#A28 INTEGERS 10 (2010), 335-350

LONG ARITHMETIC PROGRESSIONS IN SMALL SUMSETS

Itziar Bardaji
Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, c/ Llorens i Artigas 4-6. 08028 Barcelona, Espanya
ibardaji@iri.upc.edu

David J. Grynkiewicz[1]
Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria
diambri@hotmail.com

Received: 4/22/09, Revised: 2/22/10, Accepted: 3/16/10, Published: 7/19/10

Abstract

Let A, $B \subseteq \mathbb{Z}$ be finite, nonempty subsets such that $\max B - \min B \leq \max A - \min A$, $\gcd(A+B-c) = 1$, for some $c \in A+B$, and $|A+B| \leq |A|+2|B|-3-\delta(A,B)$, where $\delta(A,B)$ is 1 if $x + A \subseteq B$ for some $x \in \mathbb{Z}$, and is 0 otherwise. Assume one of the following conditions holds true:

- $\max A - \min A \leq |A| + |B| - 3$,
- $\gcd(A - a) \leq 2$, for some $a \in A$,
- $|A + B| \leq 2|A| + |B| - 3 - \delta(B, A)$.

Then $A+B$ contains a $(|A|+|B|-1)$–term arithmetic progression with difference 1.

1. Introduction

For a subset $A \subseteq \mathbb{Z}$, we let $\mathsf{diam}\, A = \max A - \min A$ denote its diameter and $|A|$ its cardinality. We let $\gcd^* A = \gcd(A - a_0)$, where $a_0 \in A$ and gcd denotes the greatest common divisor. Note that the definition of $\gcd^* A$ does not depend on the choice of a_0 and, by convention, $\gcd^*(A) = \infty$ when $|A| = 1$. For A, $B \subseteq \mathbb{Z}$, their sumset is the set of all sums of one element from A and one element from B:

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

Also, define

$$\delta(A,B) = \begin{cases} 1 & \text{if } x + A \subseteq B \text{ for some } x \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases}$$

The study of the structure of subsets with small sumset has a rich tradition (see [10] and [13] for two texts on the subject). A classical result of Freiman [4] [2] [10] [13] states that if a set A of integers satisfies $\gcd^* A = 1$ and

[1]Supported by FWF project number M1014-N13

$$|A + A| \leq 3|A| - 4, \tag{1}$$

then the diameter of A is at most $2|A| - 4$. In other words, A is an interval with at most $|A| - 3$ holes. Various versions of a generalization to distinct summands were later found [5] [7] [12] [6]. Theorem A below is a recent version [6] that combines all previous generalizations into a single (slightly) more general statement.

Theorem A. *Let* $A, B \subseteq \mathbb{Z}$ *be finite and nonempty subsets with* $\gcd^*(A + B) = 1$ *and* $\mathsf{diam}\, B \leq \mathsf{diam}\, A$. *Let* $|A + B| = |A| + |B| - 1 + r$. *Suppose either*

(i) $|A + B| \leq |A| + |B| - 3 + \min\{|B| - \delta(A, B), |A| - \delta(B, A)\}$ *or*

(ii) $|A + B| \leq |A| + 2|B| - 3 - \delta(A, B)$ *and* $\gcd^* A = 1$.

Then $\mathsf{diam}\, A \leq |A| + r - 1$ *and* $\mathsf{diam}\, B \leq \min\{|A|, |B|\} + r - 1$.

Strangely enough, Theorem A remains true if the condition $\gcd^* A = 1$ in (ii) is relaxed to $\gcd^* A \leq 2$, and this can be shown by a short argument using Theorem A as stated above. As this seems not to have been noticed before, we provide the details of the following strengthening of Theorem A in the next section.

Theorem B. *Let* $A, B \subseteq \mathbb{Z}$ *be finite subsets with* $\gcd^*(A + B) = 1$ *and* $\mathsf{diam}\, B \leq \mathsf{diam}\, A$. *Suppose either*

(i) $|A + B| \leq |A| + |B| - 3 + \min\{|B| - \delta(A, B), |A| - \delta(B, A)\}$ *or*

(ii) $|A + B| \leq |A| + 2|B| - 3 - \delta(A, B)$ *and* $\gcd^* A \leq 2$.

Then $\mathsf{diam}\, A \leq |A| + r - 1$ *and* $\mathsf{diam}\, B \leq \min\{|A|, |B|\} + r - 1$.

As later became apparent, knowing that there are only a small number of holes in a pair of sets with small sumset is not always sufficient. In part, this is because there are many subsets of small diameter that nonetheless have large sumset. Working through examples, one quickly finds that, informally speaking, it is much more difficult for the holes in a subset A with small sumset (and correspondingly the holes in $A+A$ as well) to occur in the interior of the set than near the boundary (namely, near the maximum or minimum element). However, there have been few results satisfyingly embodying this idea.

One such result occurred in a paper of Lev where long arithmetic progressions were found in hA (where $hA = A + (h - 1)A$ denotes the h–fold sumset) [8]. In particular, it was proved that if $\mathsf{diam}\, A < \frac{3}{2}|A| - 1$ and $\gcd^* A = 1$, then $A + A$ contains at least $2|A| - 1$ consecutive integers [8, Corollary 1]. Another instance occurs in a second paper of Lev characterizing large sum-free sets over $\mathbb{Z}/p\mathbb{Z}$ [9]. Namely, among other similar results from these papers, it was shown that if $\mathsf{diam}\, A < \frac{3}{2}|A| - 1$ and $\gcd^* A = 1$, then $A - A$ contains an interval of $2|A| - 1$ integers [9, Lemma 3]; see also [1] for an application of this result.

Very recently, G. Freiman showed in [3] that there is always a $(2|A|-1)$–term arithmetic progression in $A+A$ when the sumset is so small as to satisfy (1). Now $|2A| \leq 2 \cdot \mathsf{diam}\, A + 1$ holds trivially with equality only possible when $2A$ is itself an arithmetic progression with difference 1. Thus, when searching for long arithmetic progressions in $2A$, one can assume $|2A| \leq 2 \cdot \mathsf{diam}\, A$, and now the assumption $\mathsf{diam}\, A < \frac{3}{2}|A| - 1$ from Lev's paper can be used to show the assumption (1) holds in Freiman's paper except when $\mathsf{diam}\, A = \frac{3}{2}(|A|-1)$ precisely (in which case $|A|$ must be odd). Hence Freiman's result implies that of Lev except in this one particular case. However, this case can be handled by a simple separate argument using an analog of Proposition 4 proved in the paper of Freiman.

The example

$$A = \{0, 1, 2, \ldots, k-r-1, k-r+1, k-r+3, \ldots, k-r+(2r-1)\},$$

for $r = 0, 1, \ldots, k-3$, shows the bound on the arithmetic progression length to be best possible, while the example

$$A = \left\{1, 2, \ldots, \left\lceil \frac{k}{2} \right\rceil \right\} \cup \left\{x+1, x+2, \ldots, x + \left\lfloor \frac{k}{2} \right\rfloor \right\},$$

for $x \geq k+1$, shows that the assumption $|A+A| \leq 3|A| - 4$ from (1) is needed. The paper of Freiman also delved into the issue of where the holes could occur in A, but the other structural information is derivable from the bound on the length of the arithmetic progression.

The goal of our paper is to extend the aforementioned result of Freiman to pairs of distinct summands A and B.

Theorem 1. *Let A, $B \subseteq \mathbb{Z}$ be finite and nonempty with $\mathsf{diam}\, B \leq \mathsf{diam}\, A \leq |A| + |B| - 3$ and*

$$|A+B| \leq |A| + 2|B| - 3 - \delta(A, B). \tag{2}$$

Then $A+B$ contains $|A|+|B|-1$ consecutive integers.

Combining Theorem 1 with Theorem B will give the following corollary.

Corollary 2. *Let A, $B \subseteq \mathbb{Z}$ be finite and nonempty. Suppose either*

(i) $|A+B| \leq |A| + |B| - 3 + \min\{|B| - \delta(A, B),\, |A| - \delta(B, A)\}$ *or*

(ii) $\mathsf{diam}\, B \leq \mathsf{diam}\, A$, $\gcd^* A \leq 2$ *and* $|A+B| \leq |A| + 2|B| - 3 - \delta(A, B)$.

Then $A+B$ contains a $(|A|+|B|-1)$-term arithmetic progression of difference $\gcd^*(A+B)$.

The rest of the paper is organized as follows. In Section 2, we present the notation that will be assumed for the remainder of the paper. In Section 3, we give the derivations of Theorem B and Corollary 2. Section 4 is devoted to the proof of Theorem 1. The structural consequences concerning the location of holes and such will become apparent in the series of propositions and definitions leading up to the proof of Theorem 1. The paper concludes with a few additional remarks.

2. Notation

Throughout this paper, we assume A, $B \subseteq \mathbb{Z}$ are finite, nonempty subsets normalized so that

$$\min A = \min B = 0, \tag{3}$$

and with

$$M = \max A \qquad \text{and} \qquad N = \max B, \tag{4}$$

$$M \geq N, \tag{5}$$

$$|A+B| = |A|+|B|-1+r, \tag{6}$$

so that A is assumed to be the set with larger (or equal) diameter. As all questions are translation invariant, there is no loss of generality when assuming (3). Note, in view of (3) and (5), that

$$\delta(A,B) = 1 \text{ if and only if } A \subseteq B. \tag{7}$$

For a, $b \in \mathbb{Z}$, we define $[a,b] := \{x \in \mathbb{Z} \mid a \leq x \leq b\} \subseteq \mathbb{Z}$. Note $[a,b] = \emptyset$ when $b < a$. For a set X and an interval $[a,b] \subseteq \mathbb{Z}$, the number of holes of X in $[a,b]$ is denoted by

$$h_X(a,b) = |[a,b] \setminus X|.$$

When $[a,b]$ is the default interval $[\min X, \max X]$, we skip reference to the interval, that is,

$$h_X = h_X(\min X, \max X),$$

and when we refer to a hole in X without reference to an interval, we simply mean an element $x \in [\min X, \max X] \setminus X$.

Observe, in view of (3), (4) and (6), that

$$M = |A| + h_A - 1, \tag{8}$$

$$N = |B| + h_B - 1, \tag{9}$$

$$h_{A+B} = M + N + 1 - |A+B| = h_A + h_B - r. \tag{10}$$

Also remark that, using (8), we can rewrite the condition $\mathsf{diam}\, A \leq |A| + |B| - 3$ in Theorem 1 as $h_A \leq |B| - 2$ and the condition (2) as $r \leq |B| - 2 - \delta(A, B)$.

3. Proofs for Theorem B and Corollary 2

First we give the derivation of Theorem B from Theorem A.

Proof. In view of Theorem A and our hypotheses, we assume $\gcd^* A = 2 > \gcd^*(A+B) = 1$ and

$$|A + B| \leq |A| + 2|B| - 3 - \delta(A, B). \tag{11}$$

We may also use the notation and assumptions presented in Section 2. In particular, $\min A = \min B = 0$, $\max A = \mathsf{diam}\, A = M$ and $\max B = \mathsf{diam}\, B = N$. Let $B_i = B \cap (i + 2\mathbb{Z})$, for $i = 0, 1$. Note both B_0 and B_1 are nonempty, else $\gcd^*(A+B) = 2$, contrary to hypothesis. Since $\gcd^*(A) = 2$ and $\min A = 0$, we know M is even. Hence $\mathsf{diam}\, B \leq \mathsf{diam}\, A$ implies $\mathsf{diam}\, B_1 < \mathsf{diam}\, A$ and $\mathsf{diam}\, B_0 \leq \mathsf{diam}\, A$; moreover, $\delta(A, B_1) = 0$, and $\delta(A, B) = \delta(A, B_0)$.

Note that the hypothesis $\gcd^*(A + B) = 1$ in Theorems A and B is simply a normalization hypothesis; if $\gcd^*(A+B) = d \geq 2$, then A and B are both contained in arithmetic progressions with difference d, and Theorems A and B can be applied by considering A and B (appropriately translated) as subsets of $d\mathbb{Z} \cong \mathbb{Z}$. Thus, we can apply case (ii) of Theorem A to both the pairs (A, B_0) and (A, B_1) considered (appropriately translated) as subsets of $2\mathbb{Z} \cong \mathbb{Z}$ to find the following bounds (note the conclusion of Theorem A holding implies the sumset satisfies the cardinality bound corresponding to the first term in each of the minimums below, while the second bound in each of the minimums corresponds to the case when (ii) fails to hold in Theorem A):

$$|A + B_0| \geq \min\{|A| + |B_0| - 1 + h,\ |A| + 2|B_0| - 2 - \delta(A, B)\}, \tag{12}$$

$$|A + B_1| \geq \min\{|A| + |B_1| - 1 + h,\ |A| + 2|B_1| - 2\}, \tag{13}$$

where $h := |\{0, 2, \ldots, M\} \setminus A|$. Note $h_A = \frac{1}{2}M + h$ and $|A| = \frac{1}{2}M + 1 - h$, where h_A is as defined in Section 2.

First let us show that $\mathsf{diam}\, A \leq |A| + r - 1$. Assuming by contradiction this is false, then $M = \mathsf{diam}\, A \geq |A| + r$. We proceed in four cases depending on which pair of bounds from (12) and (13) holds.

If

$$|A+B| = |A+B_0| + |A+B_1| \geq 2|A| + |B_0| + |B_1| - 2 + 2h,$$

then combining with $|A+B| = |A|+|B|-1+r \leq M+|B|-1$ (in view of $M \geq |A|+r$) yields $M \geq 2|A| - 1 + 2h = M + 1$, where we use $|A| = \frac{1}{2}M + 1 - h$ for the equality, which is a contradiction. If

$$|A+B| = |A+B_0| + |A+B_1| \geq 2|A| + 2|B_0| + 2|B_1| - 4 - \delta(A,B),$$

then combining with (11) yields $|A| \leq 1$, contradicting that $\gcd^*(A) = 2 \neq \infty$. If

$$|A+B| = |A+B_0| + |A+B_1| \geq 2|A| + 2|B_0| + |B_1| - 3 - \delta(A,B) + h,$$

then combining with (11) yields $|B_1| \geq |A| + h = \frac{1}{2}M + 1$; however, since $B_1 \subseteq [1, M-1] \cap (1 + 2\mathbb{Z})$, this is impossible. Finally, if

$$|A+B| = |A+B_0| + |A+B_1| \geq 2|A| + |B_0| + 2|B_1| - 3 + h,$$

then combining with (11) yields $|B_0| \geq |A| + \delta(A,B) + h = \frac{1}{2}M + 1 + \delta(A,B)$; however, since $B_0 \subseteq [0, M] \cap 2\mathbb{Z}$, we have $|B_0| \leq \frac{1}{2}M + 1$ with equality possible only if $B_0 = [0, M] \cap 2\mathbb{Z}$, in which case $A \subseteq [0, M] \cap 2\mathbb{Z} = B_0$ and $\delta(A,B) = 1$, whence $|B_0| \geq \frac{1}{2}M + 1 + \delta(A,B)$ cannot hold, and is thus a contradiction. Therefore, we obtain a contradiction in all four cases and instead conclude that $\mathsf{diam}\, A \leq |A| + r - 1$. Since $\mathsf{diam}\, B \leq \mathsf{diam}\, A$, this also implies $\mathsf{diam}\, B \leq |A| + r - 1$.

It remains to show $\mathsf{diam}\, B \leq |B| + r - 1$, for which we may assume $|B| < |A|$, else this follows from $\mathsf{diam}\, B \leq |A| + r - 1$. But now (11) implies that $|A+B| \leq 2|A| + |B| - 4$. Consequently, the hypothesis (i) in Theorem A holds, and the result follows by applying Theorem A(i), completing the proof. □

Next, we give the proof of Corollary 2.

Proof of Corollary 2. As hypothesis (i) in Corollary 2 is symmetric with respect to A and B, we may, without loss of generality, assume $\mathsf{diam}\, B \leq \mathsf{diam}\, A$. Note both hypotheses (i) and (ii) imply

$$|A+B| = |A| + |B| - 1 + r \leq |A| + 2|B| - 3 - \delta(A,B),$$

whence $r \le |B| - 2 - \delta(A,B)$. Thus, if $\gcd^*(A+B) = 1$, then Theorem B implies that $\mathsf{diam}\, B \le \mathsf{diam}\, A \le |A| + r - 1 \le |A| + |B| - 3 - \delta(A,B) \le |A| + |B| - 3$, and now Theorem 1 completes the proof.

On the other hand, if $\gcd^*(A+B) = d \ge 2$, then we can, without loss of generality, translate A and B so that $\min A = \min B = 0$ and apply the just-proved case $\gcd^*(A+B) = 1$ in Corollary 2 to the sets $A, B \subseteq d\mathbb{Z} \cong \mathbb{Z}$ to complete the proof. □

Note that taking $A = \{id \mid i = 0, 1, \ldots, r-1\}$ to be an arithmetic progression with difference $d \ge 3$ and length $r \ge 3$ and taking $B = \{id \mid i = 0, \ldots, r-2\} + \{0, 1\}$ shows that the condition $\gcd^* A \le 2$ is needed in Corollary 2, and thus also in Theorem B. Finally, it is worth noting that

$$|A+B| \le |A| + |B| - 3 + \min\{|B| - \delta(A,B),\, |A| - \delta(B,A)\}$$

holds if and only if

$$|A+B| \le |A| + |B| - 3 + \min\{|B| - \delta(A,B),\, |A|\}.$$

Indeed, if this were not the case, then $|A+B| = 2|A| + |B| - 3$ and $\delta(B,A) = 1$; as the latter implies $|B| \le |A|$, it subsequently follows, from $2|A| + |B| - 3 = |A+B| \le |A| + 2|B| - 3 - \delta(A,B)$, that $|B| = |A|$ and $\delta(A,B) = 0$, which is impossible in view of $\delta(B,A) = 1$. The same argument, with the roles of A and B reversed, shows that both these bounds are also equivalent to

$$|A+B| \le |A| + |B| - 3 + \min\{|B|,\, |A| - \delta(B,A)\}.$$

4. Proof of Theorem 1

In this section, the proof of the main theorem is provided. We assume throughout this section all the notation and assumptions of Section 2. We begin by showing that, when A has few holes, the interval $[N, M]$ is always contained in the sumset of A and B.

Proposition 3. *If $h_A \le |B| - 1$, then*

$$[N, M] \subseteq A + B. \tag{14}$$

Proof. Let $x \in [N, M]$. Thus,

$$(x, 0),\ (x-1, 1),\ \ldots,\ (x-N, N)$$

are representations (a,b) of $x=a+b$ with $a\in[0,M]$ and $b\in[0,N]$. If $x\notin A+B$, then each of these $N+1$ pairs must either have the first element missing from A or the second element missing from B, whence $h_A+h_B\geq N+1=|B|+h_B$ (in view of (9)). But this contradicts $h_A\leq|B|-1$. □

We call a hole x in A *left-stable* (*right-stable*) if x (respectively, $x+N$) is a hole in $A+B$. Similarly, a hole x in B will be called *left-stable* (*right-stable*) if x (respectively, $x+M$) is a hole in $A+B$.

In view of Proposition 3, if x is a right-stable hole in A, then $x+N$ lies to the right of the interval $[N,M]$, and if x is a right-stable hole in B, then $x+M$ also lies to the right of this interval. These holes in $A+B$ are called *right* holes. Also, a left-stable hole in either A or B lies to the left of the interval $[N,M]$. These are *left* holes in $A+B$. Indeed, for a given integer, being a left-stable hole in A, being a left-stable hole in B, and being a left hole in $A+B$ are all equivalent.

A *stable* hole in A is one which is either right or left-stable, and likewise for B. All other holes (in either A or B) are called *unstable*. We let h_A^s and h_B^s denote the respective number of stable holes in A and B, and we let h_A^u and h_B^u denote the respective number of unstable holes in A and B.

This classification of holes into ones which contribute to a hole present in $A+B$ (the stable ones) and those which do not contribute to any hole in $A+B$ (the unstable ones) will prove to be a very useful perspective.

To every hole x in $A+B$, we associate two stable holes x_A and x_B in A and B, respectively, as follows:

- If $x<N$, we let $x_A:=x$ and $x_B:=x$; thus, both x_A and x_B are left-stable.
- If $x>M$, we let $x_A:=x-N$ and $x_B:=x-M$, so that both x_A and x_B are right-stable.

We will later see that these mappings are injective, i.e., that $x_A=y_A$ for holes x and y in $A+B$ implies $x=y$, and likewise $x_B=y_B$ implies $x=y$. However, next we prove a very important proposition—the key observation used in the proof—which shows that if we have a left hole $x\notin A+B$, then there must be many holes in $A\cap[0,x]$ and $B\cap[0,x]$, with an analogous statement holding for right holes.

Proposition 4. *If $x\in[0,N]\setminus(A+B)$, then*

$$h_A(0,x)+h_B(0,x)\geq x+1. \tag{15}$$

If $x\in[M,M+N]\setminus(A+B)$, then

$$h_A(x-N,M)+h_B(x-M,N)\geq N-(x-M)+1=M+N-x+1. \tag{16}$$

Proof. The proof is analogous to that of the previous proposition. If $x \in [0, N]$, then

$$(x, 0),\ (x-1, 1),\ \ldots, (0, x)$$

are representations (a, b) of $x = a + b$ with $a \in [0, M]$ and $b \in [0, N]$ (in view of (5)). If $x \notin A + B$, then each of these $x+1$ pairs must either have the first element missing from A or the second element missing from B, whence (15) follows. The argument for when $x \in [M, M+N]$ is analogous, considering instead

$$(M, x-M),\ (M-1, x-M+1),\ \ldots, (x-N, N).$$

□

Next, we show that no hole in A can be both left and right-stable.

Proposition 5. *Let $x \in [0, M] \setminus A$. If $h_A \leq |B| - 2$, then either $x \in A + B$ or $x + N \in A + B$.*

Proof. If both $x \notin A + B$ and $x + N \notin A + B$, then Proposition 3 implies $x \in [M - N + 1, N - 1]$, whence applying both cases of Proposition 4 yields

$$\begin{aligned} M + 2 &= (x+1) + (M - x + 1) \\ &\leq h_A(0, x) + h_B(0, x) + h_A(x, M) + h_B(x + N - M, N) \\ &\leq h_A + 1 + h_B + M - N + 1, \end{aligned} \tag{17}$$

where the second inequality follows in view of (5). Now applying (9) yields $h_A \geq |B| - 1$, contrary to assumption. □

The following shows there are also no holes in B which are both left-stable and right-stable.

Proposition 6. *Let $x \in [1, N] \setminus B$. If $h_A \leq |B| - 2$, then either $x \in A + B$ or $x + M \in A + B$.*

Proof. If both $x \notin A+B$ and $x+M \notin A+B$, then applying both cases of Proposition 4 yields

$$\begin{aligned} N + 2 &= (x+1) + (N - x + 1) \\ &\leq h_A(0, x) + h_B(0, x) + h_A(x + M - N, M) + h_B(x, N) \\ &\leq h_A + h_B + 2, \end{aligned} \tag{18}$$

where the second inequality follows by (5). Now applying (9) yields $h_A \geq |B| - 1$, contrary to assumption. □

From Proposition 5, it is easy to conclude that, when $h_A \le |B|-2$, the mapping $x \mapsto x_A$ is injective, as previously alluded. Indeed, if $x_A = y_A$ for holes x and y in $A+B$, then either $x = y \pm N$ or $x = y$; but if (without loss of generality) $x = y+N$, then $y \notin A+B$ and $y+N \notin A+B$, in contradiction to Proposition 5.

With a similar reasoning for the second mapping, it follows that $x \mapsto x_A$ is a bijection between the set of all holes in $A+B$ and the set of all stable holes in A, and $x \mapsto x_B$ is a bijection between the set of all holes in $A+B$ and the set of all stable holes in B. In consequence, we have that, when $h_A \le |B|-2$,

$$h_A^s = h_B^s = h_{A+B} = h_A + h_B - r, \tag{19}$$

where r is as defined in (10). Since $h_B = h_B^u + h_B^s$ and $h_A = h_A^u + h_A^s$, we also have

$$h_A^u = r - h_B \tag{20}$$

$$h_B^u = r - h_A. \tag{21}$$

The next proposition is the trickiest part of the proof, showing that all left-stable holes precede all right-stable holes, so there is no overlap. Recall that using (6) and (8), the conditions $h_A \le |B|-2$ and $r \le |B|-2-\delta(A,B)$ correspond to conditions $\mathsf{diam}\, A \le |A|+|B|-3$ and (2) in Theorem 1.

Proposition 7. *Suppose $h_A \le |B|-2$ and $r \le |B|-2-\delta(A,B)$. If $x_B \in [0,N]\setminus B$ is a left-stable hole and $y_B \in [0,N]\setminus B$ is a right-stable hole, then $x_B < y_B$. Likewise, if $x_A \in [0,M]\setminus A$ is a left-stable hole and $y_A \in [0,M]\setminus A$ is a right-stable hole, then $x_A < y_A$.*

Proof. If $x_A \in [0,M]\setminus A$ is a left-stable hole, $y_A \in [0,M]\setminus A$ is a right-stable hole and $x_A \ge y_A$, then $x_B = x_A \in [0,N]\setminus B$ is a left-stable hole and $y_B = y_A - (M-N) \in [0,N]\setminus B$ is a right-stable hole with $x_B \ge y_B$, in view of $x_A \ge y_A$ and (5). Therefore we see that it suffices to prove the first assertion in the proposition, as the second is an immediate consequence.

To that end, assume $x_B \in [0,N]\setminus B$ is a left-stable hole and $y_B \in [0,N]\setminus B$ is a right-stable hole with $x_B > y_B$. Note that $x_B = y_B$ cannot hold in view of Proposition 6. Moreover, assume x_B and y_B are chosen minimally, meaning that there are no stable holes $z \in [y_B+1, x_B-1]\setminus B$.

Applying both cases of Proposition 4 to x_B and $y_B + M$, respectively, we find that

$$\begin{aligned}|B| + h_B + (x_B - y_B + 1) &= (x_B + 1) + (N - y_B + 1)\\ &\le h_A(0, x_B) + h_B(0, x_B) + h_A(y_B + M - N, M)\\ &\qquad + h_B(y_B, N)\\ &\le h_A + h_B + h_A(y_B + M - N, x_B)\\ &\qquad + h_B(y_B, x_B),\end{aligned} \tag{22}$$

where we use (9) for the first equality. Note that if $y_B + M - N > x_B$, then, by definition, $h_A(y_B + M - N, x_B) = 0$. In this case, inequality (22) also holds true.

In view of the minimality of x_B and y_B, we see that

$$h_B(y_B, x_B) \le h_B^u + 2, \tag{23}$$

with equality possible only if $[y_B + 1, x_B - 1]$ contains all the unstable holes in B. We also have the trivial inequality

$$h_B(y_B, x_B) \le x_B - y_B + 1. \tag{24}$$

If $y_B + M - N > x_B$, so that $h_A(y_B + M - N, x_B) = 0$, then (22) and (24) imply $h_A \ge |B|$, contrary to hypothesis. Therefore we may assume $y_B + M - N \le x_B$, and now we also have the trivial inequality

$$h_A(y_B + M - N, x_B) \le x_B - y_B + 1 - (M - N), \tag{25}$$

with equality possible only if all the integers in $[y_B + M - N, x_B]$ are holes in A.

Applying the estimates (25) and (23) in (22) and using (8), (9) and (21), we discover that

$$|A| - 2 - r \le h_B - h_A. \tag{26}$$

In view of (5), (8) and (9), we have

$$h_B - h_A \le |A| - |B|, \tag{27}$$

with equality only possible when $M = N$. Combining (27) and (26) yields

$$r \ge |B| - 2, \tag{28}$$

whence our hypothesis $r \le |B| - 2 - \delta(A, B)$ implies that $r = |B| - 2$, that $\delta(A, B) = 0$, and that equality held in all estimates used to derive (28).

As a result, $\delta(A,B)=0$ and (7) imply $A \nsubseteq B$; equality in (27) implies $M=N$; and equality in (26) implies equality holds in both (25) and (23), whence all the integers belonging to the interval $[y_B+M-N, x_B]$ are holes in A and $[y_B+1, x_B-1]$ contains all the unstable holes in B.

Since $A \nsubseteq B$, it follows that there exists $z \in A$ with $z \notin B$. Since all the elements in $[y_B+M-N, x_B]$ are holes in A and $M=N$, it follows that $z \notin [y_B, x_B]$. Thus, since $[y_B+1, x_B-1]$ contains all the unstable holes in B, it follows that $z \notin B$ is a stable hole in B. However, from the definition of stability, this means that either $z \notin A+B$ or $z+M \notin A+B$, which are both contradictions in view of $z \in A$, $0 \in B$ and $M=N \in B$, completing the proof. □

We are now ready to finish the proof of Theorem 1, which will follow from the next proposition.

Proposition 8. *Suppose $h_A \le |B|-2$ and $r \le |B|-2-\delta(A,B)$. Then*

$$J := [e+1, M+c-1] \subseteq A+B,$$

where e is the greatest left stable hole in B (let $e=-1$ if there are no left stable holes) and c is the smallest right stable hole in B (let $c=N+1$ if there are no right stable holes). Moreover,

$$\begin{aligned} |J| &= M-1+(c-e) \\ &\ge |A|+|B|-1+h_A(e+1, c+M-N-1)+h_B(e+1, c-1) \\ &\ge |A|+|B|-1. \end{aligned} \tag{29}$$

Proof. In view of Proposition 7, we have $e<c$. By the definition of stability (and that of e and c), there are no left holes in $A+B$ greater than e and no right holes in $A+B$ less than than $M+c$. As every hole in $A+B$ is either a right or left hole (in view of Proposition 3), this means $J := [e+1, M+c-1] \subseteq A+B$. Note that

$$|J| = M-1+(c-e) = |A|+h_A-2+(c-e), \tag{30}$$

using (8). It remains to estimate $c-e$.

Let $s = h_A(e+1, c+M-N-1)+h_B(e+1, c-1)$. Applying both cases of Proposition 4 to e and $c+M$, respectively, we find that

$$\begin{aligned} (e+1)+(N-c+1) &\le h_A(0,e)+h_B(0,e)+h_A(c+M-N, M)+h_B(c,N) \\ &\le h_A+h_B-s. \end{aligned} \tag{31}$$

Note that in (31) we used $e<c$, which implies $e<c+M-N$ in view of (5).

From (9) and (31), it follows that

$$|B| + h_B + 1 + e - c \leq h_A + h_B - s,$$

yielding

$$c - e \geq |B| + 1 - h_A + s.$$

Combining the above estimate for $c - e$ with (30), we obtain

$$\begin{aligned} |J| = |A| + h_A - 2 + (c - e) &\geq |A| + h_A - 2 + (|B| + 1 - h_A + s) \\ &= |A| + |B| - 1 + s \geq |A| + |B| - 1, \end{aligned}$$

completing the proof. □

Finally, we complete the proof of Theorem 1.

Proof of Theorem 1. We may, without loss of generality, assume $\min A = \min B = 0$. Since $\mathsf{diam}\, B \leq \mathsf{diam}\, A \leq |A| + |B| - 3$, we have $h_A \leq |B| - 2$ in view of (8). Since $|A+B| := |A| + |B| - 1 + r \leq |A| + 2|B| - 3 - \delta(A, B)$, we have $r \leq |B| - 2 - \delta(A, B)$. Thus applying Proposition 8 completes the proof. □

5. Concluding Remarks

We conclude with some brief remarks, for which we assume the notation of the previous section, particularly concerning Proposition 8.

First, let us show that all the intermediary work and propositions leading up to Theorem 1, save Proposition 4, are easily deduced from Theorem 1 itself. Let $J = [m, n] \subseteq A + B$ be a maximal length arithmetic progression with difference 1, so $|J| \geq |A| + |B| - 1$ by Theorem 1 and, consequently, $n - M > m$ (in view of $h_A \leq |B| - 2$). Since

$$\begin{aligned} [m, n - N] + (B \cup [m, n - M]) &\subseteq [m, n - N] + [0, N] \subseteq [m, n] = J \subseteq A + B \quad \text{and} \\ (A \cup [m, n - N]) + [m, n - M] &\subseteq [0, M] + [m, n - M] \subseteq [m, n] = J \subseteq A + B, \end{aligned}$$

we observe that

$$(A \cup [m, n - N]) + (B \cup [m, n - M]) = A + B.$$

Thus, Theorem 1 can be applied using $(A \cup [m, n - N])$ and $(B \cup [m, n - M])$ to conclude that the bound $|J| \geq |A| + |B| - 1$ can be improved by one for each element of $[m, n - N] \setminus A$ and each element of $[m, n - M] \setminus B$.

Since $|J| \geq |A| + |B| - 1$ and

$$\mathsf{diam}\, B = N \leq M = \mathsf{diam}\, A \leq |A| + |B| - 3, \tag{32}$$

it follows that $[N-1, M+1] \subseteq J \subseteq A+B$, which yields Proposition 3. If x and $x+N$ are both holes in $A+B$, where $x \in [0, M]$, then J must lie entirely in one of the intervals $[0, x-1]$, $[x+1, x+N-1]$ or $[x+N+1, N+M]$, all of which contain less than $|A|+|B|-1$ elements in view of (32), contradicting $|J| \geq |A|+|B|-1$. This establishes Proposition 5. Likewise, if x and $x+M$ are both holes in $A+B$, where $x \in [0, N]$, then J must lie entirely in one of the intervals $[0, x-1]$, $[x+1, x+M-1]$ or $[x+M+1, N+M]$, all of which contain less than $|A|+|B|-1$ elements in view of (32), again contradicting $|J| \geq |A|+|B|-1$. This establishes Proposition 6. Now note, since e, $c+M \notin A+B$, that J must live entirely in one of the intervals $[0, e-1]$, $[c+M+1, M+N]$ or $[e+1, c+M-1]$. However, the first two intervals contain less than $|A|+|B|-1$ elements in view of $e \leq N \leq M \leq |A|+|B|-3$. Thus, since $|J| \geq |A|+|B|-1$, we conclude that $J \subseteq [e+1, c+M-1]$. In particular, we see that $e < c$ (as otherwise $|J| \leq M-1 \leq |A|+|B|-4$, a contradiction), which implies Proposition 7. Since $e < c$, the definition of e and c ensure that $[e+1, c+M-1] \subseteq A+B$. Hence, since $J \subseteq [e+1, c+M-1]$, the maximality of J implies that $J = [e+1, c+M-1]$. Thus $m = e+1$ and $n = c+M-1$, and now the improved bound on $|J|$ from the previous paragraph implies the first (and seemingly stronger) inequality from (29), namely,

$$|J| \geq |A|+|B|-1+h_A(e+1, c+M-N-1)+h_B(e+1, c-1),$$

and Proposition 8 follows.

Next, it is important to note that Theorem 1 and Proposition 8 essentially show that the sets A and B can be divided into left and right halves with each half behaving independently (with respect to the sumset $A+B$) of the other. For instance, taking the left halves $A_L = A \cap [0, e]$ and $B_L = B \cap [0, e]$ and appending on a sufficiently long interval gives a pair of subsets whose sumset has the exact same set of left holes as for the original sumset $A+B$, that is,

$$(A_L + B_L) \cap [0, e] = C_L := (A+B) \cap [0, e] \quad \text{and}$$
$$(A_L \cup [e+1, x]) + (B_L \cup [e+1, x]) = C_L \cup [e+1, 2x]$$

for sufficiently large $x \geq e+1+\min\{g_A(0,e), g_B(0,e)\}$, where $g_A(0,e)$ denotes the maximal size of a *gap* in $A_L \cup \{e+1\}$—that is, the maximal number of terms in an arithmetic progression with difference 1 contained in $[0, e] \setminus A$—and $g_B(0,e)$ is similarly defined. Note $g_A(0,e) \leq h_A(0,e) \leq h_A \leq r$ and $g_B(0,e) \leq$

$h_B(0,e) \le h_B \le r$. The right holes of $A+B$ can be independently studied in a similar manner via the right halves $A_R = A \cap [c+M-N, M]$, $B_R = B \cap [c, N]$ and $C_R = (A+B) \cap [M+c, M+N]$. Indeed, as seen from the previous two paragraphs,

$$(A_L \cup I_A \cup A_R) + (B_L \cup I_B \cup B_R) = A + B = C_L \cup J \cup C_R,$$

where $I_A = [e+1, c-1+M-N]$ and $I_B = [e+1, c-1]$.

In general, there are many possibilities for how the holes can be distributed in A_L and B_L. However, if one wishes to use holes efficiently, that is, use a large number of holes relative to the maximal bound r, then (20) and (21) show that the number of unstable holes must be small, which helps restrict the possibilities for A_L and B_L.

For instance, in the extremal case when there are no unstable holes in either A or B, then we must have $A_L = B_L$, and $A_L \cup [e+1, \infty)$ is the complement of the solution set of the Frobenius problem (see [11]) for the set A, i.e., $A_L \cup [e+1, \infty) = \bigcup_{h=1}^{\infty} hA$. In particular, if $d_1, d_2 \in A_L$, then the arithmetic progression $\{d_1 + id_2 \mid i = 0, 1, 2, \ldots,\}$ is contained in $A_L \cup [e+1, \infty)$. In fact, A_L is just the intersection of the multi-dimensional progression $\{i_1 d_1 + i_2 d_2 + \ldots + i_l d_l \mid i_j = 0, 1, 2, \ldots\}$ with $[0, e]$, where $A_L = \{0, d_1, d_2, \ldots, d_l\}$.

Acknowledgements. We wish to warmly thank the referee for the helfpful remarks.

References

[1] J. Deshouillers and V. Lev, A refined bound for sum-free sets in groups of prime order, *Bull. Lond. Math. Soc.*, 40 (2008), no. 5, 863–875.

[2] G. A. Freiman, *Foundations of a structural theory of set addition*, translated from the Russian, Translations of Mathematical Monographs, **37**, American Mathematical Society, Providence, R. I., 1973.

[3] G. A. Freiman, On the detailed structure of sets with small additive property, *Combinatorial Number Theory and Additive Group Theory*, Advanced Courses in Mathematics CRM Barcelona, Birkhäuser Vergal AG, Basel-Boston-Berlin, 2009, 233-239.

[4] G. A. Freiman, The addition of finite sets I, *Izv. Vyss. Ucebn. Zaved. Matematica* 6 (1959), no. 13, 202–213 (Russian) .

[5] G. A. Freiman, Inverse problems of additive number theory VI: On the addition of finite sets III, *Izv. Vys . S.U.cebn. Zaved. Matematika* (1962), no. 3 (28), 151–157 (Russian).

[6] D. J. Grynkiewicz and O. Serra, The Freiman $3k-2$ Theorem: Distinct Summands, preprint.

[7] V. F. Lev and P. Y. Smeliansky, On addition of two distinct sets of integers, *Acta Arith.* **70** (1995), no. 1, 85–91.

[8] V. Lev, Optimal Representations by Sumsets and Subset Sums, *Journal of Number Theory*, 62 (1997), 127-143.

[9] V. Lev, Large sum-free sets in $\mathbb{Z}/p\mathbb{Z}$, *Israel J. Math.*, 154 (2006), 221–233.

[10] M. Nathanson, *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Graduate Texts in Mathematics 165, Springer-Verlag, New York, 1996.

[11] J. L. Ramírez Alfonsín, *The Diophantine Frobenius problem*, Oxford Lecture Series in Mathematics and its Applications, 30, Oxford University Press, Oxford, 2005.

[12] Y. Stanchescu, On addition of two distinct sets of integers, *Acta Arith.* **75** (1996), no. 2, 191–194.

[13] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Studies in Advanced Mathematics 105, Cambridge University Press, Cambridge, 2006.