ACTA ARITHMETICA
137.4 (2009)

Inverse Additive Number Theory. XI. Long arithmetic progressions in sets with small sumsets

by

Gregory A. Freiman (Tel Aviv)

This paper continues the series of papers on Inverse Additive Number Theory published in 1955–1964 (see references [84]–[92], [98] in [2]).

Throughout the paper, we work with the set $A \subset \mathbb{Z}$ of cardinality $|A| = k \geq 3$. We assume that

$$A = \{a_0 = 0 < a_1 < \cdots < a_{k-1}\}$$

and that the greatest common divisor of the numbers from A is 1. Let T denote the cardinality of the set $2A = A + A$ of all pairwise sums $a + b$ of numbers from A. Notice that $T \geq 2k - 1$.

In [1] (see also the textbook [3, p. 204]), we proved the following result.

Theorem 1. *For $0 \leq b < k-2$ and $T = 2k-1+b$, the set A is contained in*

$$L = \{0, 1, 2, \ldots, k + b - 1\}. \ \blacksquare \tag{1}$$

Let us give several examples of such sets for the maximal value of $b = k - 3$, $T = 3k - 4$ and $k = 8$:

$$A = \{0, 2, 4, 6, 8, 10, 11, 12\}, \tag{2}$$

$$A = \{0, 2, 4, 6, 7, 8, 10, 12\}, \tag{3}$$

$$A = \{0, 6, 7, 8, 9, 10, 11, 12\}. \tag{4}$$

The fact that a set A with a small doubling (small T) may be included in a short interval reflects only part of the whole picture.

In order to formulate the main result of the paper we define several new notions.

Let e denote the maximal $a \in [0, a_{k-1}]$ with $a \notin 2A$; if the interval $[0, a_{k-1}]$ is included in $2A$, then we put $e = -1$.

2000 *Mathematics Subject Classification*: Primary 11P70.
Key words and phrases: inverse additive problems, doubling of the set.

DOI: 10.4064/aa137-4-2

Let c denote the minimal $a \in [0, a_{k-1}]$ with $a + a_{k-1} \notin 2A$; if the interval $[a_{k-1}, 2a_{k-1}]$ is included in $2A$, then we put $c = a_{k-1} + 1$.

In Lemma 6 we show that one always has

$$e < c.$$

We also need the following definition: the set A is called *stable* if

$$2A \cap [0, a_{k-1}] = A.$$

Examples of stable sets: $\{0,6\}$, $\{0,2,4,6\}$, $\{0,3,4,5\}$.

Define

$$B = A \cup (a_{k-1} + A).$$

We have $B \subset M_1$, where $M_1 = [0, 2a_{k-1}]$.

Let C be a set of integers. If $x \in [\min C, \max C] \setminus C$, then we say that x is a *hole in* C. For example, in (2) we have $A \subseteq M = [0, 12]$ and the set of holes is $\{1,3,5,7,9\}$. Note that if a is a hole in A, then $(a, a + a_{k-1})$ is a pair of holes in B; in what follows we will use *only* such pairs, i.e. $a \notin A$.

We can now formulate the main result of this paper:

Theorem 2. *In the setting of Theorem* 1, *we have*

$$J = [e + 1, c + a_{k-1} - 1] \subset 2A,$$

and

$$|J| \geq 2k - 1 + 2d,$$

where d is the number of holes in A in the open interval (e, c).

The most interesting result is when we assume that the interval containing A has the maximal length for a given T. The following assertion is a consequence of Theorem 2:

Corollary 1. *If $T = |A + A| = 2k - 1 + b$ where $0 \leq b < k - 2$ and if*

$$a_{k-1} = k - 1 + b,$$

then

(a) $A_1 = A \cap [0, e + 1]$ *is stable*, *i.e.* $A \cap [0, e] = 2A \cap [0, e]$,
(b) $A_2 = a_{k-1} - ([c - 1, a_{k-1}] \cap A)$ *is stable*,
(c) $J = [e + 1, c + a_{k-1} - 1] \subset 2A$,
(d) $I = [e + 1, c - 1] \subset A$.

We see that in this case the set A may be partitioned into three parts,

$$A = A_1 \cup I \cup (a_{k-1} - A_2),$$

where A_1 and A_2 are stable, and I is an interval, and the set $2A$ may be partitioned into three parts,

$$2A = A_1 \cup J \cup (2a_{k-1} - A_2),$$

where J is an interval.

We define the *length* of an interval of integers (or of an arithmetic progression) to be the number of his elements. So, the length of L in (1) is $|L| = |[0, k+b-1]| = k+b$.

We denote by M or $M(A)$ the minimal interval containing A. From Theorem 1 it follows that

$$a_{k-1} = k - 1 + b', \tag{5}$$

where

$$b' \le b. \tag{6}$$

Thus, the length of $M = [0, a_{k-1}]$ is equal to $k + b'$. We will now estimate b' from below. From $A \subset [0, k-1+b']$ it follows that

$$2A \subseteq 2[0, k-1+b'] = [0, 2k-2+2b']$$

and

$$|2A| \le |[0, 2k-2+2b']| = 2k-1+2b'.$$

Thus, from $|2A| = T = 2k-1+b$, we get

$$b' \ge b/2. \tag{7}$$

From $A \subset M = [0, a_{k-1}]$ and (5), we see that the number of holes in A is equal to b'. We have

$$B \subset M_1 = [0, 2a_{k-1}]. \tag{8}$$

From $|B| = 2k-1$ and

$$|M_1| = 2a_{k-1} + 1 = 2(k-1+b') + 1 = 2k-1+2b', \tag{9}$$

it follows that the number of holes in B is equal to $2b'$.

The following Lemmas 2–6 will be used in the proof of Theorem 2.

LEMMA 2. *For each pair* $(a, a+a_{k-1})$ *of holes in* B *we have either*

$$a \in 2A, \tag{10}$$

or

$$a + a_{k-1} \in 2A. \tag{11}$$

Proof. Let us look at A as a set of residues modulo a_{k-1}. Our modulus, a_{k-1}, has $k+b'-1 \le k+b-1 \le 2k-4$ residues, and the sets $A \pmod{a_{k-1}}$ and $a - A \pmod{a_{k-1}}$ contain $k-1$ residues each, because the numbers 0 and a_{k-1} are congruent modulo a_{k-1}. Thus, the sets of residues A and $a-A$ have a non-zero intersection, and therefore

$$a \in 2A \pmod{a_{k-1}}. \tag{12}$$

But in the set of integers the residue a is represented by a or by $a + a_{k-1}$. If neither of these numbers belongs to $2A$ then this contradicts (12). Therefore we have (10) or (11). ∎

For the pair $(a, a + a_{k-1})$ of Lemma 2, one of the numbers of the pair belongs to $2A$. And the other one?

DEFINITION. If both numbers in the pair $(a, a+a_{k-1})$ belong to $2A$, i.e. (10) and (11) are valid, we call the pair *unstable*. This pair is called *stable* if one of the numbers of the pair does not belong to $2A$, and this number will be called a *stable hole*. If

$$a \notin 2A, \tag{13}$$

the pair will be called *left*; if

$$a + a_{k-1} \notin 2A, \tag{14}$$

the pair will be called *right*.

The number and location of pairs of different types depends to a large extent, as we will see, on the structure of both $2A$ and A.

The number

$$b' = a_{k-1} - k + 1$$

represents the number of holes in A and at the same time the number of pairs of holes $(a, a + a_{k-1})$ in B.

LEMMA 3. *In B, there are $2b' - b$ stable pairs of holes and $b - b'$ unstable pairs of holes.*

Proof. The number of holes in B is equal to $2b'$. To get all $2k - 1 + b$ numbers of $2A$ we have to add, to the $2k - 1$ numbers of B, b more numbers, which are holes in B, so that the number of stable holes is equal to $2b' - b$, and the same is the number of stable pairs (one stable hole in a stable pair). The whole number of pairs of holes in B is equal to b'. The number of unstable pairs is equal to

$$b' - (2b' - b) = b - b'. \ \blacksquare$$

In the next two lemmas, which are immediate consequences of the pigeon-hole principle, we begin to explain why the holes in A under the conditions of Theorem 1 are concentrated in the neighborhoods of the endpoints of $M = [0, a_{k-1}]$.

LEMMA 4. *If $a \notin 2A$, then the number of holes of A which belong to the interval $[0, a]$ is greater than or equal to $[a/2 + 1]$.*

LEMMA 5. *The number of holes in A which belong to an interval $I = [a, a_{k-1}]$ when $a + a_{k-1}$ is a right stable hole is greater than or equal to $[(a_{k-1} - a)/2 + 1]$.*

We are now ready to prove that the numbers in the set of left stable holes are smaller than the numbers in the set of right stable holes; the set of numbers between these two sets in A contains only holes which are unstable, and this ensures the existence of a long interval in $2A$.

LEMMA 6. *We have*

$$e < c.$$

Proof. We know that e is stable in a left stable pair and so $e \notin 2A$, and from the fact that c is stable in a right stable pair $(c, c + a_{k-1})$, we get $c + a_{k-1} \notin 2A$.

If $e = c$, then the pair $(e, e + a_{k-1})$ would have neither element in $2A$, in contradiction to Lemma 1.

Suppose now, contrary to the conclusion, that $e > c$.

The number of holes in A is equal to b'. We will estimate this number from below, using estimates of the values e and c.

Now we build a finite sequence of pairs of numbers

$$(c_1, e_1), \ldots, (c_i, e_i) \tag{15}$$

in the following manner.

Define $c_1 = c$, $e_1 = e$. Suppose that the pair (c_j, e_j) is already built, where c_j is a stable point from a right stable pair and e_j is a stable point from a left stable pair. There are the following possibilities:

(i) There exists a left stable pair $(a, a + a_{k-1})$ such that $c_j < a < e_j$.
(ii) Case (i) is not valid but there exists a right stable pair $(a, a + a_{k-1})$ such that $c_j < a < e_j$.
(iii) Cases (i) and (ii) are not valid.

In case (i) put $c_{j+1} = c_j$, $e_{j+1} = a$; if (ii) is true put $c_{j+1} = a$, $e_{j+1} = e_j$; if (iii) is true put $j = i$, and the sequence is built. Let us mention that the sequence (15) was built in such a way that

$$c_j < e_j, \quad j = 1, \ldots, i, \quad [c_1, e_1] \supset \cdots \supset [c_i, e_i].$$

Denote by x the number of holes in A which belong to the interval (c_i, e_i). All these holes, because of the manner in which we built them, are unstable, and we have, because of Lemma 4, an estimate

$$x \le b - b'. \tag{16}$$

We clearly have

$$x \le e_i - c_i - 1. \tag{17}$$

The holes in A which are in the interval $[c_i, e_i]$ are perhaps counted twice when we estimate the number of holes in A belonging to $[1, e_i]$ with the help of Lemma 4 and when we estimate the number of holes in A belonging to $[c_i, a_{k-1}]$ with the help of Lemma 5. Putting all what has been said together we obtain the inequality

$$b' \ge (e_i + 1)/2 + (a_{k-1} - c_i + 1)/2 - x - 2.$$

In view of (5), we get

$$b' \geq (k+b')/2 + (e_i - c_i)/2 - 3/2 - x$$

and therefore

$$b' \geq k + e_i - c_i - 3 - 2x.$$

Using (17) we get

$$b' \geq k - 2 - x + e_i - c_i - 1 - x \geq k - 2 - x.$$

Because of (16) we have

$$0 \geq k - b + (b - b') - x - 2 \geq k - b - 2 \geq k - (k-3) - 2 = 1,$$

a contradiction. ∎

Proof of Theorem 2. We will now use Lemmas 4–6 to estimate the length of the interval contained in $2A$. We will show that

$$J = [e+1, c + a_{k-1} - 1] \subset 2A \tag{18}$$

and

$$|J| \geq 2k - 1 + 2d, \tag{19}$$

where d is the number of holes in A in the interval (e, c).

We first prove that (18) is valid. Let $f \in J$. If $f \in B$ then $f \in 2A$, because $B \subseteq 2A$. If $f \notin B$, then f is one of the numbers of the pair $(a, a + a_{k-1})$. If this pair is unstable, then both numbers in it belong to $2A$ and so $f \in 2A$. If this pair is left stable, then $a \notin 2A$ and $a \leq e$. Thus, $f = a + a_{k-1} \in 2A$. If this pair is right stable, then $a + a_{k-1} \notin 2A$ and $a \geq c$. Thus, $f = a \in 2A$.

We now prove the estimate (19). From (18) and (5) we get

$$|J| = c + a_{k-1} - 1 - e = k - 2 + b' + c - e. \tag{20}$$

We now estimate $c - e$ from below. For this we will estimate the number P of holes in A which are less than e or larger than c. For the number P_1 of holes which are less than e we have, according to Lemma 4,

$$P_1 \geq (e+1)/2, \tag{21}$$

and for the number P_2 of holes which are greater than c we have, according to Lemma 5,

$$P_2 \geq (a_{k-1} - c + 1)/2. \tag{22}$$

The sets P_1 and P_2 have an empty intersection, in view of Lemma 6, and thus, in view of (21) and (22),

$$P \geq P_1 + P_2 \geq (e+1)/2 + (k - 1 + b' - c + 1)/2. \tag{23}$$

We will get an estimate of P from above by taking the number of all pairs b' minus the number d of those a which are holes in A in the interval (e, c).

Thus,

$$b' - d \geq (e + k + b' - c + 1)/2$$

and

$$c - e \geq k + b' + 1 - 2(b' - d) = k + 2d + 1 - b'. \tag{24}$$

Because of (24) we get from (20)

$$|J| \geq k - 2 + b' + k + 2d + 1 - b' = 2k + 2d - 1. \ \blacksquare$$

Proof of Corollary 1. We have $b' = b$. Thus the set of unstable pairs is empty, every point of the interval $[e + 1, c + a_{k-1} - 1]$ belongs to $2A$.

The elements of $[0, e + 1]$ which are holes in A may belong only to a left stable pair, and so the set

$$A_1 = [0, e + 1] \cap A$$

is stable. A similar reasoning may be applied to A_2. $\blacksquare$

EXAMPLE.

$$A = \{0, 2, 4, 6, 7, 8, 9, 10, 14\}.$$

We have here $e = 5$, $c = 11$, the set $A_1 = \{0, 2, 4, 6\}$ is stable, the set $A_2 = \{0, 4\}$ is stable, $J = [6, 24]$ and $I = \{6, 7, 8, 9, 10\}$.

References

[1] G. A. Freiman, *The addition of finite sets I*, Izv. Vyssh. Uchebn. Zaved. Mat. 6 (1959), no. 13, 202–213 (in Russian).

[2] —, *Structure theory of set addition*, Astérisque 258 (1999), 1–30.

[3] M. B. Nathanson, *Additive Number Theory. Inverse Problems and the Geometry of Sumsets*, Grad. Texts in Math. 165, Springer, New York, 1996.

School of Mathematical Sciences
Tel Aviv University
Tel Aviv 69978, Israel
E-mail: grisha@post.tau.ac.il

Received on 3.2.2008
and in revised form on 27.1.2009 (5630)