GENERALIZED MORE SUMS THAN DIFFERENCES SETS

GEOFFREY IYER, OLEG LAZAREV, STEVEN J. MILLER, AND LIYANG ZHANG

ABSTRACT. A More Sums Than Differences (MSTD, or sum-dominant) set is a finite set $A \subset \mathbb{Z}$ such that $|A + A| < |A - A|$. Though it was believed that the percentage of subsets of $\{0, \ldots, n\}$ that are sum-dominant tends to zero, in 2006 Martin and O'Bryant [MO] proved that a positive percentage are sum-dominant. We generalize their result to the many different ways of taking sums and differences of a set. We prove that $|\epsilon_1 A + \cdots + \epsilon_k A| > |\delta_1 A + \cdots + \delta_k A|$ a positive percent of the time for all nontrivial choices of $\epsilon_j, \delta_j \in \{-1, 1\}$. Previous approaches proved the existence of infinitely many such sets given the existence of one; however, no method existed to construct such a set. We develop a new, explicit construction for one such set, and then extend to a positive percentage of sets.

We extend these results further, finding sets that exhibit different behavior as more sums/differences are taken. For example, we prove that for any m, $|\epsilon_1 A + \cdots + \epsilon_k A| - |\delta_1 A + \cdots + \delta_k A| = m$ a positive percentage of the time. We find the limiting behavior of $kA = A + \cdots + A$ for an arbitrary set A as $k \to \infty$ and an upper bound of k for such behavior to settle down. Finally, we say A is k-generational sum-dominant if A, $A + A$, $\ldots$, kA are all sum-dominant. Numerical searches were unable to find even a 2-generational set (heuristics indicate that the probability is at most 10^{-9}, and quite likely significantly less). We prove that for any k a positive percentage of sets are k-generational, and no set can be k-generational for all k.

CONTENTS

1. INTRODUCTION

Given a finite set of integers A, two natural sets to study are

$$
\begin{aligned}
A + A &= \{a_1 + a_2 : a_1, a_2 \in A\} \\
A - A &= \{a_1 - a_2 : a_1, a_2 \in A\}.
\end{aligned} \tag{1.1}
$$

The most natural question to ask is: As we vary A over a family of sets, how often is $|A + A| > |A - A|$ (where $|X|$ is the cardinality of X)? We call such sets More Sums Than Differences (MSTD) sets, or sum-dominant (if the two cardinalities are the same we say A is balanced, and if $|A - A| > |A - A|$ we say A is difference-dominant). As addition is commutative but subtraction is not, a typical pair contributes two differences to $A - A$ but only one sum to $A + A$. While there are numerous constructions of such sets and infinite families of such sets [He, HM2, Ma, MOS, Na2, Na3, Na4, Ru1, Ru2, Ru3], one expects sum-dominant sets to be rare; however, Martin and O'Bryant [MO] proved that a positive percentage of sets are sum-dominant. They showed the percentage is at least $2 \cdot 10^{-7}$, which was improved by Zhao [Zh2] to at least $4.28 \cdot 10^{-4}$ (Monte Carlo simulations suggest the true answer is about $4.5 \cdot 10^{-4}$). In all these arguments, each integer in $\{0, \ldots, n - 1\}$ has an equal chance of being in A or not being in A, and thus all of the 2^n subsets are equally likely to be chosen. The situation is dramatically different if we consider a binomial model where the probability parameter tends to zero. Explicitly, for each n let $p(n) \in (0, 1)$. Now assume each integer in $\{0, \ldots, n - 1\}$ is chosen with probability $p(n)$. If $p(n)$ decays to zero with n, then Hegarty and Miller [HM1] proved that with probability tending to 1 a randomly chosen set is difference-dominated. See [ILMZ] for a survey of results in the field.

Throughout this paper we use the following notations:

- $m \cdot A = \{m \cdot a : a \in A\}$.

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

- $|A|$ is the number of elements in A.

- $mA = \underbrace{A + \cdots + A}_{m \text{ times}}$ if $m \geq 1$ (if $m = 0$ we define $0A$ to be the empty set).

- $-A = \{-a : a \in A\}$, and if $m \geq 0$ then $-mA = -(mA)$; note that if $m, n \geq 0$ then $mA + nA = (m + n)A$; however, $mA - nA \neq (m - n)A$.

- $[a, b] = \{a, a + 1, \ldots, b - 1, b\}$.

The purpose of this article is to generalize the positive percentage and explicit constructions of MSTD sets. Two natural questions, which motivated much of this work, are

(1) Given non-negative integers s_1, d_1, s_2, d_2 with $s_1 + d_1 = s_2 + d_2 \geq 2$, can we find a set A with $|s_1 A - d_1 A| > |s_2 A - d_2 A|$, and if so, does this occur a positive percentage of the time?

(2) We say a set is k-generational if A, $A+A$, ..., kA are all sum-dominant. Do k-generational sets exist, and if so, do they occur a positive percentage of the time? Is there a set that is k-generational for all k?

The answer to the first question is yes, and in fact the result can be generalized. When $s_1+d_1 = 2$, the only possible sets are essentially $A + A$ and $A - A$, as $-A - A$ is just the negation of $A + A$. When $s_1 + d_1 = 3$, again there are essentially just two possibilities, $A + A + A$ and $A + A - A$, since $A - A - A = -(A + A - A)$ and thus we might as well assume $s_i \geq d_i$. New behavior emerges once the sum is at least 4. In that case, we have $A + A + A + A$, $A + A + A - A$ and $A + A - A - A$. One of our main results is that all possible orderings of these three sets happen a positive percentage of the time. This generalizes and improves results from [MOS], where large families were found with $|A + A + A| > |A + A - A|$ and large families could be found for more general binary comparisons *if* one such set could be found.

For the second question, brute force numerical explorations could not find such sets. This is not surprising, as such sets are expected to be rare (simple heuristics imply that the percentage of such sets is at most 10^{-9}, and quite likely much less). Generalizing our construction for the first problem, we find a positive percentage of sets are k-generational for any k; further, no set can be k-generational for all k.

We now state our main results and give a sketch of the proofs.

Theorem 1.1. *Let s_1, d_1, s_2, d_2 be non-negative integers such that $\{s_1, d_1\} \neq \{s_2, d_2\}$.*

(1) *There exists a finite, non-empty set A of non-negative integers such that $|s_1 A - d_1 A| > |s_2 A - d_2 A|$.*

(2) *A positive percentage of finite subsets A of non-negative integers satisfy $|s_1 A - d_1 A| > |s_2 A - d_2 A|$. Explicitly, there is a constant $c(s_1, d_1, s_2, d_2) > 0$ such that the number of subsets A of $\{0, 1, \ldots, n-1\}$ satisfying $|s_1 A - d_1 A| > |s_2 A - d_2 A|$ is at least $c(s_1, d_1, s_2, d_2)2^n$ as $n \to \infty$.*

Remark 1.2. Sketch of the proof: *The difficulty is finding one such set; after such a set has been found, we can modify the method of Martin and O'Bryant [MO] to obtain a positive percentage. To create such a set A, we decompose A into its left and right parts, denoted L and R. We pick L and R to be almost symmetric, but we have R slightly longer than L. Next, note that the left (resp. right) fringe of $xA - yA$ is given by $xL - yR$ (resp. $yL - xR$). Because of the near-symmetry of L and R, the fringes of $xA - yA$ will have similar structure for different values of x, y. However, because R is longer than L, the total length of a fringe depends on the number of copies of R, L.*

In Figures 1 and 2, we exhibit a set A where $|2A+2A| > |2A-2A|$. Figure 1 shows $A+A+A+A$, while Figure 2 shows $A + A - A - A$. Notice that in $A + A + A + A$, the right fringe intersects with the middle, which fills in all the gaps. The left fringe, on the other hand, grows too slowly to completely intersect with the middle, and is left with one gap.

In $A + A - A - A$, the left fringe is given by $L + L - R - R$, which has a length between $L + L + L + L$ and $R + R + R + R$. This is not quite long enough to intersect with the middle. Similarly, the right fringe is given by $R + R - L - L$, which is once again too short. Therefore $A + A - A - A$ is missing two elements.

Theorem 1.1 can be generalized to obtain the following.

Theorem 1.3 (Arbitrary Differences). *Let a, b, c, d be non-negative integers such that $a > b, c, d$ and $a + b = c + d = q$. If $c \neq d$, then for any non-negative integers m, ℓ such that $\ell \leq 2m$ and all sufficiently large n, there exists $A \subseteq [0, n]$ such that $|aA - bA| = qn + 1 - m$ and*
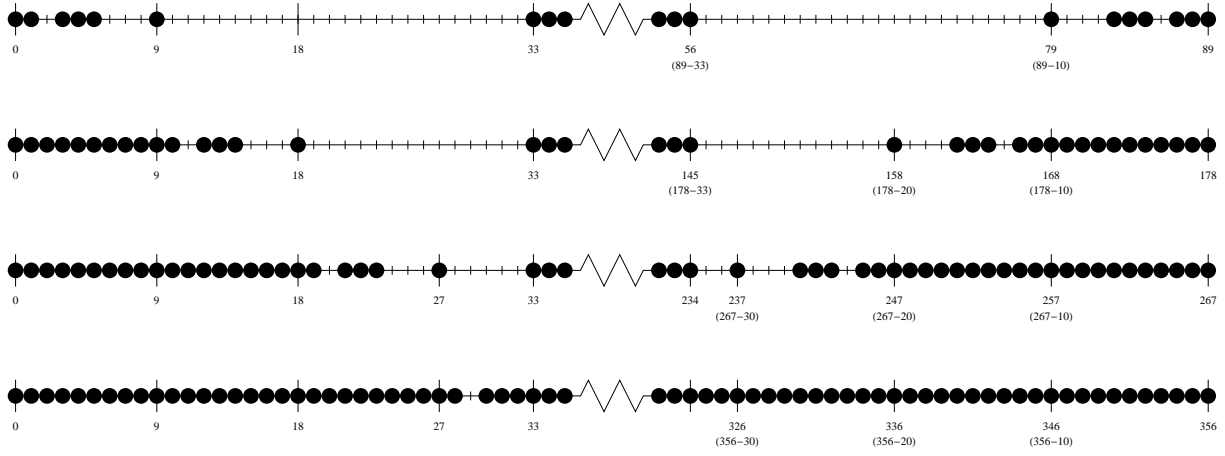
FIGURE 1. A, $A + A$, $A + A + A$, and $A + A + A + A$. The sawtooth means all elements are present in that range.
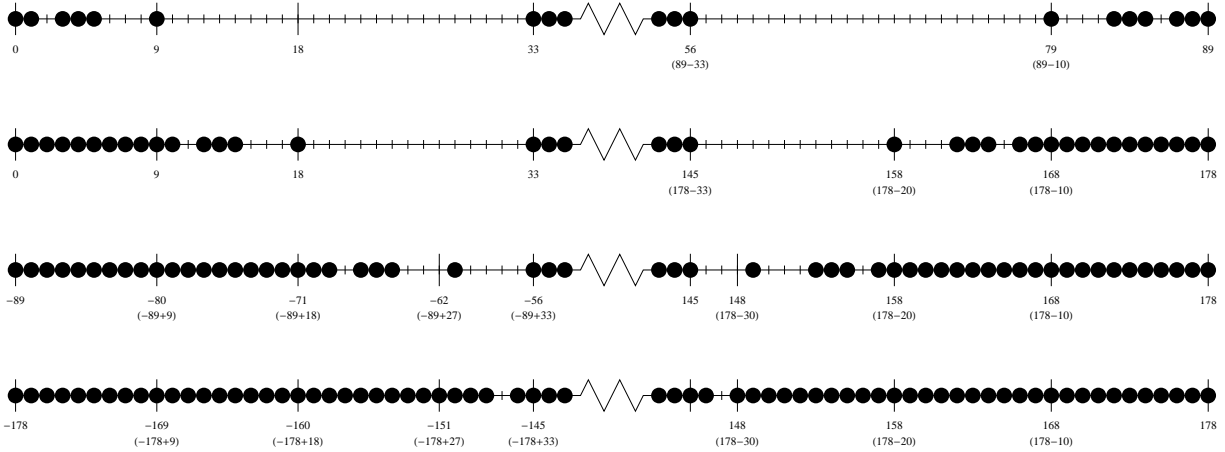

FIGURE 2. A, $A + A$, $A + A - A$, and $A + A - A - A$.

$|cA - dA| = qn + 1 - \ell$. If $c = d$, then the statement holds with the additional condition that ℓ is even.

The next theorem constructs chains of Generalized MSTD sets. We start at $k = 2$ below as there is essentially only one possibility when $k = 1$ (namely the sets A and $-A$, which must have the same cardinality).

Theorem 1.4 (Chains of Generalized MSTD Sets). *Let x_j, y_j, w_j, z_j be finite sequences of non-negative integers of length k such that $x_j + y_j = w_j + z_j = j$, and $\{x_j, y_j\} \neq \{w_j, z_j\}$ for every $2 \leq j \leq k$. A positive percentage of sets A satisfy $|x_j A - y_j A| > |w_j A - z_j A|$ for every $2 \leq j \leq k$.*

Theorem 1.5 (Simultaneous Comparisons). *Given finite, non-negative sequences of length $n \leq \left\lfloor \frac{k}{2} \right\rfloor + 1$ called s_j, d_j such that $s_j + d_j = k$ for all $1 \leq j \leq k$ and $\{s_j, d_j\} \neq \{s_i, d_i\}$ whenever $j \neq i$, there exists a set A such that $|s_n A - d_n A| > \cdots > |s_1 A - d_1 A|$*

Remark 1.6. *The bound $n \leq \lfloor \frac{k}{2} \rfloor + 1$ in the above theorem is completely artificial, as the condition $\{s_j, d_j\} \neq \{s_i, d_i\}$ is impossible for $n > \lfloor \frac{k}{2} \rfloor + 1$.*

Remark 1.7. *It is possible to combine Theorems 1.4 and 1.5 to obtain a set A that satisfies the criteria in Theorem 1.5 over many iterations of sums/differences.*

From Theorem 1.4 we deduce

Corollary 1.8 (k-Generational Sets)**.**
   (1) *For each k, there exists a k-generational set. That is, for each k, there exists a set A such that $|cA + cA| > |cA - cA|$ for all $1 \leq c \leq k$.*
   (2) *For each k, a positive percentage of sets are k-generational.*
   (3) *There is no set which is k-generational for all k.*

The paper is organized as follows. In §2 we explicitly construct one set with the properties in Theorem 1.1, obtaining only existence and not a positive percentage. For completeness we provide most of the verifications; the reader willing to accept their existence can move on to §3, where we generalize the method of Martin and O'Bryant to improve our results from the existence of one set to a positive percentage, completing the proof of Theorem 1.1. As the proof of arbitrary differences (Theorem 1.3) is not needed for the remaining results and is somewhat long and technical, we give it in Appendix A.

Section 4 contains a few lemmas required to construct the sets in Theorems 1.4 and 1.5. Once again, a reader uninterested in technical constructions may skip this section and proceed to §5. We discuss k-generational sets (and related problems) in §5, proving Theorem 1.4 and Corollary 1.8 as well as results about the limiting behavior of $|kA|$ and $|kA - kA|$ as k grows, improving earlier results of Nathanson [Na1]. We conclude in Section 6 with a proof of Theorem 1.5.

2. GENERALIZED MSTD SETS

The goal of this section is, given $k \in \mathbb{N}$ and integers with $s_1 + d_1 = s_2 + d_2 = k$ and $\{s_1, d_1\} \neq \{s_2, d_2\}$, to explicitly construct a set A such that $|s_1 A - d_1 A| = |s_2 A - d_2 A| + 1$. The existence of these sets is the key ingredient in the proof of Theorem 1.4; the proof of that theorem requires us to generalize these constructions slightly, and then modify the arguments of Martin and O'Bryant to obtain a positive percentage by showing a positive percentage of middles may be added to our set. The reader uninterested in the technical construction should skim the sketch of the method in Remark 1.2 and then continue in §3.

Let $\ell = 2k + 1$, $r = 2k + 2$, and consider the sets

$$
\begin{aligned}
L &= \{0, 1, 3, 4, \ldots, k-1, k, k+1, 2k+1\} \\
&= \{0, \ell - 2k, l - 2k - 2, \ell - 2k - 3, \ldots, \ell - k - 1, \ell - k, \ell\} \\
&= [0, \ell] \backslash (\{2\} \cup [k+2, 2k]) \\
&= [0, \ell] \backslash (\{2\} \cup [\ell - k + 1, \ell - 1]) \\
R &= \{0, 1, 2, 4, 5, \ldots, k, k+1, k+2, 2k+2\} \\
&= [0, r] \backslash (\{3\} \cup [k+3, 2k+1]).
\end{aligned}
\tag{2.1}
$$

We begin with a technical lemma. This lemma states that for any $x, y \in \mathbb{N}$, the basic structure of $xL + yR$ is the same as that of the original sets. Basically, $xL + yR$ is always missing the first k elements below the maximum, as well as the singleton element $2k - 1$ away from the maximum. Even more, it is missing no other elements.

Lemma 2.1. *For all $x, y \in \mathbb{N}$,*

$$xL + yR = [0, x\ell + yr] \setminus ([x\ell + yr - k + 1, x\ell + yr - 1] \cup \{x\ell + yr - 2k + 1\}). \qquad (2.2)$$

Proof. The proof is by double induction, first on x, then on y. As the proof of the base case $x = y = 1$ follows by a simple computation, we now assume the result for $xL + yR$ and prove it for $xL + (y + 1)R$.

We are interested in

$$([0, x\ell + yr] \setminus ([x\ell + yr - k + 1, x\ell + yr - 1] \cup \{x\ell + yr - 2k + 1\})) + R. \qquad (2.3)$$

We prove that this set contains the proper elements in several steps:

Claim 1: $[0, x\ell + yr] \subset xL + (y + 1)R$.
Proof: Clearly $x\ell + yr - 2k + 1 \in xL + (y + 1)R$, since $x\ell + yr - 2k \in xL + yR$ and $1 \in R$. Furthermore, $[x\ell + yr - k + 1, x\ell + yr - 1] \subset xL + (y + 1)R$, since $x\ell + yr - k - 4 \in xL + yR$, $x\ell + yr - k - 1 \in xL + yR$, and $[4, k + 2] \subset R$.

Claim 2: $x\ell + (y + 1)r - 2k + 1 = x\ell + yr + 3 \notin xL + (y + 1)R$.
Proof: This is equivalent to showing that $x\ell + yr + 3 - (xL + yR) \cap R = \emptyset$. This is true as $x\ell + yr + 3 - (xL + yR) \cap (\mathbb{N} \cup \{0\}) = \{3\}$, and $3 \notin R$.

Claim 3: $[x\ell + (y + 1)r - k, x\ell + (y + 1)r - 1] \cap (xL + (y + 1)R) = \emptyset$.
Proof: This is the same as showing that $[x\ell + yr + k + 2, k\ell + yr + 2k + 1] \cap (xL + (y + 1)R) = \emptyset$. I.e., we want to show that $\max(xL + (y + 1)R) - a \notin xL + (y + 1)R$ for every $1 \le a \le k$. This is true because $\max(xL + yR) - a \notin xL + yR$ and $\max(R) - a \notin R$ for every $1 \le a \le k$. Therefore the same will be true of $xL + yR + R = xL + (y + 1)R$.

Claim 4: All other elements in $[x\ell + yr, x\ell + (y + 1)r]$ are in $xL + (y + 1)R$.
Proof: This is true because each of those elements can be written as $x\ell + yr + c$ for some $c \in R$.

We have proved the inductive step for y; we omit the proof of the inductive step for x, since it is almost exactly the same as the above proof. $\qquad \square$

With the technical lemma proved, we can construct a set as in Theorem 1.1.

Theorem 2.2. *Suppose $k \in \mathbb{N}$, and $s_1 + d_1 = s_2 + d_2 = k$. Further suppose that $\{s_1, d_1\} \ne \{s_2, d_2\}$. There exists a set A such that $|s_1 A - d_1 A| = |s_2 A - d_2 A| + 1$.*

For example, the set

$$A = \{0, 1, 3, 4, 5, 9, 33, 34, 35, 50, 54, 55, 56, 58, 59, 60\}$$

has the property that

$$|A + A + A + A| > |A + A + A - A|.$$

Proof. Because $|xA - yA| = |yA - xA|$, we can assume that $s_1 \ge d_1$ and $s_2 \ge d_2$. Therefore we have either $s_1 > s_2 \ge d_2 > d_1$, or $s_2 > s_1 \ge d_1 > d_2$. We first treat the case when $s_1 > s_2$.

Case 1: $s_1 > s_2$: Take L, R, ℓ, r as in construction from Lemma 2.1, and choose $n > 4(kr - 2k + 1)$. Define

$$
\begin{aligned}
M &= [kr - 2k + 1 - d_1, kn - (kr - 2k + 1 - d_1)] \\
A &= L \cup M \cup (n - R).
\end{aligned}
\tag{2.4}
$$

To prove this, we first show that the middle of $s_1 A - d_1 A$ is full, and then we examine the fringes. We have $[(kr - 2k + 1 - d_1) - d_1 n, s_1 n - (kr - 2k + 1 - d_1)] \subset s_1 A - d_1 A$. To prove this, note that M is sufficiently large such that $(M + L) \cup (M + n - R)$ is the entire interval $[\min(M), n + \max(M)]$. Therefore, $[\min(M), n + \max(M)] \subset A + A$. Similarly, we get that $[\min(M) - n, \max(M)] \subset A - A$. The same idea shows that $(M + L) \cup (M + n - R)$ is sufficiently large such that

$$
[\min(M), 2n + \max(M)] \subset (M + L + L) \cup (M + M + L) \cup \cdots \cup (M + n - R + n - R).
$$

By induction, $s_1 A - d_1 A$ will contain $[\min(M), (k-1)n + \max(M)]$.

We first look at the left fringe of $s_1 A - d_1 A$, this is (up to translation) $sL + dR \cap [0, kr - 2k - d_1]$. Note that $kr - 2k - d_1 = s_1 \ell + d_1 r - 2k - d_1 + s_1$. Therefore, using Lemma 2.1, we get that

$$
s_1 L + d_1 R \cap [0, s_1 \ell + d_1 r - 2k - d_1 + s_1] = [0, kr - 2k - d_1] \setminus \{s_1 \ell + d_1 r - 2k + 1\}
\tag{2.5}
$$

(this is because $s_1 > d_1$). Next we look at the right fringe. This is (up to translation and a minus sign) $d_1 L + s_1 R \cap [0, kr - 2k - d_1]$, which is the same as

$$
d_1 L + s_1 R \cap [0, d_1 \ell + s_1 r - 2k - d_1 + d_1] = d_1 L + s_1 R \cap [0, d_1 \ell + s_1 r - 2k].
\tag{2.6}
$$

Using Lemma 2.1, this is just $[0, d_1 \ell + s_1 r - 2k]$ (i.e., the entire interval). Therefore $s_1 A - d_1 A$ is missing one element.

Next, we look at the left fringe of $s_2 A - d_2 A$. Once again, this is (up to translation) $s_2 L + d_2 R \cap [0, kr - sk + 1 - d_1]$. This can be rewritten as $s_2 L + d_2 R \cap [0, s_2 \ell + d_2 r - 2k - d_1 + s_2]$. Since we have that $s_2 > d_1$, we get

$$
s_2 L + d_2 R \cap [0, s_2 \ell + d_2 r - 2k - d_1 + s_2] = [0, s_2 \ell + d_2 r - 2k - d_1 + s_2] \setminus \{s_2 \ell + d_2 r - 2k + 1\}.
\tag{2.7}
$$

Therefore the left fringe is missing one element. Now we look at the right fringe. This is (up to translation and a minus sign) $d_2 L + s_2 R \cap [0, kr - 2k + 1 - d_1]$. This is the same as $d_2 L + s_2 R \cap [0, d_2 \ell + s_2 r - 2k + 1 - d_1 + d_2]$. Now, because we have $d_2 > d_1$, this intersection is $[0, d_2 \ell + s_2 r - 2k + 1 - d_1 + d_2] \setminus \{d_2 \ell + s_2 r - 2k + 1\}$. Therefore, the right fringe is missing one element. This means that $kn = |s_1 A - d_1 A| > |s_2 - d_2 A| = kn - 1$.

Case 2: $s_2 > s_1$: As $s_2 > s_1$ we have $d_1 > d_2$. Define

$$
\begin{aligned}
M &= [kr - 2k + 1 - s_1, n - (kr - 2k + 1 - s_1)] \\
A &= L \cup M \cup (n - R).
\end{aligned}
\tag{2.8}
$$

We have

$$
\begin{aligned}
s_1 L + d_1 R \cap [0, kr - 2k + 1 - s_1] &= s_1 L + d_1 R \cap [0, s_1 \ell + d_1 r - 2k + 1 + s_1 - s_1] \\
&= [0, s_1 \ell + d_1 r - 2k + 1],
\end{aligned}
\tag{2.9}
$$

so the left fringe is missing no elements. Furthermore

$$
\begin{aligned}
d_1 L + s_1 R \cap [0, kr - 2k + 1 - s_1] &= d_1 L + s_1 R \cap [0, d_1 \ell + s_1 r - 2k + 1 - s_1 + d_1] \\
&= [0, d_1 \ell + s_1 r - 2k + 1 - s_1 + d_1].
\end{aligned}
\tag{2.10}
$$

The last step is true because $s_1 \geq d_1$. Therefore, $s_1A - d_1A$ misses no elements.

Next, we look at:

$$
\begin{aligned}
s_2L + d_2R \cap [0, kr - 2k + 1 - s_1] &= s_2L + d_2R \cap [0, s_2\ell + d_2r - 2k + 1 - s_1 + s_2] \\
&= [0, s_2\ell + d_2r - 2k + 1 - s_1 + s_2] \setminus \{s_2\ell + d_2r - 2k + 1\},
\end{aligned}
$$
(2.11)

which is true because $s_2 > s_1$. This is enough to show that $|s_1A - d_1A| > |s_2A - d_2A|$, but we will go slightly further and show that $|s_1A - d_1A| = |s_2A - d_2A| + 1$. To do this, we look at the right fringe of $s_2A - d_2A$. As $s_1 > d_2$, we have

$$
\begin{aligned}
d_2L + s_2R \cap [0, kr - 2k + 1 - s_1] &= d_2L + s_2R \cap [0, d_2\ell + s_2r - 2k + 1 - s_1 + d_2] \\
&= [0, d_2\ell + s_2r - 2k + 1 - s_1 + d_2],
\end{aligned}
$$
(2.12)

which completes the proof. $\qquad\square$

Although it doesn't matter for our current purposes, the following lemma will be important later. Each of the sets constructed above is sum-difference balanced both before and after the critical point. More formally, we have the following.

Lemma 2.3. *In all the sets A defined in the proof of Theorem 2.2,*

$$
|s_1A - d_1A| = |s_2A - d_2A|
$$
(2.13)

for any $s_1 + d_1 = s_2 + d_2$ such that $s_1 + d_1 \neq k$.

Proof. In every one of the constructions above, $sA - dA$ contains all possible numbers whenever $s + d > k$, so it only remains to show this fact when $s + d < k$.

This essentially follows from Lemma 2.1. Both $s_1A - d_1A$ and $s_2A - d_2A$ contain the same middle (up to translation), so it is enough to analyze the fringes. When $s_1 + d_1 < k$, these fringes do not intersect the middle, so it suffices to show that

$$
|s_1L + d_1R| + |d_1L + s_1R| = |s_2L + d_2R| + |d_2L + s_2R|.
$$
(2.14)

Using Lemma 2.1, we know that $|sL + dR| = s\ell + dr - k$. Therefore, it is enough to show that

$$
s_1\ell + d_1r - k + d_1\ell + s_1r - k = s_2\ell + d_1r - k + d_2\ell + s_2r - k.
$$
(2.15)

This equation is the same as

$$
(s_1 + d_1)(\ell + r) - 2k = (s_2 + d_2)(\ell + r) - 2k;
$$
(2.16)

as $s_1 + d_1 = s_2 + d_2$, the above is true, which completes the proof. $\qquad\square$

3. POSITIVE PERCENTAGES

We now give a proof of Part 2 of Theorem 1.1.

Lemma 3.1. *Suppose there exists a finite set $A \subseteq \mathbb{Z}$ such that $|s_1A - d_1A| > |s_2A - d_2A|$, where $s_1 + d_1 = s_2 + d_2$. Further suppose that $s_1 \geq 2$. Then*

$$
\liminf_{n \to \infty} \frac{\#\{B \subseteq [0, n-1]; \, |s_1B - d_1B| > |s_2B - d_2B|\}}{2^n} > 0;
$$
(3.1)

in other words, a positive percentage of subsets have this structure.

Remark 3.2. *Note that the assumption $s_1 \geq 2$ only rules out the case $s_1 = d_1 = 1$, since we can always replace A with $-A$ without affecting the cardinalities. This case has already been dealt with in detail in* [MO].

Proof. By translation, we can assume that $A \subseteq [0, n-1]$, with $0, n-1 \in A$.

Choose some $m \geq 4(s_1 + d_1)n$, and define

$$
\begin{aligned}
L &= A \cup [(s_1 + d_1)n, 2(s_1 + d_1)n - 1] \\
U &= [m - 2(s_1 + d_1)n, m - (s_1 + d_1)n - 1] \cup (A + (m - n)). \qquad (3.2)
\end{aligned}
$$

Informally, our fringes consist of a copy of A at the far end, then an interval of size $(s_1 + d_1)n$ which is located $(s_1 + d_1)n$ away from the edge.

Furthermore, define $l = u = 2(s_1 + d_1)n$. Next, we note three things:

(1) $[\frac{l}{2}, 2l - 2] \subset L + L$
(2) $[2m - 2u, 2m - \frac{u}{2} - 1] \subset U + U$
(3) $[m - u, m + l - 2] \subset L + U$.

Each of these claims follows from $[\frac{l}{2}, l - 1] \subset L$ and $[m - u, m - \frac{u}{2} - 1] \subset U$, as well as the fact that $0 \in L$, and $m - 1 \in U$

Next, suppose that $B \subset [0, m-1]$ is a set with fringes L, U. Based on Proposition 8 of [MO], the probability that

$$
[2l - 1, m - u - 1] \cup [m + l - 1, 2m - 2u - 1] \subseteq B + B \qquad (3.3)
$$

is at least

$$
1 - 6(2^{-|L|} + 2^{-|U|}) > 1 - 6(2^{-(s_1+d_1)n} + 2^{-(s_1+d_1)n}) = 1 - 6 \cdot 2^{-(s_1+d_1)n+1} = c. \qquad (3.4)
$$

Therefore, if B is a set as above, then with positive probability (that is independent of m),

$$
\left[\frac{l}{2}, 2m - \frac{u}{2} - 1 \right] \subset B + B. \qquad (3.5)
$$

Essentially, we have chosen the fringes of B such that with a positive probability that is independent of m, the entire middle (here middle means everything besides the $(s_1+d_1)n$ elements on each side) of $B + B$ will be full. However, this means that the entire middle of $s_1 B - d_1 B$ will also be full. Therefore, it only remains to check the fringes of $s_1 B - d_1 B$. Each of these fringes is just a copy of $s_1 A - d_1 A$. Therefore, $s_1 B - d_1 B$ consists of a copy on $s_1 A - d_1 A$ on each fringe, and everything in between.

To show that $|s_1 B - d_1 B| > |s_2 B - d_2 B|$, it is sufficient to note that for the exact same reasons, the fringes of $s_2 B - d_2 B$ will just be copies of $s_2 A - d_2 A$. Therefore, since $s_1 B - d_1 B$ contains strictly more elements on the fringe, as well as everything not on the fringe, it must have more elements that $s_2 B - d_2 B$.

As for the probability, we have made $4(s_1 + d_1)n$ choices for the fringes of B, and making sure the middle is full accounts for a factor of c. So the probability that $|s_1 B - d_1 B| > |s_2 B - d_2 B|$ is at least $c2^{-4(s_1+d_1)n}$. $\qquad \square$

4. TECHNICAL CONSTRUCTIONS

4.1. **Multiple Fringes.** In order to prove Theorems 1.4 and 1.5, we first construct a very well behaved set. Then, in Section §5 we will use the base expansion method to create a set that combines many different copies of the below set.

Lemma 4.1. *Suppose $k \in \mathbb{N}$, and $s, d \in \mathbb{N} \cup \{0\}$ such that $s + d = k$ and $s \geq d$. There exists a set $A \subset \mathbb{N} \cup \{0\}$ such that if $s' + d' = k$, $s' \neq s$, and $s' \geq d'$, then $|sA - dA| = |s'A - d'A| + 1$.*

Proof. Have L, R as in (2.1):

$$
\begin{aligned}
L &= \{0, 1, 3, 4, \ldots, k - 1, k, k + 1, 2k + 1\} \\
&= \{2k + 1\} \cup [0, k + 1] \setminus \{2\} \\
R &= \{0, 1, 2, 4, 5, \ldots, k, k + 1, k + 2, 2k + 2\} \\
&= \{2k + 2\} \cup [0, k + 2] \setminus \{3\}.
\end{aligned}
\tag{4.1}
$$

Set $\ell = 2k + 1$ and $r = 2k + 2$ (just as before).

Before we give the main proof, there are two exceptional cases to consider. We have already proved the case where $d = 0$. If $s = d$, then choose $n > 2(kr - 2k + 1 - d)$ and take

$$
A = L \cup [kr - 2k + 1 - d, n - (kr - 2k + 1 - d)] \cup (n - R).
\tag{4.2}
$$

For this set, $sA - dA$ misses no elements, and $s'A - d'A$ misses one element for any choice of s', d' that satisfies the above. The proof of this statement is essentially the same as found in the above proofs.

Now we assume that $s, d \geq 1$ and that $s > d$. Set

$$
\begin{aligned}
A &= L \cup (L + kr - 2k + 1 - d) \cup [2kr - 4k + 2 - d - s, n - (2kr - 4k + 2 - d - s)] \\
&\quad \cup (n - (2kr - 2k + 1 - d) - R) \cup (n - R).
\end{aligned}
\tag{4.3}
$$

Essentially, A consists of an outer fringe, and inner fringe, and a full middle. Both the outer fringe and the inner fringe have the same structure (they are both made up of L and R). For simplicity, we write this as

$$
A = L_1 \cup L_2 \cup M \cup (n - R_2) \cup (n - R_1),
\tag{4.4}
$$

where $L_1 = L$, $L_2 = L + kr - 2k + 1 - d$, $R_1 = R$, $R_2 = R + kr - 2k + 1 - d$, and $M = [2kr - 4k + 2 - d - s, n - (2kr + 2 - d - s)]$.

Note first that because n is sufficiently large, $sA - dA$ and $s'A - d'A$ will contian the entire middle (the logic for this is the same as above). Further note that the fringes of $sA - dA$ are

$$
(sL_1 - d(n - R_1)) \cup (L_2 + (s - 1)L_1 - d(n - R_1))
\tag{4.5}
$$

and

$$
(s(n - R_1) - dL_1) \cup ((n - R_2) + (s - 1)(n - R_1) - dL_1.
\tag{4.6}
$$

This is because all other sums/differences fall in the large and full middle. As usual, we will translate these sets (and possibly multiply by -1), and look at

$$
(sL_1 + dR_1) \cup (L_2 + (s - 1)L_1 + dR_1) \text{ and } (sR_1 + dL_1) \cup (R_2 + (s - 1)R_1 + dL_1).
\tag{4.7}
$$

We analyze each of these four fringes one at a time.

(1) $sL_1 + dR_1$

First note that $L_2 + (s - 1)L_1 + dR_1$ contains the interval $[kr - k + 1 - s, kr - s] = [s\ell + dr - k + 1, s\ell + dr - 1]$. This means that of the potential missing elements in $sL_1 + dR_1$, all except for $kr - 2k + 1$ can be found in $L_2 + (s-1)L_1 + dR_1$. Essentially, we are interested in $sL_1 + dR_1 \cap [0, kr - 2k - d]$. This is $sL_1 + dR_1 \cap [0, s\ell + dr - 2k - d + s]$, which is just $[0, kr - 2k - d] \setminus \{s\ell + dr - 2k + 1\}$ (because $s > d$). Therefore the outer left fringe is missing one element.

(2) $L_2 + (s-1)L_1 + dR_1$

Part of this set will intersect with the full middle, so we are really only interested in $L_2 + (s-1)L_1 + dR_1 \cap [kr - 2k + 1 - d, 2kr - 4k + 1 - d - s]$. After translation, this is the same as $sL_1 + dR_1 \cap [0, kr - 2k - s]$. This is $sL_1 + dR_1 \cap [0, s\ell + dr - 2k]$, which is just $[0, s\ell + dr - 2k]$. Therefore the inner left fringe is missing no elements.

(3) $sR_1 + dL_1$

Similar to the above case, this fringe intersects with $R_2 + (s-1)R_1 + dL_1$. Therefore, we are only interested in $sR_1 + dL_1 \cap [0, kr - 2k - d]$. This is the same as $sR_1 + dL_1 \cap [0, sr + d\ell - 2k]$, which is just $[0, sr + d\ell - 2k]$. Therefore the outer right fringe is missing no elements.

(4) $R_2 + (s-1)R_1 + dL_1$

Because of the intersection with the middle, we are only interested in $R_2 + (s-1)R_1 + dL_1 \cap [kr - 2k + 1 - d, 2kr - 4k + 1 - d - s]$. After translation, this is just $sR_1 + dL_1 \cap [0, kr - 2k - s] = sR_1 + dL_1 \cap [0, sr + d\ell - 2k - s + d]$. Since $s > d$, this is just the entire interval $[0, kr - 2k - s]$. Therefore the inner right fringe is missing no elements.

Next, we run through the same analysis with $s'A - d'A$. We split this into two cases. First, if $s' > s$, then:

(1) $s'L_1 + d'R_1$

Just as above, we are interested in $s'L_1 + d'R_1 \cap [0, kr - 2k - d]$. This is $s'L_1 + d'R_1 \cap [0, s'\ell + d'r - 2k - d + s'] = [0, kr - 2k - d]\setminus\{s'\ell + dr - 2k + 1\}$. Therefore the outer left fringe is missing one element.

(2) $L_2 + (s' - 1)L_1 + d'R_1$

Similar to above, this is the same thing as $s'L_1 + d'R_1 \cap [0, kr - 2k - s]$. This is just $s'L_1 + d'R_1 \cap [0, s'\ell + d'r - 2k - s + s']$. Since $s' > s$, this is $[0, s'\ell + d'r - 2k - s + s']\setminus\{s'\ell + d'r - 2k + 1\}$.

Therefore $s'A - d'A$ is missing at least two elements. Only slightly more work shows that the set is missing exactly two elements, which means that $|sA - dA| = |s'A - d'A| + 1$.

Next, we assume that $s > s'$. In this case we have $s > s' \geq d' > d$. If we perform the same analysis as above, we will find that $sL_1 + dR_1$ and $sR_1 + dL_1$ are both missing one element. Therefore we get that $|sA - dA| = |s'A - d'A| + 1$. $\square$

Corollary 4.2. *Suppose that $k \in \mathbb{N}$, and $s, d \in \mathbb{N} \cup \{0\}$ such that $s + d = k$. Then there exists a set $A \subseteq \mathbb{N} \cup \{0\}$ such that if $s' + d' = k$, $\{s, d\} \neq \{s', d'\}$, then $|sA - dA| = |s'A - d'A| + 1$.*

Proof. This follows from the above if we just note that $|sA - dA| = |-(sA - dA)|$. $\square$

4.2. **Base expansion.** We end this section with a quick proof of the base expansion method for creating new sets. Base expansion allows us to use multiple copies of the well-behaved sets constructed in Lemma 4.1 to create the sets in Theorems 1.4 and 1.5.

Lemma 4.3. *Fix a positive integer k. Let $A, B \subset \mathbb{N} \cup \{0\}$ and choose $m > k \cdot \max(A)$. Let $C = A + m \cdot B$ (where $m \cdot B$ is the usual scalar multiplication). Then $|sC - dC| = |sA - dA| \cdot |sB - dB|$ whenever $s + d \leq k$.*

Proof. Note that each element of $A + mB$ can be written uniquely as $a + mb$ for some $a \in A$, $b \in B$. This is true because if $a_1 + mb_1 = a_2 + mb_2$, then $a_1 - a_2 = m(b_2 - b_1)$. Because we chose m sufficiently large, this is only possible when $b_1 = b_2$, in which case $a_1 - a_2 = 0$. Therefore $|C| = |A||B|$.

Furthermore, each element of $C \pm C$ can be written uniquely as $a' \pm mb'$, where $a' \in A \pm A$ and $b' \in B \pm B$. As proof, assume $a_1 \pm mb_1 = a_2 \pm mb_2$ for some $a_1, a_2 \in A \pm A$, and $b_1, b_2 \in B \pm B$. This means $a_1 - a_2 = \mp m(b_2 - b_1)$, and this is only possible when $a_1 = a_2$, and $b_1 = b_2$. Therefore, $|C \pm C| = |A \pm A| \, |B \pm B|$. A similar proof shows this fact for any $s + d \leq k$.   □

In fact, base expansion works in more generality:

Lemma 4.4. *Fix a positive integer k. Say that $A_1, \ldots, A_k \subset \mathbb{N} \cup \{0\}$. Choose some $m > k \cdot \max(\{a; a \in A_k \text{ for some } k\})$. Let $C = A_1 + m \cdot A_2 + \cdots + m^{k-1} \cdot A_k$ (where $m \cdot A_j$ is the usual scalar multiplication). Then $|sC - dC| = \prod_{j=1}^{k} |sA_j - dA_j|$ whenever $s + d \leq k$.*

Proof. This can be proved using induction and the previous lemma.   □

5. k-GENERATIONAL SETS

5.1. **Proof of Theorem 1.4.** We now have the tools required to prove our results about chains.

Theorem 1.4. For each j, choose a set A_j such that $|x_j A_j - y_j A_j| > |w_j A_j - z_j A_j|$, and $|s_1 A_j - d_1 A_j| = |s_2 A_j - d_2 A_j|$ whenever $s_1 + d_1 = s_2 + d_2 \neq j$. We know such a set exists, because of Theorem 1.1 and Lemma 2.3. Next, choose some $m > k \cdot \max(\{a \in A_j; 1 \leq j \leq k\})$. Define $A = A_1 + mA_2 + m^2 A_3 + \cdots + m^{k-1} A_k$. We have that for each $2 \leq j \leq k$

$$
\begin{aligned}
|x_j A - y_j A| &= \prod_{i=1}^{k} |x_j A_i - y_j A_i| \\
&= |x_j A_j - y_j A_j| \cdot \prod_{i \neq j} |x_j A_i - y_j A_i| \\
&= |x_j A_j - y_j A_j| \cdot \prod_{i \neq j} |w_j A_i - z_j A_i| \\
&> |w_j A_j - z_j A_j| \cdot \prod_{i \neq j} |w_j A_i - z_j A_i| \\
&= |w_j A - z_j A|.
\end{aligned}
\tag{5.1}
$$

          □

Most of Corollary 1.8 now follows automatically. The existence of a k-generational set is proven by the above theorem, and proving that a positive percentage of sets have this property only requires a slight modification of the work done in §3. It only remains to that no set can be k-generational for all k by analyzing the limiting behavior of $|kA|$ and $|kA - kA|$.

5.2. **Limiting behavior of $|kA|$ and $|kA - kA|$.** Before proving Corollary 1.8(3), we first prove two useful lemmas.

Lemma 5.1. *Let $A = \{a_1, a_2, \ldots, a_m\} \subset [0, n-1]$ be a set of integers where $a_1 < a_2 < \cdots < a_m$ and let $s = \gcd(a_1, a_2, \ldots, a_m)$. Then there exists an integer N such that for $k \geq N$ we have $|kA| = \frac{k(a_m - a_1)}{s} - C$ where C is a constant and N is bounded above by $\frac{a_m - a_1}{s}$.*

Proof. It suffices to show that a set of the form $\{0, a_1, \ldots, a_m\}$ with $\gcd(a_1, \ldots, a_m) = 1$ has the claimed properties (because of translating and rescaling).

Let $A = \{0, a_1, \ldots, a_m\}$, with $\gcd(a_1, \ldots, a_m) = 1$. We first show that in $a_1 A$ (which is the sum of a_1 copies of A) there are elements of each congruence class of a_1. Consider the set $B =$

$\{0, a_2', \ldots, a_m'\}$ where $a_i' = a_i \bmod a_1$. Clearly we also have $\gcd(a_2', \ldots, a_m') = 1$. Thus they generate the entire set $[0, a_1]$. It is clear that the largest number of times required to add B to itself is a_1 since the greatest order of any element in the set is a_1. This proves the claim.

Now consider $a_m A$, in particular we consider the set $L = a_m A \cap [0, a_1 a_m]$. We show that $kA \cap [0, a_1 a_m] = L$ for $k \geq a_m$. This is because a_1 is the smallest element in the set A, so elements that are less than $a_1 a_m$ can be written as $\sum_{i=1}^{m} \epsilon_i a_i$ where $\sum_i^m \epsilon_i \leq a_m$. We call L the stabilized left fringe of A.

We can apply the same idea to the set $a_m - A$ and show that the right fringe $R = kA \cap [(k-1)a_m, ka_m]$ is also stabilized (meaning that $ka_m - R$ stays the same for all $k \geq a_m$). Now we just need to show that for $k \geq a_m$ we have $kA \backslash (L \cup R)$ is completely filled. This can be shown by induction. With all the congruence classes of a_1, by brute force we can show that the middle part of $a_m A$ is completely filled. This serves as the base case of the induction. If $kA \backslash (L \cup R)$ is completely filled then kA contains the interval $[a_1 a_m, (k - a_m + a_{m-1})a_m]$. If we add a_m to this interval we will get the interval $[(k - a_m + a_{m-1})a_m, (k + 1 - a_m + a_{m-1})a_m]$. So in $(k+1)A$, we will have a completely filled middle $[a_1 a_m, (k + 1 - a_m + a_{m-1})a_m]$. This completes the proof. □

Lemma 5.2. *Let $A = \{a_1, a_2, \ldots, a_m\} \subset [0, n-1]$ be a set of integers where $a_1 < a_2 < \cdots < a_m$ and let $s = \gcd(a_1, a_2, \ldots, a_m)$. Then there exists an integer N such that for $k \geq N$ we have $|kA - kA| \geq |kA + kA|$ and N is bounded above by $\frac{2(a_m - a_1)}{s}$.*

Proof. Note that $kA \subset kA - kA$. This means that if $cA + cA$ has stable fringes and a full middle, then $2cA - 2cA$ will contain all those fringe elements (and maybe more) as well as the full middle. Therefore, if we choose $N = 2c$, then for any $k \geq N$, $|kA - kA| \geq |kA + kA|$ □

Corollary 1.8(3) now follows immediately; in other words, no set can be k-generational for all k. This significantly improves an earlier result of Nathanson [Na1], who proved that kA stabilizes by $k \geq a^2 m$, where a is the largest element of A and m is the largest gap between elements of A.

6. SIMULTANEOUS COMPARISON

In this section we prove that any ordering for a simultaneous comparison happens.

Proof of Theorem 1.5. We repeatedly use base expansion. For each $1 \leq j \leq n$, choose A_j such that $|s_j A_j - d_j A_j| = |sA - dA| + 1$ for every $s \neq \pm s_j$. Next, choose an $m > k \cdot \max(\{a; a \in A_j \text{ for some } 1 \leq j \leq n\})$. Let

$$A = A_1 + mA_2 + m^2 A_2 + \cdots + \underbrace{m^{\frac{j(j-1)}{2}} A_j + \cdots + m^{\frac{j(j-1)}{2}+j-1} A_j}_{j \text{ times}} + \cdots + m^{\frac{n(n-1)}{2}+n-1} A_n.$$

(6.1)

More simply, A is made of j copies of each A_j. Arguing as before (such as in Lemma 4.3), we find

$$|s_j A - d_j A| = \prod_i |s_j A_i - d_j A_i|^i.$$

(6.2)

Now, we have that $|s_j A_i - d_j A_i| = |s_\ell A_i - d_\ell A_i|$ whenever $\ell, j \neq i$. Furthermore, we have that $|s_i A_i - d_i A_i| = |s_j A_i - d_j A_i| + 1$ whenever $i \neq j$. Therefore, if we choose $j > \ell$, we obtain

$$
\begin{aligned}
|s_j A - d_j A| &= \prod_i |s_j A_i - d_j A_i|^i \\
&= |s_j A_j - d_j A_j|^j \cdot \prod_{i \neq j} |s_j A_i - d_j A_i|^i \\
&= (|s_\ell A_j - d_\ell A_j| + 1)^j \cdot (|s_\ell A_\ell - d_\ell A_\ell| - 1)^\ell \prod_{i \neq j, \ell} |s_\ell A_i - d_\ell A_i|^i \\
&= |s_\ell A_\ell - d_\ell A_\ell|^j \, |s_\ell A_j - d_\ell A_j|^\ell \prod_{i \neq j, \ell} |s_\ell A_i - d_\ell A_i|^i \\
&> |s_\ell A_\ell - d_\ell A_\ell|^\ell \, |s_\ell A_j - d_\ell A_j|^j \prod_{i \neq j, \ell} |s_\ell A_i - d_\ell A_i|^i \\
&= \prod_i |s_\ell A_i - d_\ell A_i|^i = |s_\ell A - d_\ell A| .
\end{aligned}
\tag{6.3}
$$

Informally, we have chosen the A_i such that $|s_i A_i - d_i A_i|$ is larger than all other possible combinations of sums and differences. Then we made $|s_2 A - d_2 A| > |s_1 A - d_1 A|$ by having more copies of A_2 than of A_1. Similarly, we made $|s_3 A - d_3 A| > |s_2 A - d_2 A|$ by having more copies of A_3 than of A_2. Following this process, we constructed a set A with the desired properties.

We have found an A such that $|s_n A - d_n A| > \cdots > |s_1 A - d_1 A|$, completing the proof. □

APPENDIX A. ARBITRARY DIFFERENCES

In this section we prove Theorem 1.3. Let

$$
A = L \cup [16km - 2m + 1, n - (16km - 2m + 1)] \cup (n - R) \tag{A.1}
$$

where

$$
\begin{aligned}
L &= [0, 4m] \cup [5m + 1, 6m] \cup \{8m\} \\
R &= (L + m/k) \cup [0, m/k - 1].
\end{aligned}
\tag{A.2}
$$

Note that the fringes L, R of this A are generalizations of the original fringes in (2.1). For example, this new L is obtained from the original L by extending the first gap of the original L to have length m. Also, note that this R is L shifted down by m/k, with the front filled in; this generalizes the original R in (2.1), where R is L shifted down only by 1.

We modify this A in several steps, each step bringing us closer to the full generality of Theorem 1.3. We first show that the above A has the property that $|kA + kA| = 2kn + 1 - m$ and $|kA - kA| = 2kn + 1 - 2m$ so that $|kA + kA| - |kA - kA| = m$. Note that this fringe only works if m is a multiple of k since R is shifted by m/k. In the second step, we fix this to allow m that is not a multiple of k by partially filling in the first gap of L, R. In the third step, we construct A such that $|kA + kA| = 2kn + 1 - m$ and $|kA - kA| = 2kn + 1 - \ell$ for any $\ell \leq 2m$ by extending the middle interval $[16km - 2m + 1, n - (16km - 2m + 1)]$ of A. In the last step, we get the full theorem for general a, b, c, d by changing how much R is shifted from L.

Step 1: m is a multiple of k.

We first prove that if m is a multiple of k, the above A has $|kA+kA| = 2kn+1-m, |kA-kA| = 2kn + 1 - 2m$. To find $|kA + kA|, |kA - kA|$, we need to study the fringes of $kA + kA, kA - kA$. We will use Lemma 2.1, which says that for any x, y that $xL + yR$ is a translation of L with the front filled in. In general, note that if R is shifted down from L by d, we have that $xL + yR$ ends at $x(8m) + y(8m + d) = (x + y)(8m) + yd$ and if $x + y$ is fixed, the result depends only on y and d. Hence as in Figure 1, the left fringe $kL + kL$ of $kA + kA$ moves slower than the right fringe $kR + kR$. Therefore the right fringe of $kA + kA$ reaches the middle before the left fringe of $kA + kA$, resulting in some missing elements in the left fringe but no missing elements in the right fringe. By Figure 2, the fringes $kL + kR$ of $kA - kA$ each have some missing elements since $kL + kR$ also moves slower than $kR + kR$.

To be precise, the left fringe $kL + kL$ of $kA + kA$ is

$$
\begin{aligned}
kL + kL &= [0, 16km - 4m] \cup [16km - 3m + 1, 16km - 2m] \cup \{16km\} \\
&= (L + 16km - 8m) \cup [0, 16km - 8m - 1].
\end{aligned} \tag{A.3}
$$

Note by (A.1) that the middle of $kA + kA$ on the left side starts at $16km - 2m + 1$. Therefore, $kA + kA$ is missing the m elements in $[16km - 4m + 1, 16km - 3m]$ in its left fringe.

The right fringe of $kA + kA$ is $2kn - (kR + kR)$ and so after reflection, we only need to study $kR + kR$, which is

$$
kR + kR = [0, 16km - 2m] \cup [16km - m + 1, 16km] \cup \{16km + 2m\}. \tag{A.4}
$$

Again by (A.1) note that the middle of $kA + kA$ on the right side starts at $2kn - (16km - 2m + 1)$, which is $16km - 2m + 1$ after reflection. This covers the missing elements of $kR + kR$ and so $kA + kA$ has no missing elements in its right fringe.

Since the middle of $kA + kA$ is filled in, $kA + kA$ has all elements except for the m missing elements in its left fringe and so $|kA + kA| = 2kn + 1 - m$.

Now we need to study the fringes of $kA - kA$. Note that $kA - kA$ is symmetric so the left and right fringes are the same. The left fringe of $kA - kA$ is $kL - k(n - R) = kL + kR - kn$. After translation, we can study $kL + kR$, which is

$$
kL + kR = [0, 16km - 3m] \cup [16km - 2m + 1, 16km - m] \cup \{16km + m\}. \tag{A.5}
$$

After translation, the middle of $kA - kA$ starts on the left side at $16km - 2m + 1$ as before. Therefore, the middle covers the first gap $[16km - m + 1, 16km + m - 1]$ in $kL + kR$ but not the second gap $[16km - 3m + 1, 16km - 2m]$, which has m elements. Therefore, the left fringe of $kA - kA$ has m missing elements. By symmetry, the right fringe of $kA - kA$ also has m missing elements. Since the middle of $kA - kA$ is filled in, $kA - kA$ has all elements except for $2m$ elements and so $|kA - kA| = 2kn + 1 - 2m$.

Finally, we note that it is sufficient to take n such that $n - 2(16km - 2m + 1) > 16m$. We make n large enough so that the middle of A has size at least $16m$, the size of the original fringes L, R. In fact, we just need that the middle of $kA + kA, kA - kA$ has enough elements to cover the second gap of the $kL + kL, kR + kR$, and $kL + kR$.

Step 2: m is not a multiple of k.

To do the case when m is not a multiple of k, we use the same fringes as before but partially fill in their gaps. Let m' be the smallest multiple of k that is greater than or equal to m. By (A.1) and Step 1, we can construct A' such that $|kA' + kA'| = 2kn + 1 - m'$ and $|kA' - kA'| = 2kn + 1 - 2m'$.

That is, the left fringe of A' is

$$L' = [0, 4m'] \cup [5m' + 1, 6m'] \cup \{8m'\} \tag{A.6}$$

so that L' is defined like the original L but for m' instead of m. Now we note that since the middle of A' starts at $16km' - 2m' + 1$, the first gap of L' accounts for all the missing elements of $kL' + kL'$ and $kL' + kR'$. In fact, a copy of $L' \cap [4m' + 1, 5m']$ appears identically in the left fringe of $kA' + kA'$ and $kA' - kA'$. Therefore, we can fill in the first $m' - m$ elements of the first gap of L' by considering

$$L'' = L' \cup [4m' + 1, 4m' + (m' - m)]. \tag{A.7}$$

and do the same to construct R'' from R'. Then $kL'' + kL''$ will have only m missing elements since $kL' + kL'$ has m' missing elements and we filled in $m' - m$ elements. Also note that $kR'' + kR''$ has no missing elements since $kR' + kR'$ did not have any missing elements. Thus, if we construct A from L'' and R'', we have $|kA + kA| = 2n + 1 - m$. Note that for this construction, we can fill in *any* $m' - m$ elements of the first gap of L', not necessarily the first $m' - m$ elements.

Similarly, $kL'' + kR''$ now misses only m elements since it also has a copy of $L'' \cap [4m' + 1, 5m']$. Therefore $kA - kA$ has m missing elements in each fringe and so $|kA - kA| = 2kn + 1 - 2m$.

Step 3: Arbitrary $m, \ell \leq 2m$.

Now we further modify A so that for any m and $\ell \leq m$, we have $|kA + kA| = kn + 1 - m$ and $|kA - kA| = kn + 1 - 2\ell$. Note that again we must do the cases when m is multiple of k and when m is not a multiple separately. However, we only do the case where m is a multiple of k since from Step 2, it is clear how to extend to other case.

In particular, we will modify A by extending the middle section in both directions by $m - \ell$. Therefore the middle of $kA + kA$ now starts at $16km - 2m + 1 - (m - \ell)$. Recall that the missing elements in $kL + kL$ occur only from the first gap $[16km - 4m + 1, 16km - 3m]$. Since $\ell \geq 0$, we have $16km - 2m + 1 - (m - \ell) \geq 16km - 3m + 1$ and so $kL + kL$ is still missing m elements. As before, $kR + kR$ has no missing elements and so we still have $|kA + kA| = 2kn + 1 - m$.

On the other hand, $kL + kR$ has fewer missing elements than it usually would. Note that now the middle of $kA - kA$ also starts at $16km - 2m + 1 - (m - \ell) = 16km - 3m + \ell + 1$. Since the missing elements in $kL + kR$ occur only from the first gap $[16km - 3m + 1, 16km - 2m]$ of $kL + kR$, then $kL + kR$ has only the missing ℓ elements $[16km - 3m + 1, 16km - 3m + \ell]$. Therefore, we get that $kA - kA$ is missing only ℓ elements in each fringe and so $|kA - kA| = 2kn + 1 - 2\ell$.

Note that we cannot do better than having $|kA - kA| = 2kn + 1 - 2\ell$ with $\ell \leq m$ with this approach. Shortening the middle does not help since although it increases the number of missing elements in $kA - kA$, it also increases the number of missing elements in $kA + kA$.

Step 4: Arbitrary a, b, c, d.

Finally, we modify A to prove the desired theorem for arbitrary a, b, c, d. In particular, we will modify A by changing how much R is shifted from L. This changes the speed at which the right fringe approaches the middle. We adjust the speed so that the right fringe of $aA - bA$ has no missing elements while all the other fringes still have some missing elements.

We again make some simplifying assumptions. We will only construct A such that $|aA - bA| = qn + 1 - m$ and $|cA - dA| = qn + 1 - 2m$ since we can use the methods from Step 3 to extend

to the case with $|aA - bA| = qn + 1 - m, |cA - dA| = qn + 1 - \ell$, where $\ell \leq 2m$. Since $|aA - bA| = |bA - aA|$, we can assume $a > b$ and $c > d$. Furthermore, since $a + b = c + d = q$ and a is the maximal element, we have that $a > c$ and $b < d$. We first assume that $c \neq d$ and then discuss how to do case when $c = d$; note that in the case $c = d$, we must have ℓ be even. We must also break up the proof into the case when $c - d \leq d - b$ and when $c - d > d - b$. We will only do the case when $c - d \leq d - b$ and then discuss how to do the other case. Finally, we must consider separately the case when m is a multiple of $c - d$ and when m is not; we will only do the former since the latter follows as in Step 2.

We now construct A such that $|aA - bA| = qn + 1 - m$ and $|cA - dA| = qn + 1 - 2m$, with $c \neq d$ and m a multiple of $c - d$.

We first let $\Delta = m/(c - d)$ and

$$\begin{aligned} L &= [0, 2\Delta(a - b)] \cup [2\Delta(a - b) + \Delta(c - d) + 1, 3\Delta(a - b)] \cup \{4\Delta(a - b)\} \\ R &= (L + \Delta) \cup [0, \Delta - 1]. \end{aligned} \tag{A.8}$$

These fringes are similar to the fringes in (A.1) except that the middle block of L, R has a different size and R is shifted from L by a different amount. Also let

$$A = L \cup [4\Delta(a-b)(a+b)+\Delta a-2\Delta(a-b)+1, n-(4\Delta(a-b)(a+b)+\Delta a-2\Delta(a-b)+1)] \cup (n-R). \tag{A.9}$$

The middle is chosen to start at $4\Delta(a - b)(a + b) + \Delta a - 2\Delta(a - b) + 1$, which is 1 after the end of the first block of $bL + aR$, the right fringe of $aA - bA$.

We first study $aA - bA$. The left fringe of $aA - bA$ is $aL - (b(n - R))$, which is $aL + bR$ after translation. The maximum element of $aL + bR$ is

$$4\Delta(a - b)a + (4\Delta(a - b) + \Delta)b = 4\Delta(a - b)(a + b) + \Delta b \tag{A.10}$$

and the pattern to the left of the maximum element is exactly the same as in L and R. That is,

$$\begin{aligned} &aL + bR \\ &= [0, 4\Delta(a - b)(a + b) + \Delta b - 2\Delta(a - b)] \\ &\cup [4\Delta(a - b)(a + b) + \Delta b - 2\Delta(a - b) + \Delta(c - d) + 1, 4\Delta(a - b)(a + b) + \Delta b - \Delta(a - b)] \\ &\cup \{4\Delta(a - b)(a + b) + \Delta b\}. \end{aligned} \tag{A.11}$$

Since the middle of A starts at

$$4\Delta(a - b)(a + b) + \Delta a - 2\Delta(a - b) + 1 = 4\Delta(a - b)(a + b) + \Delta b - \Delta(a - b) + 1, \tag{A.12}$$

we see that $aL + bR$ is missing the $\Delta(c - d) = m$ elements

$$[4\Delta(a-b)(a+b)+\Delta b-2\Delta(a-b)+1, 4\Delta(a-b)(a+b)+\Delta b-2\Delta(a-b)+\Delta(c-d)], \tag{A.13}$$

which are all the elements in its first gap.

Now we must consider the right fringe $a(n - R) - bL$ of $aA - bA$, which $bL + aR$ after reflection. Note that

$$\begin{aligned} bL + aR &= aL + bR + \Delta(a - b) \\ &= [0, 4\Delta(a - b)(a + b) + \Delta a - 2\Delta(a - b)] \\ &\cup [4\Delta(a - b)(a + b) + \Delta a - 2\Delta(a - b) + \Delta(c - d) + 1, 4\Delta(a - b)(a + b) + \Delta a - \Delta(a - b)] \\ &\cup \{4\Delta(a - b)(a + b) + \Delta a\}, \end{aligned} \tag{A.14}$$

with the first equality since R is L shifted down by Δ. Note that $bL + aR$ is not missing any elements since the middle of A starts at

$$4\Delta(a-b)(a+b) + \Delta a - 2\Delta(a-b) + 1, \tag{A.15}$$

which is exactly where the first gap of $bL + aR$ starts. Therefore $|aA - bA| = 2kn + 1 - m$ since $aA - bB$ is missing only m elements in its left fringe.

Now we will consider $cA - dA$. Its left fringe is $cL - d(n-R)$, which is $cL + dR$ after translation. Then as before,

$$
\begin{aligned}
cL + dR &= aL + bR + \Delta(d - b) \\
&= [0, 4\Delta(a-b)(a+b) + \Delta d - 2\Delta(a-b)] \\
&\quad \cup [4\Delta(a-b)(a+b) + \Delta d - 2\Delta(a-b) + \Delta(c-d) + 1, 4\Delta(a-b)(a+b) + \Delta d - \Delta(a-b)] \\
&\quad \cup \{4\Delta(a-b)(a+b) + \Delta d\}.
\end{aligned}
\tag{A.16}
$$

Then $cL + dR$ is only missing the $\Delta(c-d) = m$ elements

$$[4\Delta(a-b)(a+b) + \Delta d - 2\Delta(a-b) + 1, 4\Delta(a-b)(a+b) + \Delta d - 2\Delta(a-b) + \Delta(c-d)], \tag{A.17}$$

which are all the elements in its first gap. This is because the middle of A starts in the middle of the second block of $cL + dR$ since

$$4\Delta(a-b)(a+b) + \Delta d - 2\Delta(a-b) + \Delta(c-d) + 1 \leq 4\Delta(a-b)(a+b) + \Delta a - 2\Delta(a-b) \tag{A.18}$$

as $c < a$ and

$$4\Delta(a-b)(a+b) + \Delta a - 2\Delta(a-b) \leq 4\Delta(a-b)(a+b) + \Delta d - \Delta(a-b) \tag{A.19}$$

as $b > d$.

Similarly, the right fringe of $cA - dA$ is $c(n-R) - dL$, which is $dL + cR$. Then

$$
\begin{aligned}
dL + cR &= aL + bR + \Delta(c - b) \\
&= [0, 4\Delta(a-b)(a+b) + \Delta c - 2\Delta(a-b)] \\
&\quad \cup [4\Delta(a-b)(a+b) + \Delta c - 2\Delta(a-b) + \Delta(c-d) + 1, 4\Delta(a-b)(a+b) + \Delta c - \Delta(a-b)] \\
&\quad \cup \{4\Delta(a-b)(a+b) + \Delta c\},
\end{aligned}
\tag{A.20}
$$

and as before, $dL + cR$ is only missing the m elements

$$[4\Delta(a-b)(a+b) + \Delta c - 2\Delta(a-b) + 1, 4\Delta(a-b)(a+b) + \Delta c - 2\Delta(a-b) + \Delta(c-d)], \tag{A.21}$$

which are all the elements in its first gap. This is because the middle of A starts in the middle of the second block of $dL + cR$ since

$$4\Delta(a-b)(a+b) + \Delta c - 2\Delta(a-b) + \Delta(c-d) + 1 \leq 4\Delta(a-b)(a+b) + \Delta a - 2\Delta(a-b) \tag{A.22}$$

and

$$4\Delta(a-b)(a+b) + \Delta a - 2\Delta(a-b) \leq 4\Delta(a-b)(a+b) + \Delta c - \Delta(a-b). \tag{A.23}$$

To verify the first inequality, we note that $2c - d < a$ holds since $a + b = c + d$ and $c - b < d - b$ in this case. The second inequality follows from $b \leq c$. Therefore, $cA - dA$ is missing m elements in each fringe and so $|cA - dA| = 2kn + 1 - 2m$.

To do the case $c - d > d - b$, we need to change the fringes slightly. However, the only real difference occurs when we extend the middle to get $|cA - dA| = 2kn + 1 - \ell$, where $\ell \leq 2m$, as in Step 3. We do this by first extending the middle one element at a time (to decrease ℓ one element

at a time). However, at a certain point we need to extend the middle by adding a whole block; at this point extending one element does not change the value of $|cA - dA|$ and so we just extend by a whole block. Afterwards, we continue extending the middle one element at a time as before.

Finally, we note that the case when $c = d$ is similar to the result achieved in Step 3, except that now the left fringe $aL + bR$ of $aA - bA$ is closer to the middle; therefore we need to make the middle shorter so that the middle misses the first gap in the left fringe of $aL + bR$. This completes the proof of Theorem 1.3. □

REFERENCES

[FP]   G. A. Freiman and V. P. Pigarev, *The relation between the invariants R and T*, Number theoretic studies in the Markov spectrum and in the structural theory of set addition (Russian), Kalinin. Gos. Univ., Moscow, 1973, 172–174.

[He]   P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007), Acta Arithmetica **130** (2007), no. 1, 61–77.

[HM1]   P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136. http://arxiv.org/abs/0707.3417

[HM2]   P. V. Hegarty and S. J. Miller, Appendix 2 of *Explicit constructions of infinite families of MSTD sets* (by S. J. Miller and D. Scheinerman), Additive Number Theory: Festschrift In Honor of the Sixtieth Birthday of Melvyn B. Nathanson (David Chudnovsky and Gregory Chudnovsky, editors), Springer-Verlag, 2010.

[ILMZ] G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Finding and Counting MSTD sets*, preprint (2011). http://arxiv.org/abs/1107.2719.

[Ma]   J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.

[MO]   G. Martin and K. O'Bryant, *Many sets have more sums than differences*, in Additive Combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287ï£¡305.

[MOS]   S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010) 1221–1233.

[MPR]   S. J. Miller, S. Pegado and S. L. Robinson, *Explicit constructions of infinite families of generalized MSTD sets*, preprint.

[Na1]   M. B. Nathanson, *Sums of finite sets of integers*, The American Mathematical Monthly, Vol. 79, No. 9 (Nov., 1972), pp. 1010-1012.

[Na2]   M. Nathanson, *Additive Number Theory: The Classical Bases*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1996.

[Na3]   M. B. Nathanson, *Problems in additive number theory, 1*, Additive combinatorics, 263–270, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007.

[Na4]   M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp).

[Ru1]   I. Z. Ruzsa, *On the cardinality of $A + A$ and $A - A$*, Combinatorics year (Keszthely, 1976), vol. 18, Coll. Math. Soc. J. Bolyai, North-Holland-Bolyai Tàrsulat, 1978, 933–938.

[Ru2]   I. Z. Ruzsa, *Sets of sums and differences*, Séminaire de Théorie des Nombres de Paris 1982-1983 (Boston), Birkhüser, 1984, 267–273.

[Ru3]   I. Z. Ruzsa, *On the number of sums and differences*, Acta Math. Sci. Hungar. **59** (1992), 439–447.

[Zh1]   Y. Zhao, *Constructing MSTD Sets Using Bidirectional Ballot Sequences*, Journal of Number Theory **130** (2010), no. 5, 1212–1220.

[Zh2]   Y. Zhao, *Sets Characterized by Missing Sums and Differences*, Journal of Number Theory **131** (2011), 2107–2134.

E-mail address: geoff.iyer@gmail.com

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109

E-mail address: olazarev@Princeton.edu

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NJ 08544

E-mail address: sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: lz1@williams.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267