RESTRICTED SET ADDITION IN GROUPS, I. THE CLASSICAL SETTING

VSEVOLOD F. LEV

Abstract. We survey the existing and prove several new results for the cardinality of the restricted doubling $2\hat{}A = \{a' + a'' : a', a'' \in A,\ a' \neq a''\}$, where $A \subseteq G$ is a subset of the set of elements of an (additively written) group G. In particular, we improve known estimates for $G = \mathbb{Z}$ and $G = \mathbb{Z}/p\mathbb{Z}$ and give a first-of-a-kind general estimate valid for arbitrary G.

1. Background, motivation and summary of results

Let G be an arbitrary group. We use additive notation for the group operation in G, as all particular groups that appear in this paper (excluding the Appendix) are Abelian; however, no commutativity is assumed in general, unless indicated explicitly.

Let $A \subseteq G$ and $B \subseteq G$ be finite non-empty subsets of the set of all elements of G. How small can be the set

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

of all elements representable as a sum of an element of A and an element of B? Though this and related problems are studied in numerous papers, almost nothing is known about the cardinality of the set

$$A \mathbin{\widehat{+}} B = \{a + b : a \in A,\ b \in B,\ a \neq b\}$$

of all sums with *distinct* summands. We are primarily interested in $B = A$ and we abbreviate $2A = A + A$ and $2\hat{}A = A \mathbin{\widehat{+}} A$.

The first case one might think of is $G = \mathbb{Z}$, the group of integers. Here we can shift A to make its minimum element 0 and then divide through all the shifted elements by their greatest common divisor — this normalization, clearly, does not affect the cardinalities of $2A$ and $2\hat{}A$. We denote then by l the maximum element of, and by n the cardinality of A. Thus, there is no loss of generality in writing

$$A \subseteq [0, l], \quad 0, l \in A, \quad \gcd(A) = 1, \quad |A| = n.$$

It was proved by G. Freiman over 30 years ago (see [5]) that under this notation

$$|2A| \geq \min\{l, 2n-3\} + n = \begin{cases} l + n, & \text{if } l \leq 2n-3, \\ 3n-3 & \text{if } l \geq 2n-2, \end{cases}$$

1991 *Mathematics Subject Classification.* Primary: 11B75; Secondary: 05D99, 20F99.

equality being attained, for instance, for $A = \{0, \dots, n-2\} \cup \{l\}$ in which case $2A = \{0, \dots, 2n-4\} \cup \{l, \dots, l+n-2\} \cup \{2l\}$.

Oddly enough, no parallel result for $2\hat{}A$ has ever been obtained, though one was conjectured by Freiman (personal communication) and independently by the present author.

Conjecture 1. *Let A be a set of $n > 7$ integers such that $A \subseteq [0, l]$, $0, l \in A$ and $\gcd(A) = 1$. Then*

$$|2\hat{}A| \ge \min\{l, 2n-5\} + n - 2 = \begin{cases} l+n-2, & \text{if } l \le 2n-5, \\ 3n-7, & \text{if } l \ge 2n-4. \end{cases}$$

This is the strongest possible assertion of this kind, as letting $A = \{0, \dots, n-3\} \cup \{l-1, l\}$ we get $2\hat{}A = \{1, \dots, 2n-7\} \cup \{l-1, \dots, l+n-3\} \cup \{2l-1\}$. (More generally, choose $A = \{0, d, \dots, (n_1-1)d\} \cup \{l-(n-n_1-1)d, \dots, l-d, l\}$ with $2 \le n_1 \le n-2$ and $d \ge 1$ small enough.) The condition $n > 7$ is necessary due to a singularity for $n = 7$: consider $A = \{0, 1, m-1, m, m+1, 2m-1, 2m\}$ with $m = l/2$ sufficiently large.

The trivial estimate is this.

Lemma 1. *Let $A, B \subseteq \mathbb{Z}$ be finite sets of integers. Then*

$$|A \hat{+} B| \ge |A| + |B| - 3.$$

Proof. Write $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_m\}$, the elements being arranged in increasing order. Then among the $n+m-1$ distinct sums

$$a_1 + b_1, \dots, a_1 + b_m, a_2 + b_m, \dots, a_n + b_m$$

at most two do *not* belong to $A \hat{+} B$. $\square$

Notice, that equality is attained when $A = B$ is an arithmetic progression.

Not counting Lemma 1, the only known result in this direction is proved in [6].

Theorem A (Freiman, Low, Pitman [6])**.** *Let A be a set of $n \ge 2$ integers such that $A \subseteq [0, l]$, $0, l \in A$ and $\gcd(A) = 1$. Then*

$$|2\hat{}A| \ge \frac{1}{2}\min\{l, 2n-3\} + \frac{3}{2}n - \frac{7}{2} = \begin{cases} 0.5(l+n) + n - 3.5, & \text{if } l \le 2n-3, \\ 2.5n - 5, & \text{if } l \ge 2n-2. \end{cases}$$

In this paper we get somewhat nearer to Conjecture 1.

Theorem 1. *Let A be a set of $n \ge 3$ integers such that $A \subseteq [0, l]$, $0, l \in A$, and $\gcd(A) = 1$. Then*

$$|2\hat{}A| \ge \begin{cases} l+n-2, & \text{if } l \le 2n-5, \\ (\theta+1)n - 6, & \text{if } l \ge 2n-4, \end{cases}$$

where $\theta = (1+\sqrt{5})/2 \approx 1.61$ is the "golden mean."

This theorem will be proved in Sections 2 and 3 using our reduction method developed in [10, 11, 12]. Very roughly, this method can be described as follows. We consider the image $\bar{A} \subseteq \mathbb{Z}/l\mathbb{Z}$ of A under the canonical homomorphism of $\mathbb{Z}$ onto $\mathbb{Z}/l\mathbb{Z}$ (the set of residues modulo l), and we derive estimates for $|2\hat{\ }A|$ from those for $|2\hat{\ }\bar{A}|$. The problem, however, is that very little is known about $|2\hat{\ }\bar{A}|$. All papers published so far concentrate on the Erdős-Heilbronn conjecture, proposed in [4] and proved in [3]; in particular, the moduli under consideration are prime.

Theorem B (Dias da Silva, Hamidoune; conjectured by Erdős and Heilbronn)**.** *Let $\bar{A} \subseteq \mathbb{Z}/p\mathbb{Z}$ and $\bar{B} \subseteq \mathbb{Z}/p\mathbb{Z}$ be sets of residues modulo a* prime *number p. Then*

$$|\bar{A} \mathbin{\widehat{+}} \bar{B}| \geq \min\{|\bar{A}| + |\bar{B}| - 3, p\}.$$

In [1, 2], Alon, Nathanson, and Ruzsa gave another and easier proof, which also yields similar estimates for the number of sums $a + b$ with $P(a, b) \neq 0$, where P is an arbitrary, fixed polynomial. (Theorem B is obtained for $P(x, y) = x - y$.)

For *sparse* sets of residues, Freiman, Low, and Pitman were able to go further: using their Theorem A, they described all $\bar{A} \subseteq \mathbb{Z}/p\mathbb{Z}$ such that the cardinality of $2\hat{\ }\bar{A}$ slightly exceeds the minimum possible value.

Theorem C (Freiman, Low, Pitman [6])**.** *Let $\bar{A} \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set of n residues modulo a prime p, where $60 < n < p/50$. Suppose that $|2\hat{\ }\bar{A}| \leq 2.06n - 3$. Then $\bar{A}$ is contained in an arithmetic progression of at most $2|2\hat{\ }\bar{A}| - 3n + 8$ terms.*

Using our Theorem 1 (instead of Theorem A) and following the method of Freiman, Low, and Pitman otherwise, we prove in Section 4 the following.

Theorem 2. *Let $\bar{A} \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set of n residues modulo a prime p, where $200 \leq n \leq p/50$. Suppose that $|2\hat{\ }\bar{A}| \leq 2.18n - 6$. Then $\bar{A}$ is contained in an arithmetic progression of at most $|2\hat{\ }\bar{A}| - n + 3$ terms.*

Here the expression $|2\hat{\ }\bar{A}| - n + 3$ is best possible, as the above example shows, when reduced modulo p:

$$\bar{A} = \{0, 1, \ldots, n-3\} \cup \{l-1, l\} \subseteq \mathbb{Z}/p\mathbb{Z}, \quad n - 1 \leq l \leq 1.18n - 4.$$

Conjecturally, the restriction $|2\hat{\ }A| \leq 2.18n - 6$ can be relaxed to $|2\hat{\ }A| < 3n - 7$, and the bound $p/50$ can be replaced by $(p - C)/2$ with a relatively small absolute constant C; this, however, is far beyond the reach of our methods. In fact, the constants 50 and 200 are of a technical nature and can be varied in a certain range. In particular, $1/50$ can be increased to $1/35$ at the expense of increasing 200 to a very large number, like $15,000$.

The attentive reader may have noticed a logical problem: to prove a result for residues (Theorem 2) we need a result for integers (Theorem 1), while the proof of the latter is based on a reduction back to the residues case. For the "regular" (not restricted) doubling $2A$ this problem is overcome by an application of Kneser's theorem to the set $\bar{A} \subseteq \mathbb{Z}/l\mathbb{Z}$.

Given an Abelian group G and a set $C \subseteq G$, the *period* $H = H(C)$ of C is defined by

$$H = \{h \in G \colon C + h = C\}.$$

Observe that $H \subseteq G$ is a subgroup and that C is a union of a number of H-cosets. If $H \neq \{0\}$, then C is said to be *periodic*.

Kneser's theorem (see [8, 9]) is the following.

Theorem D (Kneser). *Let A and B be finite sets of elements of an Abelian group G, and write $H = H(A+B)$. Suppose that $|A+B| \le |A| + |B| - 1$. Then*

$$|A+B| = |A+H| + |B+H| - |H|.$$

Therefore, $H \neq \{0\}$ when $|A+B| < |A| + |B| - 1$:

Corollary 1. *Let A and B be finite sets of elements of an Abelian group G, and suppose that $|A+B| < |A| + |B| - 1$. Then $A+B$ is periodic.*

Though this is not obvious at first glance, Theorem D is equivalent to Corollary 1 in the sense that the former can be easily derived from the latter.

No analogue of Kneser's theorem is known for the restricted sum $A \mathbin{\hat{+}} B$. However, heuristic arguments suggest that non-trivial conclusions about $A \mathbin{\hat{+}} B$ can be drawn, provided $|A \mathbin{\hat{+}} B| < |A| + |B| - (L+2)$, where $L = L(G)$ is the maximum number of pairwise distinct elements of G that share a common doubling:

$$L(G) = \max_{\substack{g_1,\dots,g_\lambda \in G \\ g_i \neq g_j \ (1 \le i < j \le \lambda) \\ 2g_1 = \dots = 2g_\lambda}} \lambda. \tag{1}$$

We call $L(G)$ the *doubling constant* of the group G. Note that for $G = \mathbb{Z}/l\mathbb{Z}$ we have $L(G) = \delta_2(l) + 1$,

$$\delta_2(l) = \begin{cases} 1, & \text{if } 2 \mid l, \\ 0, & \text{if } 2 \nmid l. \end{cases}$$

Some group-theoretic properties of the constant $L(G)$ are discussed in the Appendix.

Actually, Kneser's theorem can be successfully applied in the restricted doubling context if $A \mathbin{\hat{+}} B = A + B$; and in the remaining case $A \mathbin{\hat{+}} B \neq A + B$ (meaning that there exists an element in $A+B$ *not* representable as $a+b$ with $a \neq b$) we conjecture the following.

Conjecture 2. *Let G be an Abelian group with the doubling constant $L = L(G)$. Then*

$$|A \mathbin{\hat{+}} B| \ge |A| + |B| - (L+2)$$

for any finite $A, B \subseteq G$ such that $A \mathbin{\hat{+}} B \neq A + B$.

We show that this conjecture, and even its special case $B = A$, yields

$$|2\hat{}A| \geq \min\{l, 2n - 5 - \delta_2(l)\} + n - 2$$

(where A is a set of n integers such that $A \subseteq [0, l]$, $0, l \in A$ and $\gcd(A) = 1$) which is almost as strong as Conjecture 1. We believe that the "lost" 1 can be recovered, but perhaps it makes no sense to hunt for it until Conjecture 2 is proved.

We note that a classical result of Kemperman [7, Theorem 2] implies that $|A + B| \geq |A| + |B| - L$ under the same condition $A \widehat{+} B \neq A + B$, but his method doesn't seem to be applicable to estimates of $|A \widehat{+} B|$.

In the present paper (Section 3) we use a combinatorial argument to give a partial proof of Conjecture 2 in the case $B = A$. Actually, we go somewhat further by omitting the commutativity requirement.

Theorem 3. *Let G be an arbitrary group with the doubling constant $L = L(G)$. Then*

$$|2\hat{}A| > \theta|A| - (L + 2); \quad \theta = (1 + \sqrt{5})/2$$

for any finite $A \subseteq G$ such that $2\hat{}A \neq 2A$.

We can now outline the plan of attack on Theorems 1 and 2. Given a set of integers $A \subseteq [0, l]$, we consider its reduction $\bar{A} \subseteq \mathbb{Z}/l\mathbb{Z}$. By Theorem 3 as applied to $G = \mathbb{Z}/l\mathbb{Z}$, either $|2\hat{}\bar{A}|$ is large, or $2\hat{}\bar{A} = 2\bar{A}$. In the former case it is not difficult to see that $|2\hat{}A|$ is also large; in the latter case we use Kneser's theorem to derive structure information about $2\bar{A}$ which allows us to complete the proof of Theorem 1. Once Theorem 1 is proved, we use it in conjunction with the method of Freiman, Low and Pitman to prove Theorem 2.

2. Reduction method

With possible generalizations in mind, we formulate and prove the following lemma for h-fold restricted addition, $h \geq 2$. The corresponding restricted sums are defined in the natural way:

$$h\hat{}A = \{a_1 + \cdots + a_h \colon a_i \in A \ (1 \leq i \leq h),\ a_i \neq a_j \ (1 \leq i < j \leq h)\}.$$

Our convention for using overlined symbols: given an integer $l \geq 2$, overlined characters are used for objects (sets or elements) in $\mathbb{Z}/l\mathbb{Z}$, and same characters without the overline sign are used for their pre-images in $\mathbb{Z}$.

Lemma 2. *Let $A \subseteq [0, l]$ be a set of integers such that $0, l \in A$. Write $A' = A \setminus \{0, l\}$ and let $\bar{A} \subseteq \mathbb{Z}/l\mathbb{Z}$ be the image of A under the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/l\mathbb{Z}$. Then*

$$|h\hat{}A| \geq |h\hat{}\bar{A}| + |(h-1)\hat{}A'|.$$

Proof. We have:

$$
\begin{aligned}
|h\hat{}A| &\overset{(1)}{\geq} \sum_{\bar{c}\in h\hat{}\bar{A}} \#\{\text{pre-images of } \bar{c} \text{ in } h\hat{}A\} \\
&= \sum_{\bar{c}\in h\hat{}\bar{A}} \big(\#\{\text{pre-images of } \bar{c} \text{ in } h\hat{}A\} - 1\big) + |h\hat{}\bar{A}| \\
&\overset{(2)}{\geq} \sum_{\bar{c}\in (h-1)\hat{}\bar{A}'} (\#\{\text{pre-images of } \bar{c} \text{ in } h\hat{}A\} - 1) + |h\hat{}\bar{A}| \\
&\overset{(3)}{\geq} \sum_{\bar{c}\in (h-1)\hat{}\bar{A}'} \#\{\text{pre-images of } \bar{c} \text{ in } (h-1)\hat{}A'\} + |h\hat{}\bar{A}| \\
&\overset{(4)}{=} |(h-1)\hat{}A'| + |h\hat{}\bar{A}|.
\end{aligned}
$$

Explanations follow.

(1) Each element of $h\hat{}\bar{A}$ has at least one pre-image in $h\hat{}A$ and distinct elements, plainly, have distinct pre-images;
(2) $(h-1)\hat{}\bar{A}' \subseteq h\hat{}\bar{A}$, as $0 \in \bar{A} \setminus \bar{A}'$;
(3) If $c_1 < \dots < c_s$ are distinct pre-images of $\bar{c}$ in $(h-1)\hat{}A'$, then $c_1 < \dots < c_s < c_s+l$ are distinct pre-images of $\bar{c}$ in $h\hat{}A$;
(4) Each element of $(h-1)\hat{}\bar{A}'$ has a pre-image in $(h-1)\hat{}A'$, and each element of $(h-1)\hat{}A'$ has an image in $(h-1)\hat{}\bar{A}'$.

□

Below, we need only a particular case of Lemma 2.

Corollary 2. *Let $A \subseteq [0, l]$ be a set of $n = |A|$ integers such that $0, l \in A$, and let $\bar{A}$ be the canonical image of A in $\mathbb{Z}/l\mathbb{Z}$. Then*

$$|2\hat{}A| \geq |2\hat{}\bar{A}| + n - 2.$$

Proof. Apply Lemma 2 with $h = 2$ and observe that $|A'| = n - 2$. □

In fact, a straightforward proof of Corollary 2 would be a bit simpler than the proof of Lemma 2: just notice that each element of $2\hat{}\bar{A}$ has a pre-image in $2\hat{}A$, and that the $n-2$ elements $\bar{a}$ (where $a \in A \setminus \{0, l\}$) each have two distinct origins in $2\hat{}A$, namely a and $a + l$.

The following is a striking, but probably useless consequence.

Corollary 3. *Conjecture 1 holds if l is a prime number. That is,*

$$|2\hat{}A| \geq \min\{l, 2n-5\} + n - 2,$$

where $A \subseteq [0, l],\ 0, l \in A,\ n = |A|$ and l is prime.

Proof. Apply Corollary 2 and observe that

$$|2\hat{\ }\bar{A}| \geq \min\{l, 2n-5\}$$

by Theorem B (as $2|\bar{A}|-3 = 2(n-1)-3 = 2n-5$). □

Our next corollary is less impressive, but more important; it establishes Conjecture 1 in the particular case of "large" restricted doubling $2\hat{\ }\bar{A}$.

Corollary 4. *Let $A \subseteq [0, l]$ be a set of $n = |A| \geq 3$ integers such that $0, l \in A$, and let $\bar{A} \in \mathbb{Z}/l\mathbb{Z}$ be the canonical image of A in $\mathbb{Z}/l\mathbb{Z}$. Suppose that $|2\hat{\ }\bar{A}| \geq 2|\bar{A}|-3$. Then*

$$|2\hat{\ }A| \geq 3n-7.$$

Proof. This follows immediately from Corollary 2. □

Another conclusion is that Conjecture 1 holds if $l \leq 2n-5$.

Corollary 5. *Let $A \subseteq [0, l]$ be a set of $n = |A|$ integers such that $0, l \in A$, and suppose that $l \leq 2n-5$. Then*

$$|2\hat{\ }A| \geq l+n-2.$$

Proof. In view of Corollary 2, it suffices to show that $2\hat{\ }\bar{A} = \mathbb{Z}/l\mathbb{Z}$. Indeed, let $\bar{c}$ be an arbitrary element of $\mathbb{Z}/l\mathbb{Z}$. Since $|\bar{A}| + |\bar{c}-\bar{A}| = 2|\bar{A}| = 2n-2 \geq l+3$, the sets $\bar{A}$ and $\bar{c}-\bar{A}$ intersect by at least three distinct elements and therefore we have at least three distinct representations

$$\bar{c} = \bar{a}'_i + \bar{a}''_i; \quad \bar{a}'_i, \bar{a}''_i \in \bar{A} \quad (i = 1, 2, 3).$$

Now, at least one of the pairs $(\bar{a}'_i, \bar{a}''_i)$ satisfies $\bar{a}''_i \neq \bar{a}'_i$ — otherwise $2\bar{a}'_1 = 2\bar{a}'_2 = 2\bar{a}'_3$, which contradicts $L(\mathbb{Z}/l\mathbb{Z}) \leq 2$. □

While Lemma 2 provides a relatively simple application of the reduction method, the following result is more technical. Establishing Conjecture 1 for $2\hat{\ }\bar{A} = 2\bar{A}$, it will allow us to restrict attention to the case when $2\hat{\ }\bar{A}$ is distinct from $2\bar{A}$.

Lemma 3. *Let $A \subseteq [0, l]$ be a set of $n = |A| \geq 3$ integers such that $0, l \in A$ and* $\gcd(A) = 1$. *Furthermore, let $\bar{A}$ be the canonical image of A in $\mathbb{Z}/l\mathbb{Z}$, and suppose that $2\hat{\ }\bar{A} = 2\bar{A}$. Then*

$$|2\hat{\ }A| \geq \min\{l, 2n-5\} + n - 2.$$

Proof. By Corollary 4, we can assume

$$|2\hat{\ }\bar{A}| = |2\bar{A}| \leq 2|\bar{A}| - 4. \tag{2}$$

Let $H = H(2\bar{A})$ be the period of $2\bar{A}$. By Kneser's theorem (Theorem D),

$$|H| = 2|\bar{A}+H| - |2\bar{A}| \geq 2|\bar{A}| - (2|\bar{A}|-4) = 4.$$

Obviously, $\bar{A}+H\subseteq 2\bar{A}$ (as $\bar{A}\subseteq 2\bar{A}$), and without loss of generality we confine ourselves to the case $\bar{A}+H\subsetneqq 2\bar{A}$: otherwise,

$$2\bar{A}=2\bar{A}+H=\bar{A}+(\bar{A}+H)=3\bar{A}=2\bar{A}+H+\bar{A}=4\bar{A}=\cdots=\mathbb{Z}/l\mathbb{Z}$$

(as $\bar{A}$ generates $\mathbb{Z}/l\mathbb{Z}$ in view of $\gcd(A)=1$) and by Corollary 2,

$$|2\hat{\ }A|\geq|2\hat{\ }\bar{A}|+n-2=|2\bar{A}|+n-2=l+n-2.$$

We now write

$$2\hat{\ }\bar{A}=\bar{A}\cup\big((\bar{A}+H)\setminus\bar{A}\big)\cup\big(2\hat{\ }\bar{A}\setminus(\bar{A}+H)\big)$$

and we subdivide the elements $c\in 2\hat{\ }A$ into three classes depending on the set in the right-hand side that $\bar{c}$ falls into. The total number of elements c in each class is then counted separately.

(i) $\#\{c\in 2\hat{\ }A\colon \bar{c}\in\bar{A}\}\geq 2n-3$: write $A=\{a_1,\dots,a_n\}$ (where $a_i<a_{i+1},\ i=1,\dots,n-1$) and consider

$$a_1+a_2<\cdots<a_1+a_{n-1}<a_1+a_n<a_2+a_n<\cdots<a_{n-1}+a_n.$$

(ii) $\#\{c\in 2\hat{\ }A\colon \bar{c}\in(\bar{A}+H)\setminus\bar{A}\}\geq|\bar{A}+H|-|\bar{A}|$: indeed, as $(\bar{A}+H)\setminus\bar{A}\subset 2\hat{\ }\bar{A}$, any element of $(\bar{A}+H)\setminus\bar{A}$ has at least one pre-image in $2\hat{\ }A$.

(iii) $\#\{c\in 2\hat{\ }A\colon \bar{c}\in 2\hat{\ }\bar{A}\setminus(\bar{A}+H)\}\geq 2|\bar{A}|-|\bar{A}+H|-3$. This is *the* estimate. To prove it, we first notice that $2\hat{\ }\bar{A}\setminus(\bar{A}+H)$ consists of $N=(|2\hat{\ }\bar{A}|-|\bar{A}+H|)/|H|>0$ H-cosets, each of the form $\bar{a}_1+\bar{a}_2+H$ with some $a_1,a_2\in A\setminus\{0,l\}$ and $\bar{a}_i\notin H$. Let $A_i=\{a\in A\colon \bar{a}\in\bar{a}_i+H\}$ $(i=1,2)$. Then

$$\begin{aligned}|A_i|=|\bar{A}_i|=|H|-|(\bar{a}_i+H)\setminus\bar{A}_i|&=|H|-|(\bar{a}_i+H)\setminus\bar{A}|\\&\geq|H|-|(\bar{A}+H)\setminus\bar{A}|=|\bar{A}|+|H|-|\bar{A}+H|.\end{aligned}$$

Thus by Lemma 1,

$$|A_1\hat{+}A_2|\geq|A_1|+|A_2|-3\geq 2|\bar{A}|+2|H|-2|\bar{A}+H|-3,$$

and we denote the right-hand side by K. Therefore, the elements of each H-coset (of the N that comprise $2\hat{\ }\bar{A}\setminus(\bar{A}+H)$) have together at least K pre-images in $2\hat{\ }A$. Moreover, by (2) and Kneser's theorem,

$$K\geq|2\bar{A}|+1+2|H|-2|\bar{A}+H|\geq|H|+1,$$

and we can bound the quantity in question from below by

$$\begin{aligned}KN&=(K-|H|)(N-1)+K+N|H|-|H|\\&\geq K+N|H|-|H|\\&=|2\hat{\ }\bar{A}|-3|\bar{A}+H|+2|\bar{A}|+|H|-3\\&=2|\bar{A}|-|\bar{A}+H|-3.\end{aligned}$$

(We used here Kneser's theorem a third time.)

Finally, putting together the estimates of (i), (ii), and (iii) we get

$$\begin{aligned}|2\hat{\ }A| &\geq (2n-3) + (|\bar{A}+H| - |\bar{A}|) + (2|\bar{A}| - |\bar{A}+H| - 3)\\ &= 2n + |\bar{A}| - 6\\ &= 3n-7,\end{aligned}$$

completing the proof. □

The connection between Conjectures 1 and 2 now becomes apparent.

Corollary 6. *Let $A \subseteq [0,l]$ be a set of $n = |A| \geq 3$ integers such that $0, l \in A$ and* $\gcd(A) = 1$. *Then, assuming Conjecture 2, we have*

$$|2\hat{\ }A| \geq \min\{l, 2n-5-\delta_2(l)\} + n - 2.$$

Proof. Let $\bar{A} \in \mathbb{Z}/l\mathbb{Z}$ be defined as usual. By Conjecture 2 as applied to $\bar{A}$, either $|2\hat{\ }\bar{A}| \geq 2|\bar{A}| - \delta_2(l) - 3 = 2n-5-\delta_2(l)$ (since $L = 1+\delta_2(l)$), or $2\hat{\ }\bar{A} = 2\bar{A}$. In the former case the assertion follows by Corollary 2, in the latter case by Lemma 3. □

Above, we proved unconditionally Conjecture 1 in the very special case of prime l (Corollary 3). Now we consider another particular case, specifically symmetric sets: $A = l - A$. Of course, both these situations are fairly artificial. However, they support the general conjecture and the proofs are good illustrations of our method.

Corollary 7. *Let $A \subseteq [0,l]$ be a set of $n = |A| \geq 3$ integers such that $0, l \in A$ and* $\gcd(A) = 1$, *and suppose that A is symmetric: $A = l - A$. Then*

$$|2\hat{\ }A| \geq \min\{l, 2n-5+\delta_2(l)\} + n - 2.$$

Proof. Define $\bar{A}$ in the usual way. By Lemma 3 we can restrict ourselves to the situation $2\hat{\ }\bar{A} \neq 2\bar{A}$, in which case there exists $\bar{a} \in \bar{A}$ such that $2\bar{a} \notin 2\hat{\ }\bar{A}$. Then

$$|(\bar{a} - \bar{A}) \cap (\bar{A} - \bar{a})| \leq 1 + \delta_2(l)$$

as $\bar{c} = \bar{a} - \bar{a}' = \bar{a}'' - \bar{a}$ implies $2\bar{a} = \bar{a}' + \bar{a}''$, whence $\bar{a}' = \bar{a}''$ and $2\bar{a} = 2\bar{a}'$. It follows that

$$|(\bar{a} - \bar{A}) \cup (\bar{A} - \bar{a})| \geq 2|\bar{A}| - 1 - \delta_2(l).$$

On the other hand, in view of the symmetry of A, the set $\bar{A}$ is also symmetric in the sense that $-\bar{A} = \bar{A}$, and therefore both $\bar{a} - \bar{A}$ and $\bar{A} - \bar{a}$ are subsets of $2\hat{\ }\bar{A} \cup \{2\bar{a}\} \cup \{-2\bar{a}\}$. Thus,

$$\begin{gathered}|2\hat{\ }\bar{A}| + 2 \geq 2|\bar{A}| - 1 - \delta_2(l),\\ |2\hat{\ }\bar{A}| \geq 2|\bar{A}| - 3 - \delta_2(l) = 2n - 5 - \delta_2(l)\end{gathered}$$

and the result follows from Corollary 2. □

3. The combinatorial kernel: proofs of Theorems 3 and 1

We now turn to the proof of Theorem 3 — the core result, on which both Theorems 1 and 2 are based. In the course of the proof, we use the notation $A - A$ for the set of all elements of G, representable as $a' - a''$ with $a', a'' \in A$, and $-A + A$ for the set of all elements of G, representable as $-a' + a''$. We have to distinguish these two sets, since G is not assumed to be Abelian.

Proof of Theorem 3. Assume the conclusion fails:

$$|2\hat{\ }A| < \theta n - L - 2, \tag{3}$$

where we write $n = |A|$.

(i) Obviously, there exists $\sigma \in 2\hat{\ }A$ such that the number of representations of σ in the form $\sigma = a' + a''$ $(a', a'' \in A)$ is at least

$$\frac{n^2 - n}{\theta n - L - 2} > \frac{n(n-1)}{\theta(n-1)} = (\theta - 1)n.$$

(ii) Given $c = a_0'' - a_0' \in A - A$, we write $A' = A \setminus \{a_0'\}$ and $A'' = A \setminus \{a_0''\}$. As both $A' + a_0'$ and $A'' + a_0''$ are subsets of $2\hat{\ }A$, we have

$$|(A' + a_0') \cap (A'' + a_0'')| \ge |A'| + |A''| - |2\hat{\ }A| > 2(n-1) - (\theta n - L - 2) = (2 - \theta)n + L.$$

Now any solution of $a' + a_0' = a'' + a_0''$ in $a', a'' \in A$ yields a representation $c = -a'' + a'$. This shows that any $c \in A - A$ has more than $(2 - \theta)n + L$ representations of the form $c = -a'' + a'$ $(a', a'' \in A)$.

An immediate conclusion is that $A - A \subseteq -A + A$. Similarly, $-A + A \subseteq A - A$ (fix $c = -a_0'' + a_0' \in A - A$ and estimate $|(a_0' + A) \cap (a_0'' + A)|$). Therefore,

$$A - A = -A + A.$$

(iii) Given $c \in A - A$, consider the subset of A consisting of all those a'' which can appear as a first term in $c = -a'' + a'$. On the other hand, consider the subset of A comprised of all those a'' which can appear as a second term in $\sigma = a' + a''$. By (i) and (ii) the sum of the cardinalities of these two subsets is greater than

$$(\theta - 1)n + (2 - \theta)n + L = n + L,$$

and we conclude that there are more than L common values of a'', resulting in $L + 1$ equalities

$$\sigma = a_i + a_i'',\ c = -a_i'' + a_i'; \quad a_i, a_i', a_i'' \in A \quad (i = 1, \dots, L + 1)$$

and further to $L + 1$ equalities

$$\sigma + c = a_i + a_i'.$$

We observe that at least one index i satisfies $a_i \neq a'_i$: otherwise $2a_1 = \cdots = 2a_{L+1}$. It follows that $\sigma + c \in 2\hat{}A$ for any $c \in A - A$, that is

$$\sigma + (A - A) \subseteq 2\hat{}A. \tag{4}$$

(iv) To get a contradiction, we show that $A - A$ is "too large" to satisfy (4). Indeed, fix $a \in A$ such that $2a \notin 2\hat{}A$. (Such an a exists in view of the hypothesis $2A \neq 2\hat{}A$ which had not yet been used.) Then

$$|(a - A) \cap (-a + A)| \leq L$$

as any solution of $a - a' = -a + a''$ in $a', a'' \in A$ necessarily satisfies $a' = a''$, and therefore there exist at most L solutions. Now both $a - A$ and $-a + A$ are subsets of $A - A = -A + A$, hence by (3)

$$|A - A| \geq |a - A| + |-a + A| - L = 2n - L > |2\hat{}A|,$$

contradicting (4). □

It will be noted that for commutative G certain simplifications of the proof are possible. Theorem 1 now follows easily.

Proof of Theorem 1. By Corollary 5, it suffices to consider the case $l \geq 2n - 4$. Define $\bar{A}$ as usual. By Theorem 3, either $2\hat{}\bar{A} = 2\bar{A}$, or $|2\hat{}\bar{A}| > \theta n - 3 - \delta_2(l)$. In the former case the result follows immediately from Lemma 3, in the latter case from Corollary 2. □

4. Proof of Theorem 2

Since the proof follows closely that of [6, Theorem 2], which in turn is very similar to the proof of [5, Theorem 2.1] where non-restricted doubling $2A$ was considered, we only sketch here the argument omitting technical details.

(i) Given $C \subseteq \mathbb{Z}/p\mathbb{Z}$ and $z \in \mathbb{Z}/p\mathbb{Z}$, we define

$$S_C(z) = \sum_{c \in C} e^{2\pi i \frac{cz}{p}},$$

so that $S_C(0) = |C|$ and

$$\sum_{z=0}^{p-1} |S_C(z)|^2 = p|C|,$$

whence

$$\sum_{z=1}^{p-1} |S_C(z)|^2 = |C|(p - |C|). \tag{5}$$

(ii) We write $n = |A|$ and consider the sum

$$\sum_{\substack{a',a''\in A\\ a\in 2A}} \sum_{z=0}^{p-1} e^{2\pi i \frac{a'+a''-a}{p} z}.$$

On the one hand, this sum equals $n^2 p$ as to any $a', a'' \in A$ there corresponds precisely one $a \in 2A$ such that $a' + a'' = a$. On the other hand, it can be represented as

$$\sum_{\substack{a',a''\in A\\ a\in 2\hat{}A}} + \sum_{\substack{a',a''\in A\\ a\in 2A\setminus 2\hat{}A}},$$

and here the second summand equals $p|2A \setminus 2\hat{}A| \le np$, as any $a \in 2A \setminus 2\hat{}A$ has precisely one representation $a = a' + a''$. Thus, the first summand is at least $(n^2 - n)p$, which can be rewritten as

$$\sum_{z=0}^{p-1} S_A^2(z) S_{2\hat{}A}(-z) \ge (n^2 - n)p.$$

We further single out the term with $z = 0$ to obtain

$$\sum_{z=1}^{p-1} S_A^2(z) S_{2\hat{}A}(-z) \ge (n^2 - n)p - n^2 T, \tag{6}$$

where for brevity we write $T = |2\hat{}A|$.

(iii) Define

$$M = \max_{1\le z\le p-1} |S_A(z)|.$$

Then (6) implies

$$M \sum_{z=1}^{p-1} |S_A(z)||S_{2\hat{}A}(z)| \ge (n^2 - n)p - n^2 T$$

and using Cauchy-Schwarz and (5) (as applied to $C = A$ and $C = 2\hat{}A$) we conclude that

$$\frac{M}{n} \ge \frac{(n-1)p - nT}{\sqrt{n(p-n)}\sqrt{T(p-T)}}.$$

As the right-hand side is easily seen to be a decreasing function of T on $[0, p/2]$, and $2.18n < p/2$, it follows that

$$\begin{aligned}\frac{M}{n} &> \frac{(n-1)p - 2.18n^2}{\sqrt{n(p-n)}\sqrt{2.18n(p-2.18n)}} \\ &= \frac{1}{\sqrt{2.18}} \frac{(n-1)p - 2.18n^2}{np\sqrt{1-n/p}\sqrt{1-2.18n/p}} \\ &= \frac{1}{\sqrt{2.18}} \frac{\alpha p - 1 - 2.18\alpha^2 p}{\alpha p\sqrt{1-\alpha}\sqrt{1-2.18\alpha}} \\ &= \frac{1}{\sqrt{2.18}} \frac{1 - 2.18\alpha - 1/n}{\sqrt{1-3.18\alpha+2.18\alpha^2}} \\ &\ge \frac{1}{\sqrt{2.18}} \frac{1 - 2.18\alpha - 0.005}{\sqrt{1-3.18\alpha+2.18\alpha^2}}\end{aligned}$$

where $\alpha \in [0, 1/50]$ is defined by $n = \alpha p$. A tedious but straightforward calculation establishes the convexity of the right-hand side as a function of α and thus shows that the minimum is attained at one of the endpoints. However, the values at the endpoints are both greater than 0.6655, whence $M > 0.6655n$ and

$$|S_A(z_0)| \ge 0.6655n$$

for some $z_0 \in \mathbb{Z}/p\mathbb{Z},\ z_0 \neq 0$.

(iv) By [5, Lemma 2.2] (or [6, Lemma 3.2]) there exists a subset $A_0 \subseteq A$ of cardinality

$$|A_0| > \frac{1+0.6655}{2}|A| > 0.8327n$$

and a residue $u \in \mathbb{Z}/p\mathbb{Z}$ such that

$$A_0 \subseteq \{u, u+z_0', \ldots, u+((p-1)/2)z_0'\},$$

where z_0' is the inverse of z_0 in $\mathbb{Z}/p\mathbb{Z}$.

Let $B_0 \subseteq [0, (p-1)/2]$ be the set of all integers b such that $u + \bar{b}z_0' \in A_0$. (As usual, $\bar{b}$ stands for the residue (mod p) corresponding to the integer b.) Applying appropriate affine transformations $x \mapsto \lambda x + \mu$ with $\lambda \not\equiv 0 \,(\mathrm{mod}\ p)$ to A and B_0, we can ensure that

$$\min(B_0) = 0,\ \gcd(B_0) = 1 \quad \text{and} \quad A_0 = B_0 \pmod p. \tag{7}$$

It is worth pointing out that neither the cardinality $|2\hat{\ }A|$, nor the property of A to be contained in an arithmetic progression of a given length are affected by non-singular ($\lambda \not\equiv 0 \,(\mathrm{mod}\ p)$) affine transformations. Thus, without loss of generality we assume (7).

Clearly, B_0 is isomorphic to A_0 in the sense that $a_1 + a_2 = a_3 + a_4$ holds in $\mathbb{Z}/p\mathbb{Z}$ for some elements $a_i \in A_0$ if and only if $b_1 + b_2 = b_3 + b_4$ holds in $\mathbb{Z}$ for the corresponding elements $b_i \in B_0$; the crucial point here is $B_0 \subseteq [0, p/2)$. It follows that B_0 is a "large" set of integers ($|B_0| = |A_0| > 0.8327n$) with a "small" restricted doubling ($|2\hat{\ }B_0| = |2\hat{\ }A_0| \le$

$|2\hat{\ }A| \le 2.18n - 6$), and we can use Theorem 1 to show that B_0 is contained in a short interval. Formally, we write

(8) $|2\hat{\ }B_0| = |2\hat{\ }A_0| \le |2\hat{\ }A| \le 2.18n - 6 < 2.6180 \cdot 0.8327n - 6 < (1+\theta)|B_0| - 6.$

Let l_0 be the maximal element of B_0. By Theorem 1 and in view of (8), we have

$$l_0 \le |2\hat{\ }B_0| - |B_0| + 2 \le 2.18n - 4 < p/6.$$

(v) The next step is to verify that $A \subseteq [-l_0, 2l_0] \pmod p$. Indeed, otherwise we could pick any element $a \in A$ outside the indicated interval and notice that $(2\hat{\ }A_0) \cap (a + A_0) = \varnothing$ implying

$$|2\hat{\ }A| \ge |2\hat{\ }A_0| + |a + A_0| = |2\hat{\ }B_0| + |B_0| \ge 3|B_0| - 3 > 2.4981n - 3 > 2.18n - 6,$$

a contradiction.

(vi) Now we have the whole set A embedded in the interval $[-l_0, 2l_0] \pmod p$ of length $3l_0 < p/2$ and we essentially repeat the argument above.

Applying an appropriate affine transformation, we can assume that A is an image in $\mathbb{Z}/p\mathbb{Z}$ of a set of co-prime integers $B \subseteq [0, l]$ such that $0, l \in B$ and $l < p/2$. Then A is isomorphic to B and

$$|2\hat{\ }B| = |2\hat{\ }A| \le 2.18n - 6 < (1+\theta)|B| - 6;$$

hence by Theorem 1

$$l \le |2\hat{\ }B| - |B| + 2 = |2\hat{\ }A| - |A| + 2.$$

This completes the proof.

Appendix. A group-theoretic property of the doubling constant.

The doubling constant $L(G)$ is, perhaps, of some group-theoretic interest. A closely related characteristic of G which might be easier to understand and compute is $L_1(G)$, the number of pairwise distinct elements of G whose doubling is the identity element 0:

$$L_1 = \max_{\substack{g_1,\dots,g_\lambda \in G \\ g_i \ne g_j \ (1 \le i < j \le \lambda) \\ 2g_1 = \dots = 2g_\lambda = 0}} \lambda \tag{9}$$

(cf. (1)). Clearly, $L_1(G) \le L(G)$ for any group G. Moreover, if G is Abelian, then

$$L_1(G) = L(G) \tag{10}$$

as $2g_1 = \dots = 2g_L$ leads to $2(g_1 - g_L) = \dots = 2(g_L - g_L) = 0$ in the commutative case. Therefore, if $G = \bigoplus G_i$ is a direct sum of cyclic groups, then $L(G) = L_1(G) = 2^s$, where s is the number of components of even order.

The non-commutative case is subtler. Here we can prove (10), provided that all elements of G are of finite order, not divisible by 4. Indeed, suppose that

$$2g_1 = \cdots = 2g_L, \tag{11}$$

and assume first that g_L (say) has odd order m. Multiplying (11) by $(m+1)/2$, we obtain

$$(m+1)g_1 = \cdots = (m+1)g_{L-1} = g_L.$$

Thus, g_L commutes with each g_i ($i = 1, \dots, L-1$), and it is easy to deduce, as in the commutative case, that $2(g_1 - g_L) = \cdots = 2(g_L - g_L) = 0$; this proves (10). Assume now that (11) holds, and that g_L is of order $2m$, where m is odd. Then

$$2(mg_1) = \cdots = 2(mg_L) = 0,$$

and all mg_i are distinct, since for m odd, $mg_i = mg_j$ along with $2g_i = 2g_j$ implies $g_i = g_j$.

It may come as a surprise that the non-divisibility by 4 condition is, indeed, essential: for the group $\mathcal{Q} = \{\pm 1, \pm i, \pm j, \pm k\}$ of quaternion units we have $L(\mathcal{Q}) = 6$: $(\pm i)^2 = (\pm j)^2 = (\pm k)^2 = -1$, while $L_1(\mathcal{Q}) = 2$: $(\pm 1)^2 = 1$. (Here and in the next example we retain the traditional multiplicative notation.)

Moreover, there are groups G, all elements of which, save for the identity element, are of infinite order (so that $L_1(G) = 1$), yet for any non-identity element $g \in G$ there exist infinitely many $f \in G$ such that $f^2 = g^2$ (so that $L(G) = \infty$). Specifically, fix an open interval I and consider the group G of all monotonically increasing continuous bijections of I onto itself, with composition as group operation. No element $f \neq \mathrm{id}$ is of finite order: if $f(x_0) > x_0$ for some $x_0 \in I$, then by monotonicity

$$x_0 < f(x_0) < f^2(x_0) < \cdots < f^n(x_0)$$

for any positive integer n, whence $f^n \neq \mathrm{id}$; and $f(x_0) < x_0$ is dealt with similarly. On the other hand, for any $g \neq \mathrm{id}$ there exist infinitely many $f \in G$ satisfying

$$f^2(x) = g^2(x). \tag{12}$$

We show this assuming that g has no fixed points (to which the general case easily reduces). We first fix an arbitrary $x_0 \in I$ and for $n \in \mathbb{Z}$ define $I_n = [g^n(x_0), g^{n+1}(x_0)]$. It is a routine exercise to verify that I is a union of all I_n, disjoint except for the endpoints where the neighboring I_n abut. Next, we define f to be any continuous, increasing bijection of I_0 onto I_1. Finally, we extend the definition of f onto the whole interval I by letting for $x \in I_n$

$$f(x) = \begin{cases} g^n f g^{-n}(x), & \text{if } 2 \mid n, \\ g^{n+1} f^{-1} g^{-n+1}(x), & \text{if } 2 \nmid n. \end{cases}$$

It is easily seen that this definition is correct and produces a continuous, monotonically increasing function f on I, that maps each interval I_n onto the "next" interval I_{n+1}. By a direct substitution one can then verify that f satisfies the required identity $f^2(x) = g^2(x)$ for all $x \in I$.

Acknowledgment

The observations in the Appendix (connections between the characteristics $L(G)$ and $L_1(G)$) resulted from a discussion with Ed Azoff. It is our pleasure to thank him for his interest in the problem and valuable suggestions.

References

[1] N. Alon, M.B. Nathanson and I.Z. Ruzsa, Adding distinct congruence classes modulo a prime, *American Math. Monthly*, **102** (1995), 250–255.
[2] N. Alon, M.B. Nathanson and I.Z. Ruzsa, The Polynomial Method and Restricted Sums of Congruence Classes, *J. Number Theory*, **56** (1996), 404–417.
[3] J.A. Dias da Silva and Y.O. Hamidoune, Cyclic space for Grassmann derivatives and additive theory, *Bull. London Math. Soc.* **26** (1994), 140–146.
[4] P. Erdős and H. Heilbronn, On the addition of residue classes $\pmod p$, *Acta Arithmetica* **9** (1964), 149–159.
[5] G. Freiman, *Foundations of a structural theory of set addition*, Kazan 1966 [Russian]. English translation in: *Translations of Math. Monographs* **37** (1973), American Math. Soc., Providence.
[6] G. Freiman, L. Low and J. Pitman, Sumsets with distinct summands and the conjecture of Erdős-Heilbronn on sums of residues, *Astèrisque*, to appear.
[7] J.H.B. Kemperman, On complexes in a semigroup, *Indag. Math.* **18** (1956), 247–254.
[8] M. Kneser, Abschätzung der asymptotischen Dichte von Summenmengen, Math. Z. **58** (1953), 459–484.
[9] ———, Ein Satz über abelschen Gruppen mit Anwendungen auf die Geometrie der Zahlen, *Math. Z.* **61** (1955), 429–434.
[10] V. Lev and P. Smeliansky, On addition of different sets of integers, *Acta Arithmetica* **LXX.1** (1995), 85–91.
[11] V. Lev, Structure theorem for multiple addition and the Frobenius problem, *J. Number Theory,* **58**(1) (1996), 79–88.
[12] ———, Addendum to "Structure theorem for multiple addition", *J. Number Theory,* **65**(1) (1997), 96–100.

Department of Mathematics, The University of Georgia, Athens, GA 30602, USA
E-mail address: `seva@@math.uga.edu`