series. Let us list some useful properties of arithmetic series.
(i) For every $c, t \in \mathbb{Z}_p$ with $c \neq 0$ and $A, B \subseteq \mathbb{Z}_p$ holds

$$(t+A)+B = t+(A+B), \qquad cA+cB = c(A+B).$$

(ii) For any $c \neq 0$, $A \sim B$ implies $cA \sim cB$ and also the chain of relations:

$$B \sim A \sim t+A \sim -A.$$

(iii) Translation and multiplication do not change the number of elements:

$$|A| = |t+A| = |cA|.$$

Our goal is to find a lower bound for the number of elements in a set which is a sum of two other sets. We will formulate this in the next theorem:

Theorem 1 *For all nonempty subsets $A, B \subseteq \mathbb{Z}_p$,*

$$|A+B| \geq \min(|A|+|B|-1, p).$$

Proof. Let $|A| = m, |B| = n$. We will consider two possibilities.
Case 1: $m+n > p$. Suppose $|A+B| < p$. Then there exists $t \notin A+B$, hence

$$\begin{cases} (t+(-B)) \cap A = \emptyset \\ |t+(-B)| + |A| = m+n > p \end{cases}$$

which is impossible. Therefore

$$|A+B| = p = \min(m+n-1,\, p).$$

Case 2: $m+n \leq p$. We will conduct an induction on $\min(m, n)$. Without loss of generality, assume $m \leq n \leq p$. If $m = 1$, then

$$|A+B| = |B| = |A|+|B|-1 \tag{1}$$

and the statement holds.

Now, suppose that $1 < m \leq n < p$ and take an arbitrary element u of $\mathbb{Z}_p \setminus B$ and two distinct elements a_1, a_2 of A. Since the series

$$u + i(a_2 - a_1), \quad i = 1, 2, \ldots, p-1$$

runs through all elements of $\mathbb{Z}_p \setminus \{u\}$, there exists a smallest i_0 for which the element $b = u + i_0(a_2 - a_1) \in B$. From properties (i) and (iii), we can replace A by the set $(b - a_2) + A$, which now contains $b + (a_1 - a_2) \notin B$ and $b \in B$. Let

$$F = A \cap B, \quad G = A \cup B, \tag{2}$$

with $0 < |F| < m \leq n < |G|$. We have

$$|F| + |G| = |A| + |B| = m + n. \tag{3}$$

At the same time, if $f \in F$ and $g \in G$, then, depending on whether $g \in A$ or $g \in B$, either $(g, f) \in A \times B$ or $(f, g) \in A \times B$. Indeed, if $g \in A$ then $(g, f) \in A \times B$, while if $g \in B$ then $(f, g) \in A \times B$. Both possibilities lead to $f + g \in A + B$. This means that

$$F + G \subseteq A + B \tag{4}$$

and therefore

$$|A + B| \geq |F + G|. \tag{5}$$

Taking into account (3), it remains to prove the statement for the couple F and G. Since $|F| < |A|$, we can apply the induction and the proof is complete. □

Corollary 1

$$|A_1 + A_2 + \cdots + A_k| \geq \min(|A_1| + |A_2| + \cdots + |A_k| - (k-1),\, p).$$

Now we shall establish when equality does hold in the statement of Theorem 1.

Theorem 2 *Let $A, B \subseteq \mathbb{Z}_p$ with $|A|, |B| > 1$ and $|A| + |B| < p$, then*

$$|A + B| = |A| + |B| - 1 \iff A \sim B.$$

Proof. Let $m = |A|, n = |B|$, where $m > 1, n > 1$ and $m + n < p$. Without loss of generality, we can assume that $m \leq n$.

The backward direction is straightforward: if $A \sim B$ and

$$A = \{a, a + d, \ldots, a + (m - 1)d\}, \qquad B = \{b, b + d, \ldots, b + (n - 1)d\},$$

then from $m + n < p$ we have

$$A + B = \{a + b, a + b + d, \ldots, a + b + (m + n - 2)d\},$$

hence $|A + B| = m + n - 1 = |A| + |B| - 1$.

To prove the forward direction, let $|A + B| = |A| + |B| - 1$. We want to prove that $A \sim B$. Observe that from properties (ii) and (iii) it suffices to prove the statement for any translations of A or B or for the sets cA and cB for some $c \neq 0$.

If $a_1 \neq a_2 \in A$ are two arbitrary elements, replace A and B by $c(-a_1 + A)$ and cB respectively, where $c = (a_2 - a_1)^{-1}$. Now we have the convenient fact that

$$0, 1 \in A. \tag{6}$$

We will prove the theorem by induction on m, the size of the smaller set.

If $m = 2$, from (6) we get $A = \{0, 1\}$. Notice that if there existed more than one element $b \in B$ with $b + 1 \notin B$, we would have

$$|A + B| \geq |B| + 2 > |A| + |B| - 1,$$

contradicting the assumption. Therefore B must be a set of consecutive elements, i.e. of the form $x, x + 1, x + 2, \ldots$. Obviously, in this case $A \sim B$ and the statement is proved.

Now, let $3 \leq m \leq n$. We want to apply the earlier idea of replacing A and B with their union and intersection, but this time the induction hypothesis has a stronger requirement, namely $|A \cap B| > 1$. We examine the following cases:

Case 1: There exists a translation of A whose intersection with B has between 2 and $m - 1$ elements, i.e.

$$\exists t \in \mathbb{Z}_p : 2 \leq |(t + A) \cap B| \leq m - 1.$$

Replace A by $t + A$ and define

$$F = A \cap B, \quad G = A \cup B.$$

From the choice of t we have

$$2 \leq |F| \leq m-1. \tag{7}$$

As in the proof of Theorem 1, from (3) and (5) we have

$$|F|+|G|-1 \leq |F+G| \leq |A+B| = |A|+|B|-1 = |F|+|G|-1,$$

so all the inequalities must be equalities, hence

$$|F+G| = |F|+|G|-1 = |A+B| \tag{8}$$

From (7), the induction hypothesis applies to F and G, yielding $F \sim G$. Furthermore, $|G| = |A \cup B| \leq |A|+|B| = m+n < p$, so $G \neq \mathbb{Z}_p$. Therefore if $k = |F|, r = |G|$, we have $1 < k < m \leq n \leq r < p$ and

$$F = \{f, f+d, \ldots, f+(k-1)d\}, \qquad G = \{g, g+d, \ldots, g+(r-1)d\}.$$

where $f-d \notin F, f+kd \notin F, g-d \notin G, g+rd \notin G$. From $F \subset G$ it follows that

$$f = g+qd \tag{9}$$

for some $q \geq 0$, and since $g+rd \notin G$, it must be that $q \leq r-k$. We can now represent G as $G = L \cup F \cup R$, where

$$L = \{g, g+d, \ldots, g+(q-1)d\},$$

$$F = \{g+qd, g+(q+1)d \ldots, g+(q+k-1)d\},$$

$$R = \{g+(q+k)d, g+(q+k+1)d, \ldots, g+(r-1)d\}.$$

Basically L and R consist of the first q and the last $r-k-q$ elements of G respectively. Notice that $L, R \subseteq G \setminus F$, so each of the elements in L and R belongs to either A or B but not to both A and B at the same time.

We shall show that all elements of L (and similarly of R) belong entirely either to A or to B.

If $|L| \leq 1$ there is nothing to prove, so assume $|L| = q > 1$.

First, we shall show that for any two indices $i, j \in [0, q-1]$ with $i+j \in \{q-2, q-1\}$ holds

$$g+id, g+jd \in A \text{ or } g+id, g+jd \in B. \qquad (*)$$

Assume the contrary, and let $g+id \in A, g+jd \in B$ without loss of generality. Again, as in the proof of Theorem 1, we have $F+G \subseteq A+B$ as in (4). From (8) we obtain $A+B = F+G$, hence

$$(g+id)+(g+jd) = (f+i'd)+(g+j'd)$$

for some $i' \in [0, k-1]$ and $j' \in [0, r-1]$. Then, from (9) and $d \neq 0$, we obtain:

$$i + j \equiv i' + j' + q \pmod{p}. \tag{10}$$

Now, using (8), from $m + n < p$ we have $k + r = m + n < p$, and from $i + j \in [q-2, q-1]$ we obtain

$$\begin{aligned} i + j &\leq q - 1 \\ &< i' + j' + q \leq (k-1) + (r-1) + q \\ &< p + q - 2 \leq p + (i + j). \end{aligned}$$

Subtracting $i + j$, we have

$$0 < i' + j' + q - (i + j) < p,$$

which contradicts (10) and completes the proof of $(*)$.

Consider the elements of L arranged in the following order:

$$g + (q-1)d, g, g + (q-2)d, g + d, g + (q-3)d, g + 2d, g + (q-4)d, g + 3d, \ldots$$

Observe that each two consecutive elements are of the form $g + id, g + jd$ where $i + j = q - 1$ or $i + j = q - 2$, so according to $(*)$ they must belong to either A or B. Therefore all elements of L belong simultaneously to either A or to B, which is what we wanted. The proof for R is analogous to the proof of L.

Now, we have $L \subseteq A$ or $L \subseteq B$ and $R \subseteq A$ or $R \subseteq B$. Examine the four possible cases:

- $L, R \subseteq A \Rightarrow (A, B) = (L \cup F \cup R, F)$;
- $L \subseteq A, R \subseteq B \Rightarrow (A, B) = (L \cup F, F \cup R)$;
- $L \subseteq B, R \subseteq A \Rightarrow (A, B) = (F \cup R, L \cup F)$;
- $L, R \subseteq B \Rightarrow (A, B) = (F, L \cup F \cup R)$.

In all examined cases, A and B consist of consecutive elements in G, and therefore $A \sim B$.

Case 2: There *is no* translation of A whose intersection with B has between 2 and $m - 1$ elements. Therefore, any translation that has at least two elements in common with B must have all m elements in B. Formally,

$$|(t + A) \cap B| \geq 2 \Rightarrow |(t + A) \cap B| \geq m = |t + A| \Rightarrow (t + A) \subseteq B. \tag{11}$$

The idea is to remove an element from A so as to use the induction hypothesis.
Let k, $1 \leq k \leq n < p$ be the maximum length of a block of consecutive elements in B, i.e.

$$w, w+1, \ldots, w+k-1 \in B \text{ and } w-1, w+k \notin B$$

for some $w \in \mathbb{Z}_p$. Next, replace B by its translation $-w+B$ so that now

$$0, 1, \ldots, k-1 \in B, \qquad -1, k \notin B. \tag{12}$$

Suppose first that there exists an element $q \neq 0$ in $\mathbb{Z}_p$ with

$$q \in A, \qquad -q \in B, \tag{13}$$

and let $A_1 = -q + A$. From (6), (12), and (13) we have

$$\{-q, -q+1, 0\} \subseteq A_1, \qquad \{-q, 0\} \subseteq B. \tag{14}$$

Then, $|A_1 \cap B| \geq 2$, and from (11) we get $A_1 \subseteq B$, thus $-q+1 \in B$.
Next, let $A_2 = -q+1+A$, so

$$\{-q+1, -q+2, 1\} \subseteq A_2, \qquad \{-q+1, 1\} \subseteq B.$$

Again, $|A_2 \cap B| \geq 2$ hence $A_2 \subseteq B$, thus $-q+2 \in B$.
We proceed the same way with $A_3, A_4, \ldots$ until we reach $A_k = -q+k-1+A$ with

$$\{-q+k-1, -q+k, k-1\} \subseteq A_k, \qquad \{-q+k-1, k-1\} \subseteq B.$$

One last time, $|A_k \cap B| \geq 2$, and from (11) we get $A_k \subseteq B$, thus $-q+k \in B$. Now, we have found $k+1$ consecutive elements in B, namely $-q, -q+1, \ldots, -q+k$, which contradicts the maximality choice of k. Therefore, the existence of q in (13) is not possible, and from (6) and (12)

$$A \cap (-B) \quad = \quad \{0\} \tag{15}$$

Define

$$A^* = A \setminus \{0\},$$

where $|A^*| = |A| - 1 = m - 1 > 1$ since $m \geq 3$. Notice that from $0 \notin A^*$ and (15) we get

$$0 \notin A^* + B, \tag{16}$$

else there would exist a nonzero element in both A and $-B$ in contradiction to (15).

We know that $0 \in A$ and $0 \in B$, so $0 \in A + B$, while at the same time $0 \notin A^* + B$. Therefore

$$A^* + B \subseteq A + B \setminus \{0\}. \tag{17}$$

From Theorem 1, we have

$$(m-1) + n - 1 \leq |A^* + B| < |A + B| = m + n - 1,$$

whence

$$|A^* + B| = m + n - 2 = |A^*| + |B| - 1 = |A + B| - 1. \tag{18}$$

Applying the induction hypothesis to A^* and B yields $A^* \sim B$. Let

$$A^* = \{a, a + d, \ldots, a + (m-2)d\}, \quad B = \{b, b + d, \ldots, b + (n-1)d\} \tag{19}$$

for some $a, b, d \in \mathbb{Z}_p$ where $d \neq 0$. We have

$$A^* + B = \{a + b + id \mid i = 0, 1, \ldots, (m + n - 3)\}. \tag{20}$$

From (17) and (18) we know that $A^* + B \subseteq A + B \setminus \{0\}$ and $|A^* + B| = |A + B \setminus \{0\}| = m + n - 2$, hence

$$A^* + B = A + B \setminus \{0\}. \tag{21}$$

From (6) we have $0 \in A$, so $B = 0 + B \subseteq A + B$, and together with (21) we get

$$B \setminus \{0\} \subseteq (A + B) \setminus \{0\} = A^* + B. \tag{22}$$

From (12) we have $0 \in B$, so from (16) and (19) there exists $r \in [0, n-1]$ such that

$$b + rd = 0 \notin A^* + B \tag{23}$$

We shall show that $r = 0$ or $r = n - 1$. Assume the contrary, so $0 < r < n-1$. In that case, from (19) and (23) the elements $b+(r-1)d = -d$ and $b + (r+1)d = d$ are nonzero and belong to B, so using (22) we obtain

$$-d, d \in A^* + B. \tag{24}$$

From $d \in A^* + B$ and (20) there must be $i \in [0, m + n - 3]$ such that $a + b + id = d$. But if $i > 0$, we would have $a + b + (i-1)d = 0 \in A^* + B$

in contradiction to (16). Therefore $i = 0$, hence $a = d - b = (r+1)d$ owing to (23). Similarly, from $-d \in A^* + B$ and (20), there must be $j \in [0, m+n-3]$ such that $a + b + jd = -d$. But if $j < m+n-3$, we would have $a + b + (j+1)d = 0 \in A^* + B$ in contradiction to (16). Therefore $j = m+n-3$, hence $a = -b - (m+n-2)d = (r-m-n+2)d$ owing to (23). Finally, coupling $a = (r-m-n+2)d$ with $a = (r+1)d$ and $d \neq 0$ we get $m + n \equiv 1 \pmod{p}$, contradicting $2 \leq m+n < p$.

Consequently, $r = 0$ or $r = n-1$.
Consider first the case $r = 0$, so $b = 0$ due to (23). Since $n \geq 2$ and $d \neq 0$, using (19) we have $d = b + d \in B \setminus \{0\}$. From (22) we now have

$$d \in A^* + B.$$

Therefore, from (20) there must be $i \in [0, m+n-3]$ such that $a+b+id = d$. But if $i > 0$, we would have $a + b + (i-1)d = 0 \in A^* + B$ in contradiction to (16). Therefore $i = 0$, hence $a = d - b = d$. Now, using (19) and $A = A^* \cup \{0\} = A^* \cup \{a - d\}$, we obtain

$$A = \{a - d, a, a + d, \ldots, a + (m-2)d\},$$

hence $A \sim B$.
Similarly, we consider the case $r = n-1$, where $b = -(n-1)d$. Since $n \geq 2$ and $d \neq 0$, using (19) we have $-d = b + (n-2)d \in B \setminus \{0\}$. From (22) we now have

$$-d \in A^* + B.$$

Therefore, from (20) there must be $j \in [0, m+n-3]$ such that $a+b+jd = -d$. But if $j < m+n-3$, we would have $a + b + (j+1)d = 0 \in A^* + B$ in contradiction to (16). Therefore $j = m+n-3$, hence $a = -(m-1)d$. Again, using (19) and $A = A^* \cup \{0\} = A^* \cup \{a + (m-1)d\}$, we obtain

$$A = \{a, a + d, \ldots, a + (m-2)d, a + (m-1)d\},$$

and again $A \sim B$.
This completes the proof of the theorem. □

Here is a question to explore: describe all non-empty $A, B \subseteq \mathbb{Z}_p$ such that

$$|A + B| = |A| + |B| - 1. \qquad (**)$$

- If $|A| = 1$, then $|A + B| = |B| = |A| + |B| - 1$, so $(**)$ holds for any $B \subseteq \mathbb{Z}_p$. Likewise, $(**)$ holds for any A and any B with $|B| = 1$.

- If $|A|+|B| > p+1$, from $|A+B| \le p < |A|+|B|-1$ we see that $(**)$ never holds.

- If $|A|+|B| = p+1$, then Theorem 1 gives us $|A+B| \ge p$, so $|A+B| = p = |A|+|B|-1$, and hence $(**)$ is true for any $A, B \subseteq \mathbb{Z}_p$ with $|A|+|B| = p+1$.

- If $|A| > 1$ and $|B| > 1$ and $|A|+|B| < p$, then Theorem 2 tells us that $(**)$ iff $A \sim B$.

- Finally, if $|A|+|B| = p$, then $(**)$ is equivalent to $|A+B| = p-1$, i.e. $A+B = \mathbb{Z}_p \setminus \{t\}$ for some $t \in \mathbb{Z}_p$, which in turn is equivalent to

$$A \cap (t+(-B)) = \emptyset. \tag{25}$$

  Since $|t+(-B)| = |B| = p-|A|$, then (25) implies that A and $t-B$ are complementary sets. Therefore

$$t - B = (\mathbb{Z}_p \setminus A) \Rightarrow B = t - (\mathbb{Z}_p \setminus A).$$

  Conversely, we shall see that every $A \subset \mathbb{Z}_p$ and $B = t-(\mathbb{Z}_p \setminus A)$ satisfy $(**)$. Indeed, in this case $p-1 = |A|+|B|-1 \le |A+B|$. At the same time, $t = a+b$ does not have a solution in $a \in A$ and $b = t-a' \in B$, otherwise we would have $t = a+(t-a') \leftrightarrow a = a'$ for some $a' \in \mathbb{Z}_p \setminus A$, a contradiction. Therefore $t \notin A+B$, hence $|A+B| < p$, and so $|A+B| = p-1 = |A|+|B|-1$, satisfying $(**)$.

3 Applications

We are ready to solve the problem we formulated in the beginning.
Solution: Let $A_d = \{0, 1^d, 2^d, \ldots, (p-1)^d\}$. Since the equation $x^d - c = 0, c \neq 0$ has no more than d roots in $\mathbb{Z}_p$, among the $p-1$ numbers $1^d, 2^d, \ldots, (p-1)^d$ at least $\dfrac{p-1}{d}$ are distinct, whence $|A_d| \ge 1 + \dfrac{p-1}{d}$. Now, from the corollary of Theorem 1 we have

$$|\underbrace{A_d + A_d + \cdots + A_d}_{d\text{-times}}| \ge \min(d|A_d| - (d-1), p) = p.$$

In particular, $n \in (A_d + A_d + \cdots + A_d)$ and we are done. □

Here are several more problems for practice.

Problem[1] **2** Let p be a prime and consider a regular $2p$-gon. Let A and B be opposite vertices. Among the remaining vertices, $k < p$ are marked. Draw all segments joining pairs of marked vertices that intersect AB, and assume that at least one such segment exists. Prove that at least $k-1$ of these segments can be chosen so that no two are parallel or perpendicular.

Solution Let us number the points clockwise from A inclusive to B respectively by $0, 1, \ldots, p-1$ and call them "left" points. We do the same with the "right" points, i.e. from B to A in the same direction. Let us take one of the mentioned diagonals with left end a_1 and right end b_1. It is parallel to another diagonal with ends a_2 and b_2 whenever $a_1 + b_1 = a_2 + b_2$, and it is perpendicular to it iff $a_1 + b_1 = a_2 + b_2 \pm p$. In order to find $k-1$ diagonals with the desired property we need at least $k-1$ incongruent modulo p sums of the type $a+b$, which is actually the result of Theorem 1. □

Problem 3 Let $p > 2$ be prime and let $a(x)$, $b(x)$, and $c(x)$, be three nonzero polynomials of degree less than p having all coefficients in $\{0, 1\}$.

Prove that if $a(1) + b(1) + c(1) = p + 2$, then

$$\sum_{j=0}^{p-1} a(\omega^j) b(\omega^j) c(\omega^j) > 0,$$

where ω is a primitive p-th root of 1.

Solution: Since

$$\sum_{j=0}^{p-1} \omega^{jr} = \begin{cases} p, & p \mid r, \\ 0, & p \nmid r, \end{cases}$$

the sum in the problem statement equals $p\times$ the sum of the coefficients of those monomials in $s(x) = a(x)b(x)c(x)$ whose exponents are divisible by p. Since all coefficients of $s(x)$ are nonnegative coefficients, we just need to show that one of the monomials $x^{jp}, j \geq 0$ appears in $s(x)$.

Let A be the set of exponents n such that the monomial x^n appears in $a(x)$. Same for B and C. From $a(1) + b(1) + c(1) = p + 2$ it follows that $|A| + |B| + |C| = p + 2$. Then, from the corollary of Theorem 1, we get $A + B + C = \mathbb{Z}_p$, and in particular $0 \in A + B + C$. That means there exist exponents $\alpha \in A, \beta \in B$, and $\gamma \in C$ such that

$$\alpha + \beta + \gamma \equiv 0 \pmod{p}.$$

[1] This and the next problems are equivalent to Theorem 1

Since $0 \leq \alpha, \beta, \gamma \leq p-1$, it follows that $\alpha+\beta+\gamma = kp$ for some integer $k \in [0,2]$. Therefore the polynomial $s(x)$, contains $x^{\alpha+\beta+\gamma} = x^{kp}$ and the proof is complete. □

Problem 4 (Selection test for JBMO 1998) Given are seven integers not divisible by seven. Prove that all their sums represent all seven remainders modulo seven.

Hint: If the numbers are a_i, consider the sets $A_i = \{0, a_i\}$.

Problem 5 Find all prime numbers p with the property: There exists only one element among $\{0, 1, \ldots, p-1\}$ which is not a sum of a quadratic residue and a non-quadratic residue modulo p.

4 Special cases

4.1 Geometric series

We will examine the sets $A_d = \{0, 1^d, 2^d, \ldots, (p-1)^d\}$ from the first problem more closely. Let ξ be a primitive root in $\mathbb{Z}_p$ and define

$$S_n = \underbrace{A_d + A_d + \cdots + A_d}_{n\text{-times}}.$$

Next, define

$$h(d) = \min n : S_n = \mathbb{Z}_p.$$

We want to find an upper bound on $h(d)$.

If $\delta = (p-1, d)$, we shall see that $A_d = A_\delta$. There exist integers a and b, such that $(p-1)a + db = \delta$. Therefore, for every $x \in \mathbb{Z}_p \setminus \{0\}$ and some u we have

$$x^\delta = (\xi^u)^\delta = \xi^{(p-1)au+dbu} = (\xi^{bu})^d = y^d,$$

hence every non-zero element in A_δ is also in A_d.

Conversely, $x^d = (x^{d/\delta})^\delta = y^\delta$, so every non-zero element in A_d is also in A_δ. Finally, given that $0 \in A_d$ and $0 \in A_\delta$, we have $A_\delta = A_d$. Therefore it suffices to examine only those d that divide $p-1$. Let

$$m = \frac{p-1}{d}.$$

Notice that

$$A_d = \{0, \xi^d, \xi^{2d}, \ldots, \xi^{md}\}. \tag{26}$$

There are several possibilities for d and m where either $d \leq 2$ or $m \leq 2$:

- $m = 1, d = p - 1 \Rightarrow A_d \equiv \{0, 1\} \Rightarrow h(d) = p - 1$;
- $m = 2, d = \dfrac{p-1}{2} \Rightarrow A_d = \{-1, 0, 1\} \Rightarrow h(d) = \dfrac{p-1}{2}$;
- $d = 1 \Rightarrow A_d = \mathbb{Z}_p \Rightarrow h(d) = 1$;
- $d = 2 \Rightarrow |A_d| = m + 1 = \dfrac{p+1}{2} \Rightarrow h(d) = 2$.

In all four cases we have $h(d) = d$. Now, consider $m, d > 2$, i.e.

$$3 \leq d \leq \frac{p-1}{3}, \qquad 3 \leq m \leq \frac{p-1}{3}. \tag{27}$$

We shall show that A_d is *not* an arithmetic series. Assume the contrary, and let

$$\{a, a + t, \ldots, a + mt\} = A_d = \{0, \xi^d, \xi^{2d}, \ldots, \xi^{md}\} \tag{28}$$

for some $a, t \in \mathbb{Z}_p$. From (27) we have $\xi^d \neq \pm 1$. Operating in $\mathbb{Z}_p$ and using (28) we obtain:

$$0 = \sum_{i=1}^{m} \xi^{id} = \sum_{i=0}^{m} (a + it) = (m + 1)a + \frac{m(m+1)}{2} t, \tag{29}$$

$$0 = \sum_{i=1}^{m} (\xi^{id})^2 = \sum_{i=0}^{m} (a+it)^2 = (m+1)a^2 + \frac{m(m+1)(2m+1)}{6} t^2 + m(m+1)at. \tag{30}$$

Since $m + 1 \neq 0$, from (29) it follows $a = -mt/2$, which we plug into (30) to obtain

$$0 = -\frac{m^2}{4} + \frac{m(2m+1)}{6}$$

after dividing by $(m+1)t^2 \neq 0$. From $m \neq 0$, this leads to $m \equiv -2 \pmod{p}$, hence $m = p-2$, which is a contradiction with (27). Consequently the equivalence (28) is not possible, and we can conclude that A_d is not an arithmetic series.

We know that $|S_k| < p$ for $k = 1, 2, \ldots, h(d) - 1$ by the definition of $h(d)$. Consider the bijection $g : (S_k \setminus \{0\}) \to (S_k \setminus \{0\})$, defined by $g(x) = \xi^d x$. Each orbit in this bijection has length m, therefore

$$m \mid |S_k| - 1. \tag{31}$$

We will prove by induction on k that $|S_k| \geq (2k-1)m+1$ for $k = 1, 2, \ldots, h(d)-1$.

For $k=1$, we get $|S_k|-1 \geq m$ from (31), hence the start of the induction is established.
From now on, assume $1 < k < h(d)$ and $|S_{k-1}| \geq (2k-3)m+1$. From Theorem 1 we have

$$|S_1| + |S_{k-1}| - 1 \leq |S_1 + S_{k-1}| = |S_k| \leq p-1,$$

so

$$|S_1| + |S_{k-1}| \leq p. \tag{32}$$

Next, from (31), we have $|S_i| \equiv 1 \pmod{m}$, so $|S_1| + |S_{k-1}| \equiv 2 \pmod{m}$, where $m > 2$ owing to (27). At the same time, $p \equiv 1 \pmod{m}$, hence $|S_1| + |S_{k-1}| \neq p$, which means the inequality in (32) is strict.

Now we have $|S_1| + |S_{k-1}| < p$, and from $|S_1| > 1, |S_{k-1}| > 1$ we can apply Theorem 2 to these two sets. We already know that $S_1 = A_d$ is not an arithmetic series, hence

$$|S_k| \geq |S_1| + |S_{k-1}| \geq m+1+(2k-3)m+1 = (2k-2)m+2.$$

Now, from (31) we conclude $|S_k| \geq (2k-1)m+1$ and the induction is complete.

Now, applying the result of the induction to $k = h(d)-1$,

$$\begin{aligned} p > |S_k| \geq (2k-1)m+1 = (2k-1)\frac{p-1}{d} + 1 \\ \Rightarrow d > 2k-1 \Rightarrow d \geq 2k = 2(h(d)-1) \\ \Rightarrow h(d) \leq \frac{d}{2} + 1 \Rightarrow h(d) \leq \left\lfloor \frac{d}{2} \right\rfloor + 1. \end{aligned}$$

Since we used the limits on d specified in (27), this result is not guaranteed for the 4 cases $d \leq 2$ and $m \leq 2$ that we examined earlier. However, we can verify manually that it also holds for $d = 1, 2$, where $h(1) = 1$ and $h(2) = 2$. We summarize this result in a theorem:

Theorem 3 *For every positive integer d,*

$$h(d) \leq \left\lfloor \frac{(p-1, d)}{2} \right\rfloor + 1,$$

with the exception of $(p-1, d) = \dfrac{p-1}{2}, p-1$, where $h(d) = (p-1, d)$.

Comment The sets $A_d \setminus \{0\}$ are the subgroups of $\mathbb{Z}_p^*$.

The next problem illustrates the result achieved.

Problem 6 Let n be the smallest positive integer for which $2^n - 1$ is divisible by the prime $p > 3$. Prove that for every integer d, there exists an integer $m, 1 \leq m \leq 2^n$ such that p divides $d + m$ and the binary representation of m has no more than $\left\lfloor \frac{p-1}{2n} \right\rfloor + 1$ units.

Comment. If we change $\left\lfloor \frac{p-1}{2n} \right\rfloor + 1$ to $\frac{p-1}{n}$, the problem becomes much easier and the solution requires only Theorem 1.

4.2 Arithmetic series

The arithmetic series sets A are easily described by complex numbers. Let ω be a primitive p-th root of 1. Define the function $\tau(A) : \{A : A \subset \mathbb{Z}_p\} \to \mathbb{C}$ by

$$\tau(A) = \sum_{a \in A} \omega^a.$$

It can be shown that the minimal polynomial of $\tau(A_d)$ over $\mathbb{Q}$ has degree $m = (p-1, d)$, which allows us to solve the first problem without using the theorems. Unfortunately, in the general case $h(A) < \deg_{\mathbb{Q}} \tau(A)$. We know that the polynomial $\Phi_p(x) = 1 + x + \cdots + x^{p-1}$ is irreducible over $\mathbb{Q}$. Therefore, we are able to identify the sets by their τ-values, namely $A = B \iff \tau(A) = \tau(B)$. Consequently, the assertion that A is an arithmetic series implies that there exist $a \in [0, p-1]$ and $d \in [1, p-1]$ with

$$\tau(A) = \omega^a + \omega^{a+d} + \omega^{a+2d} + \cdots + \omega^{a+(m-1)d} = \frac{\omega^{a+md} - \omega^a}{\omega^d - 1},$$

where $m = |A|$. When $m = 1$ or $m = p-1$ everything is clear. Assume now $1 < m < p-1$. We will prove that there exist exactly two couples $\{a, d\}$ generating the arithmetic series A. Let

$$\frac{\omega^{a+md} - \omega^a}{\omega^d - 1} = \tau(A) = \frac{\omega^{a'+md'} - \omega^{a'}}{\omega^{d'} - 1}.$$

Move everything to the left and notice that every term with sign "+" needs to be cancelled by the *same* term with sign "-":

$$\omega^{a+md+d'} + \omega^a - \omega^{a+md} - \omega^{a+d'} - \omega^{a'+md'+d} - \omega^{a'} + \omega^{a'+md'} + \omega^{a'+d} = 0.$$

Since $\omega^a \neq \omega^{a+md}, \omega^{a+d'}$ and $\omega^{a+md+d'} \neq \omega^{a+md}, \omega^{a+d'}$, then either

$$\left(\omega^{a+md+d'}, \omega^a\right) = \left(\omega^{a'+md'+d}, \omega^{a'}\right),$$

or

$$\left(\omega^{a+md+d'}, \omega^a\right) = \left(\omega^{a'}, \omega^{a'+md'+d}\right),$$

where the pairs are ordered. The first case leads to $\omega^{(m-1)(d-d')} = 1$, hence $d = d'$ and $a = a'$; and the second case leads to $\omega^{(m+1)(d+d')} = 1$, hence $d' = p - d$ and $a' \equiv a + (m-1)d \pmod{p}$. This shows that each arithmetic series A with $1 < |A| < p - 1$ is given by the pairs (a, d) and $(a', p - d)$.

Corollary 2 *If $A, B, C \subset \mathbb{Z}_p$ and $1 < |B| < p - 1$, then*

$$A \sim B \sim C \quad \Rightarrow \quad A \sim C.$$

Corollary 3 *For a given p, the number of all arithmetic series except $\mathbb{Z}_p$ is*

$$p + \frac{p(p-1)(p-3)}{2} + p = \frac{p(p^2 - 4p + 7)}{2}.$$

We can define arithmetic series for every positive integer p, not necessarily prime. However, when $p = mn$ for $m, n > 1$ all theorems become false, as shown by the following example:

$$A = \{0, n, 2n, \ldots, (m-1)n\}.$$

We have $A + A = A$ and $m = |A| = |A + A| < |A| + |A| - 1$, but $A + A \neq \mathbb{Z}_p$. On the other hand, the first two theorems remain true when we remove p from the stem and consider instead the numbers in $\mathbb{Z}$:

Theorem 4 *For every two sets A and B of integers holds*

$$|A + B| \geq |A| + |B| - 1,$$

with equality holding for $|A|, |B| > 1$ whenever A and B are arithmetic series with the same common difference.

We conclude with a nice problem about arithmetic series in $\mathbb{Z}_p$.

Problem 7 Let $p > 5$ be prime, and a, d, n be integers for which $2 \leq d \leq p - 2$ and $3 \leq n \leq p - 3$. It is known that for each $j = 1, 2, \ldots, n$ there exists an integer $b_j \in [0, n]$ with $b_j \equiv a + jd \pmod{p}$. Prove that $n = p - 3$.

Hint Consider the numbers $a + jd$ geometrically as points on a circle.