

Counting MSTD Sets in Finite Abelian Groups

Yufei Zhao

Massachusetts Institute of Technology
yufeiz@mit.edu

August 8, 2010

Abstract

In an abelian group G , a more sums than differences (MSTD) set is a subset $A \subset G$ such that $|A + A| > |A - A|$. We provide asymptotics for the number of MSTD sets in finite abelian groups, extending previous results of Nathanson. The proof contains an application of a recently resolved conjecture of Alon and Kahn on the number of independent sets in a regular graph.

1 Introduction

Let G be an abelian group. A *more sums than differences* set is a finite subset $A \subset G$ such that $|A + A| > |A - A|$, where the sum set $A + A$ and the difference set $A - A$ are subsets of G defined as

$$\begin{aligned}A + A &= \{a + b : a, b \in A\}, \\A - A &= \{a - b : a, b \in A\}.\end{aligned}$$

We will use $\text{MSTD}(G)$ to denote the collection of all MSTD sets in G , and $|\text{MSTD}(G)|$ the number of MSTD sets in G . Since addition is commutative while subtraction is not, two generic elements generate one sum but two differences. So we should generally expect there to be at least as many differences as sums. However, this is not always the case, and the exceptions are the MSTD sets.

There has been extensive of research on MSTD sets of integers. The first example of an MSTD set of integers was found by Conway in the 1960's: $\{0, 2, 3, 4, 7, 11, 12, 14\}$. The name MSTD was later given by Nathanson [12]. For recent papers on MSTD sets, see [3, 4, 8, 10, 11, 12, 22, 21]. For older papers see [5, 7, 13, 15, 16, 17, 18]. We refer the readers to [11, 12] for the history of the problem.

Almost all previous research on MSTD sets focused exclusively on the integers, as opposed to other abelian groups. The only paper where MSTD sets in finite abelian groups are considered is by Nathanson [12], who showed that families of MSTD sets of integers can be constructed from MSTD sets in finite abelian groups. As an illustration, if $A \subset \mathbb{Z}/n\mathbb{Z}$ satisfies $|A + A| > |A - A|$,

then $\{a \in \mathbb{Z} : a \bmod n \in A, 0 \leq a < kn\}$ is an MSTD set of integers for sufficiently large k . In general, starting from an MSTD set in a finite abelian group, we can produce an MSTD subset of the lattice \mathbb{Z}^d in an analogous manner, and then we can obtain an MSTD set of the integers through the group homomorphism $\phi : \mathbb{Z}^d \rightarrow \mathbb{Z}$ defined by

$$\psi(a) = \sum_{i=1}^d a_i m^{i-1}$$

where m is some sufficiently large integer. This connection to MSTD sets of integers is the initial motivation for studying MSTD sets in finite abelian groups. Nathanson showed that

$$|\text{MSTD}(\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z})| \geq \begin{cases} 2^n \left(1 - \frac{2n}{2^{n/2}}\right) & \text{if } n \text{ is even,} \\ 2^n \left(1 - \frac{\sqrt{2n}}{2^{n/2}}\right) & \text{if } n \text{ is odd.} \end{cases}$$

Our main result improves and generalizes Nathanson's result. We give asymptotics for $|\text{MSTD}(G)|$ for large finite abelian groups G . Recall that the notation $f_n \sim g_n$ means that $f_n/g_n \rightarrow 1$ as $n \rightarrow \infty$.

Theorem 1.1. *Let $\{G_n\}$ be a sequence of finite abelian groups with $|G_n| \rightarrow \infty$.*

1. *(Even case) If G_n has $k_n > 0$ elements of order 2, and $\limsup_{n \rightarrow \infty} \frac{k_n}{|G_n|} < 1 - \frac{1}{2} \log_3 7 = 0.114\dots$, then*

$$|\text{MSTD}(G_n)| \sim k_n \cdot 3^{|G_n|/2}.$$

2. *(Odd case) If every $|G_n|$ is odd, and the proportion of elements in G_n with order less than $\log_\varphi |G_n|$ approaches 0 as $n \rightarrow \infty$, then*

$$|\text{MSTD}(G_n)| \sim \frac{1}{2} |G_n| \varphi^{|G_n|},$$

where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

The hypotheses on G_n in both cases are rather mild. For instance, if the rank of G_n is uniformly bounded, then k_n is uniformly bounded so that the hypotheses for the even case hold, and Proposition 4.3 will show that the hypotheses for the odd case hold as well. As examples, here are two corollaries, the second of which is an improvement of Nathanson's result stated before Theorem 1.1.

Corollary 1.2. $|\text{MSTD}(\mathbb{Z}/n\mathbb{Z})| \sim \begin{cases} 3^{n/2}, & \text{if } n \text{ is even,} \\ \frac{1}{2} n \varphi^n, & \text{if } n \text{ is odd.} \end{cases}$

Corollary 1.3. $|\text{MSTD}(\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z})| \sim \begin{cases} 3^{n+1} & \text{if } n \text{ is even,} \\ 3^n & \text{if } n \text{ is odd.} \end{cases}$

Let us mention some analogous results for MSTD sets of integers. Martin and O’Byrant [8] showed that, as $n \rightarrow \infty$, there is a positive lower bound to the proportion of subsets of $\{1, 2, \dots, n\}$ which are MSTD sets. Recently the author [21] showed that this proportion in fact approaches a limit greater than 4×10^{-4} as $n \rightarrow \infty$. Monte Carlo experiments suggest that the limit should be around 4.5×10^{-4} . Families of MSTD subsets of $\{1, 2, \dots, n\}$ were constructed by Miller, Orosz, and Scheinerman [10] and the author [22].

The rest of the paper contains the proof of Theorem 1.1. One feature of the proof is that it contains an application of a graph theory conjecture of Alon [1] and Kahn [6] which was recently resolved by the author [23]; see Theorem 3.2.

The paper is structured as follows. In Section 2 we describe our strategy for bounding the number of MSTD sets. We reduce the problem into counting the number subsets of G that avoid specific sums and/or differences, and they are analyzed using forbiddance graphs, which are discussed in Section 3. In Section 4 we put all the individual forbiddance graph inequalities together to give a bound for the number of MSTD sets. Section 5 contains some further questions.

2 Union bounds

First, let us give some intuition on why MSTD sets in finite abelian groups behave differently from MSTD sets of integers. In the case of integers, say if we were interested in subsets of $\{1, 2, \dots, n\}$ that are MSTD sets, the “middle” sums and differences are almost always present since they can each be represented as a sum or difference in many ways. Thus it is the fringe elements that matter the most. It was recently shown [21] that a “typical” MSTD subset of $\{1, 2, \dots, n\}$ consists of a well-controlled fringe and an almost unrestricted middle. On the other hand, in the case of finite abelian groups, there are no longer any fringe elements. Consequently, almost all subsets A of a finite abelian group G satisfy $A + A = A - A = G$, and MSTD sets occupy a diminishing fraction of the subsets of G , unlike the integers case.

Now, returning to the case of finite abelian groups, if $A \subset G$ is an MSTD set, then we must have $A - A \neq G$. On the other hand, if $A - A \neq G$ but $A + A = G$, then A is necessarily an MSTD set. Therefore

$$\{A \subset G : A - A \neq G, A + A = G\} \subset \text{MSTD}(G) \subset \{A \subset G : A - A \neq G\}. \quad (1)$$

We will use union bounds to estimate the sizes of the sets in (1). Let G' denote a subset of G such that for every nonzero $d \in G$, exactly one of $d, -d$ appears in G' . Also assume that $0 \notin G'$. For example, if $G = \mathbb{Z}/8\mathbb{Z}$, then we could use $G' = \{1, 2, 3, 4\} \subset G$ (though $G' = \{1, 3, 4, 6\}$ is an equally valid choice).

Using (1), we obtain the following upper bound

$$|\text{MSTD}(G)| \leq |\{A \subset G : A - A \neq G\}| \leq \sum_{d \in G'} |\{A \subset G : d \notin A - A\}|. \quad (2)$$

For $d \in G$, let

$$D_d = \{A \subset G : d \notin A - A, A + A = G\}.$$

Then

$$\{A \subset G : A - A \neq G, A + A = G\} = \bigcup_{d \in G'} D_d.$$

We have

$$\left| \bigcup_{d \in G'} D_d \right| \geq \sum_{d \in G'} |D_d| - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |D_{d_1} \cap D_{d_2}|$$

Also,

$$\begin{aligned} |D_d| &= \{A \subset G : d \notin A - A, A + A = G\} \\ &\geq |\{A \subset G : d \notin A - A\}| - \sum_{s \in G} |\{A \subset G : d \notin A - A, s \notin A + A\}|, \end{aligned}$$

and

$$|D_{d_1} \cap D_{d_2}| = |\{A \subset G : d_1, d_2 \notin A - A, A + A = G\}| \leq |\{A \subset G : d_1, d_2 \notin A - A\}|.$$

Putting everything together, we obtain the following lower bound

$$\begin{aligned} |\text{MSTD}(G)| &\geq |\{A \subset G : A - A \neq G, A + A = G\}| \\ &= \left| \bigcup_{d \in G'} D_d \right| \\ &\geq \sum_{d \in G'} |D_d| - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |D_{d_1} \cap D_{d_2}| \\ &\geq \sum_{d \in G'} \left(|\{A \subset G : d \notin A - A\}| - \sum_{s \in G} |\{A \subset G : d \notin A - A, s \notin A + A\}| \right) \\ &\quad - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |\{A \subset G : d_1, d_2 \notin A - A\}|. \end{aligned} \tag{3}$$

In the next section, we explain how to compute the individual terms on the RHS of (2) and (3).

3 Fibonacci indices of forbiddance graphs

3.1 Forbiddance graph

We would like to compute the cardinalities of collections of subsets of forms

$$\{A \subset G : d \notin A - A\}, \{A \subset G : d \notin A - A, s \notin A + A\}, \text{ and } \{A \subset G : d_1, d_2 \notin A - A\}.$$

More generally, for a collection of the form

$$\mathcal{A} = \{A \subset G : d_1, \dots, d_p \notin A - A; s_1, \dots, s_q \notin A + A\},$$

we call d_1, \dots, d_p the *forbidden differences* and s_1, \dots, s_q the *forbidden sums*. We define the *forbiddance graph* $\mathcal{G}(\mathcal{A})$ to be the graph with G as its sets of vertices, and an edge between two vertices (possibly coinciding) whenever their sum or difference is forbidden. In particular, if $2x$ is a forbidden sum, then there is a loop at x . See Figure 1 for an example of a forbiddance graph. If there are no forbidden sums, then the forbiddance graph is simply the undirected Cayley graph of the forbidden differences. The significance of forbiddance graphs is given in the following key observation. Recall that an independent set in a graph is a subset of the vertices with no two adjacent.

Observation 3.1. *The elements of \mathcal{A} are in natural bijective correspondence with independent sets of vertices in the forbiddance graph $\mathcal{G}(\mathcal{A})$.*

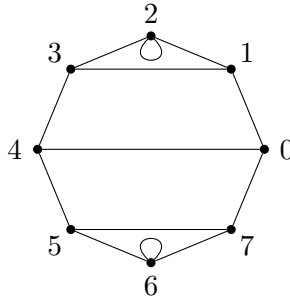


Figure 1: The forbiddance graph of $\{A \subset \mathbb{Z}/8\mathbb{Z} : 1 \notin A - A, 4 \notin A + A\}$. Note that the loops at 2 and 6 mean that neither vertex can be included in any independent set.

The number of independent sets in a graph G is known as the *Fibonacci index* of G , denoted $i(G)$. The above observation implies that

$$|\mathcal{A}| = i(\mathcal{G}(\mathcal{A})).$$

Fibonacci indices were introduced by Prodinger and Tichy [14]. In chemistry, $i(G)$ is known as the Merrifield-Simmons index [9]. Next we state some results about the Fibonacci index that we will use in the sequel.

3.2 Fibonacci index

We use F_n to denote the n -th Fibonacci number ($F_1 = F_2 = 1, F_{n+2} = F_{n+1} + F_n$) and L_n the n -th Lucas number ($L_1 = 1, L_2 = 3, L_{n+2} = L_{n+1} + L_n$). The path graph with n vertices is denoted by P_n , the cycle graph with n vertices is denoted by C_n , and the Cartesian graph product is denoted by \square .

The Fibonacci indices of many families of graphs are known (e.g., [20]). Here we state some results that are relevant to us. The derivation of the formulas can be found in the Appendix. (Assume $n \geq 1$; C_1 is the graph with one vertex and a loop.)

- Path: $i(P_n) = F_{n+2}$.
- Cycle: $i(C_n) = L_n = \varphi^n + (-\varphi)^{-n}$.
- Ladder: $i(P_n \square P_2) = \frac{1}{2}((1 + \sqrt{2})^{n+1} + (1 - \sqrt{2})^{n+1})$.
- Prism: $i(C_n \square P_2) = (1 + \sqrt{2})^n + (1 - \sqrt{2})^n + (-1)^n$.
- The Fibonacci index of a graph equals the product of the Fibonacci indices of its connected components.

We will also need an upper bound to the number of independent sets in a d -regular graph, which are graphs where vertex has degree d . This bound is provided by the following theorem which was conjectured implicitly by Alon [1], explicitly by Kahn [6], and recently proved by the author [23]. See also [2] for a different proof of Theorem 3.2 when $d \leq 5$. We will only use the theorem for $d = 3$ and 4. Note that we do not allow loops or multiple edges here.

Theorem 3.2. *For any N -vertex, d -regular simple graph G*

$$i(G) \leq (2^{d+1} - 1)^{N/(2d)}.$$

3.3 Forbidding one difference

We would like to determine the size of the collection

$$\mathcal{A} = \{A \subset G : d \notin A - A\}.$$

The forbiddance graph $\mathcal{G}(\mathcal{A})$ is easy to describe. Recall that L_n is the n -th Lucas number.

Lemma 3.3. *Let $d \in G$ be a nonzero element of order ℓ . Let $\mathcal{A} = \{A \subset G : d \notin A - A\}$. Then the forbiddance graph $\mathcal{G}(\mathcal{A})$ is a disjoint union of ℓ -cycles C_ℓ , and $|\mathcal{A}| = L_\ell^{|G|/\ell}$.*

Proof. The first statement is clear. The second statement follows from the facts listed in Section 3.2. □

Let $\text{ord}(d)$ denote the order of d in G . The next result follows from the above lemma. Recall that G' was defined in Section 2.

Lemma 3.4. *We have*

$$\sum_{d \in G'} |\{A \subset G : d \notin A - A\}| = \sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)}.$$

The next lemma is an easy fact about Lucas numbers.

Lemma 3.5. *The sequence $(L_{2n}^{1/2n})_{n \geq 1}$ is decreasing and the sequence $(L_{2n-1}^{1/(2n-1)})_{n \geq 1}$ is increasing. Both sequences approach the limit φ . In particular, $L_2^{1/2} > L_4^{1/4} \geq L_n^{1/n}$ for all $n > 2$.*

Proof. Recall that $L_n = \varphi^n + (-\varphi)^{-n}$, where $\varphi = \frac{1+\sqrt{5}}{2}$. So

$$L_{2n}^{1/2n} = (\varphi^{2n} + \varphi^{-2n})^{1/2n} = \varphi(1 + \varphi^{-4n})^{1/2n} \searrow \varphi,$$

and

$$L_{2n-1}^{1/(2n-1)} = (\varphi^{2n-1} - \varphi^{-(2n-1)})^{1/(2n-1)} = \varphi(1 - \varphi^{-2(2n-1)})^{1/(2n-1)} \nearrow \varphi. \quad \square$$

The above fact is central to the dichotomy between the even case and the odd case in the theorem. In the sequel, we will show that, under the hypotheses of Theorem 1.1, as $|G| \rightarrow \infty$,

$$|\text{MSTD}(G)| \sim \sum_{d \in G'} |\{A \subset G : d \notin A - A\}| = \sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)}, \quad (4)$$

In the case of even groups, the terms with $\text{ord}(d) = 2$ dominate the RHS sum. For odd groups, every summand can be approximated from above by $\varphi^{|G|}$, and the error is significant only when $\text{ord}(d)$ is small. In the remainder of this section, we will analyze the other terms in (3) and show that they are insignificant compared to the RHS sum in (4).

3.4 Other forbiddance graphs: even case

In this section we consider the case when $|G|$ is even. We will show that the $L_2^{|G|/2} = 3^{|G|/2}$ term in (4) asymptotically dominates the other terms in the sum in (3).

Lemma 3.6. *If $d \in G$ has order greater than 2, then $|\{A \subset G : d \notin A - A\}| \leq 7^{|G|/4}$.*

Proof. This follows from Lemmas 3.3 and 3.5. Recall that $L_4 = 7$. □

Now let us consider the case when we forbid one sum and one difference. The structure of the forbiddance graph is given by the following lemma.

Lemma 3.7. *Let $d, s \in G$, where d has order 2. Let k denote the number of elements of G of order 2. Let $\mathcal{A} = \{A \subset G : d \notin A - A, s \notin A + A\}$. Then, after removing vertices with loops, the forbiddance graph $\mathcal{G}(\mathcal{A})$ is a disjoint union of 4-cycles and at most $(k+1)/2$ copies of P_2 .*

Proof. Every connected component of $\mathcal{G}(\mathcal{A})$ consists of the elements $\{x, x+d, s-x-d, s-x\}$, for $x \in G$, and possibly with some repeats. If all four elements are distinct, then this component is a 4-cycle. If $x = s-x$, then also $x+d = s-x-d$ (recall that d has order 2), then this component consists of two connected vertices both with loops. If $x = s-x-d$, then also $x+d = s-x$, so the connected component is isomorphic to P_2 . The number of $x \in G$ satisfying $2x = s-d$ is at most $k+1$, and hence at most $(k+1)/2$ components can be isomorphic to P_2 . □

Lemma 3.8. *Let $d, s \in G$, where $d \neq 0$. Let k denote the number of elements of G of order 2. Then*

$$|\{A \subset G : d \notin A - A, s \notin A + A\}| \leq \begin{cases} 3^{(k+1)/2} \cdot 7^{|G|/4}, & \text{if } d \text{ has order 2,} \\ 7^{|G|/4}, & \text{if } d \text{ has order greater than 2.} \end{cases}$$

Proof. In Lemma 3.7, note that the number of 4-cycles cannot exceed $|G|/4$. Also recall that $i(P_2) = 3$, $i(C_4) = 7$. The first case then follows immediately.

The second case follows from Lemma 3.6 since

$$|\{A \subset G : d \notin A - A, s \notin A + A\}| \leq |\{A \subset G : d \notin A - A\}| \leq 7^{|G|/4}. \quad \square$$

Next we move onto the case when we forbid two differences.

Lemma 3.9. *Let $d_1, d_2 \in G$ be distinct elements with order 2. Then $\{A \subset G : d_1, d_2 \notin A - A\}$ has exactly $7^{|G|/4}$ elements.*

Proof. The connected components of the forbiddance graph are 4-cycles with vertices $x, x + d_1, x + d_1 + d_2, x + d_2$ for $x \in G$. Note that all four elements are distinct. Since $i(C_4) = 7$, we see that the Fibonacci index of the forbiddance graph is $7^{|G|/4}$. \square

Lemma 3.10. *Assume that $|G|$ is even. Let $d_1, d_2 \in G$ be distinct nonzero elements. Then*

$$|\{A \subset G : d_1, d_2 \notin A - A\}| \leq 7^{|G|/4}.$$

Proof. If both d_1 and d_2 have order 2, then this follows from Lemma 3.9. Otherwise, suppose without loss of generality that d_1 has order greater than 2, then from Lemma 3.6 we get

$$|\{A \subset G : d_1, d_2 \notin A - A\}| \leq |\{A \subset G : d_1 \notin A - A\}| \leq 7^{|G|/4}. \quad \square$$

3.5 Other forbiddance graphs: odd case

In this section we consider the case when $|G|$ is odd. We will show that the RHS sum in (4) dominates the RHS sum in (3).

First we analyze the case when we forbid one sum and one difference.

Lemma 3.11. *Let $d \in G$ be of odd order $\ell > 1$, and $s \in G$. Let $\mathcal{A} = \{A \subset G : d \notin A - A, s \notin A + A\}$. Then each connected component of the forbiddance graph $\mathcal{G}(\mathcal{A})$, after removing vertices with loops, is either a prism $C_\ell \square P_2$ or a ladder $P_{(\ell-1)/2} \square P_2$. Furthermore, if n_P is the number of prism components and n_L the number of ladder components, then $2n_P + n_L = |G|/\ell$.*

Proof. Let us start with the forbiddance graph of $\{A \subset G : d \notin A - A\}$, which consists of $|G|/\ell$ disjoint ℓ -cycles, and then add edges of the form $(x, s - x)$, $x \in G$, to obtain $\mathcal{G}(\mathcal{A})$. Let $x \in G$. There are two cases to consider.

Case 1. $s - x$ does not belong to the same cycle as x . Then the connected component of x in $\mathcal{G}(\mathcal{A})$ is a prism $C_\ell \square P_2$. The elements are connected as illustrated in Figure 2.

Case 2. $s - x$ belongs to the same cycle as x . Then the connected component of x in $\mathcal{G}(\mathcal{A})$ is the ℓ -cycle with some “parallel” chords of the form $(x + jd, s - x - jd)$ added in. See Figure 3. Since we start with an odd cycle, there is exactly one vertex with a loop, and removing this vertex gives us the ladder $P_{(\ell-1)/2} \square P_2$.

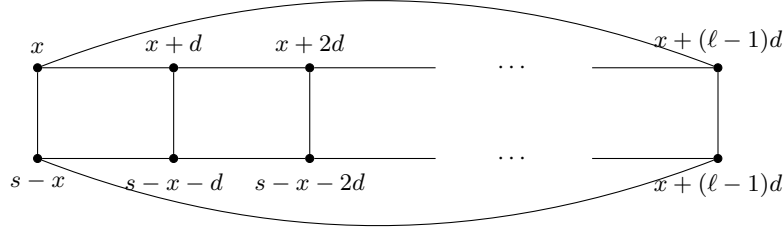


Figure 2: The connected component of x is the prism $C_\ell \square P_2$ in Case 1 of the proof of Lemma 3.11.

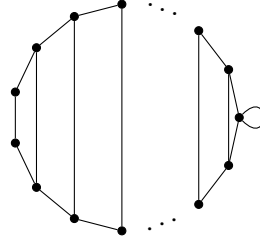


Figure 3: The connected component of x in Case 2 of the proof of the proof of Lemma 3.11.

Therefore, after removing the vertices with loops, all connected components are prisms or ladders, thereby establishing the first claim. In the first case, two cycles combine to form a prism, and in the second case, a cycle transforms into a ladder. Since we start with $|G|/\ell$ cycles, the claim $2n_P + n_L = |G|/\ell$ follows. \square

Using the notation from the lemma, we see that

$$\begin{aligned} |\{A \subset G : d \notin A - A, s \notin A + A\}| &= i(P_{(\ell-1)/2} \square P_2)^{n_L} \cdot i(C_\ell \square P_2)^{n_P} \\ &\leq \max \left\{ i(P_{(\ell-1)/2} \square P_2)^{|G|/\ell}, i(C_\ell \square P_2)^{|G|/(2\ell)} \right\}. \end{aligned} \quad (5)$$

The following lemmas show that the above quantity is insignificant.

Lemma 3.12. *Let $\ell \geq 3$ be an odd integer. Then $i(P_{(\ell-1)/2} \square P_2)^{1/\ell} < \sqrt{1 + \sqrt{2}}$.*

Proof. Using results from Section 3.2 and routine algebraic manipulation, we obtain

$$i(P_{(\ell-1)/2} \square P_2) = \frac{1}{2} \left((1 + \sqrt{2})^{(\ell+1)/2} + (1 - \sqrt{2})^{(\ell+1)/2} \right) < (1 + \sqrt{2})^{\ell/2}. \quad \square$$

Lemma 3.13. *Let $\ell \geq 3$, then $i(C_\ell \square P_2)^{1/(2\ell)} < 15^{1/6}$.*

Proof. Since $C_\ell \square P_2$ is a 3-regular simple graph, the lemma follows directly from Theorem 3.2. \square

We could have directly used the formula for $i(C_\ell \square P_2)$ in Section 3.2 to give a sharper bound, but the proof would be longer and the final conclusion would not be changed.

Using (5), Lemmas 3.12 and 3.13, and noting that $\sqrt{1 + \sqrt{2}} < 15^{1/6}$ we obtain the following upper bound for the case of forbidding one sum and one difference.

Lemma 3.14. *Suppose that $|G|$ is odd, and $d, s \in G$ with $d \neq 0$, then*

$$|\{A \subset G : d \notin A - A, s \notin A + A\}| < 15^{|G|/6}.$$

Next we consider the case of forbidding two differences.

Lemma 3.15. *Let $d_1, d_2 \in G$ be two nonzero elements, such that $2d_1 \neq 0$, $2d_2 \neq 0$, $d_1 \neq d_2$, $d_1 \neq -d_2$, then*

$$|\{A \subset G : d_1, d_2 \notin A - A\}| \leq 31^{|G|/8}.$$

Proof. The forbiddance graph is simple and 4-regular, since each vertex x is adjacent to four distinct vertices: $x + d_1, x - d_1, x + d_2, x - d_2$. The result then follows from Theorem 3.2. \square

4 Bounding the number of MSTD sets

In this section we prove Theorem 1.1. We use (2) and (3) to bound the number of MSTD sets in G and use the results from the previous section to bound the individual terms. The even case and the odd case are analyzed separately.

4.1 Even case

Assume that $|G|$ is even. Let k denote the number of elements of order 2 in G . Then from (2) and Lemmas 3.4 and 3.5, we have

$$\begin{aligned} |\text{MSTD}(G)| &\leq \sum_{d \in G'} |\{A \subset G : d \notin A - A\}| \\ &= \sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)} \\ &\leq k \cdot 3^{|G|/2} + |G| \cdot 7^{|G|/4} \\ &= k \cdot 3^{|G|/2} \left(1 + \frac{|G|}{k} \left(\frac{7}{9} \right)^{|G|/4} \right). \end{aligned} \quad (6)$$

For the lower bound, we use some facts from Section 3.4. For any element $d \in G$ with order 2, Lemma 3.8 gives

$$\begin{aligned} & |\{A \subset G : d \notin A - A, A + A = G\}| \\ & \geq |\{A \subset G : d \notin A - A\}| - \sum_{s \in G} |\{A \subset G : d \notin A - A, s \notin A + A\}| \\ & \geq 3^{|G|/2} - |G| 3^{(k+1)/2} \cdot 7^{|G|/4}. \end{aligned}$$

From Lemma 3.10 we know that for any distinct nonzero $d_1, d_2 \in G$,

$$|\{A \subset G : d_1, d_2 \notin A - A, A + A = G\}| \leq |\{A \subset G : d_1, d_2 \notin A - A\}| \leq 7^{|G|/4}.$$

Let $G_{(2)}$ be the subset of G containing the elements of order 2, so that $|G_{(2)}| = k$. Then, from (3) we get

$$\begin{aligned} |\text{MSTD}(G)| & \geq \sum_{d \in G'} |\{A \subset G : d \notin A - A, A + A = G\}| \\ & \quad - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |\{A \subset G : d_1, d_2 \notin A - A, A + A = G\}| \\ & \geq \sum_{d \in G_{(2)}} |\{A \subset G : d \notin A - A, A + A = G\}| \\ & \quad - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |\{A \subset G : d_1, d_2 \notin A - A, A + A = G\}| \\ & \geq k \left(3^{|G|/2} - |G| 3^{(k+1)/2} \cdot 7^{|G|/4} \right) - |G|^2 7^{|G|/4} \\ & = k \cdot 3^{|G|/2} \left(1 - \left(|G| \cdot 3^{(k+1)/2} + \frac{|G|^2}{k} \right) \left(\frac{7}{9} \right)^{|G|/4} \right). \end{aligned} \tag{7}$$

By combining (6) and (7), and using the notation of the even case of Theorem 1.1, we obtain that

$$1 - \left(|G_n| \cdot 3^{(k_n+1)/2} + \frac{|G_n|^2}{k_n} \right) \left(\frac{7}{9} \right)^{|G_n|/4} \leq \frac{|\text{MSTD}(G_n)|}{k_n \cdot 3^{|G_n|/2}} \leq 1 + \frac{|G_n|}{k_n} \left(\frac{7}{9} \right)^{|G_n|/4}.$$

If $|G_n| \rightarrow \infty$ and $\limsup_{n \rightarrow \infty} \frac{k_n}{|G_n|} < 1 - \frac{1}{2} \log_3 7$, then letting $n \rightarrow \infty$ gives us

$$\lim_{n \rightarrow \infty} \frac{|\text{MSTD}(G_n)|}{k_n \cdot 3^{|G_n|/2}} = 1,$$

thereby proving the even case of Theorem 1.1.

4.2 Odd case

Now assume that $|G|$ is odd. We use (2) and Lemma 3.3 to obtain an upper bound for $|\text{MSTD}(G)|$,

$$|\text{MSTD}(G)| \leq \sum_{d \in G'} |A \subset G : d \notin A - A| = \sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)}. \quad (8)$$

For the lower bound, we use (3) and Lemmas 3.3, 3.14, and 3.15 to get

$$\begin{aligned} |\text{MSTD}(G)| &\geq \sum_{d \in G'} |\{A \subset G : d \notin A - A\}| - \sum_{d \in G'} \sum_{s \in G} |\{A \subset G : d \notin A - A, s \notin A + A\}| \\ &\quad - \sum_{\substack{d_1, d_2 \in G' \\ d_1 \neq d_2}} |\{A \subset G : d_1, d_2 \notin A - A\}|. \\ &\geq \left(\sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)} \right) - |G|^2 \cdot 15^{|G|/6} - |G|^2 \cdot 31^{|G|/8}. \end{aligned} \quad (9)$$

By Lemma 3.5 and the fact that $\text{ord}(d)$ is odd for every $d \in G$, we have $L_{\text{ord}(d)}^{|G|/\text{ord}(d)} \geq L_3^{|G|/3} = 4^{|G|/3}$. Since $4^{1/3} > 15^{1/6} > 31^{1/8}$, (8) and (9) imply that

$$\begin{aligned} |\text{MSTD}(G)| &\sim \sum_{d \in G'} L_{\text{ord}(d)}^{|G|/\text{ord}(d)} \\ &\sim \frac{1}{2} \sum_{d \in G} L_{\text{ord}(d)}^{|G|/\text{ord}(d)} \\ &= \frac{1}{2} \sum_{d \in G} \left(\varphi^{\text{ord}(d)} + (-\varphi)^{-\text{ord}(d)} \right)^{|G|/\text{ord}(d)} \\ &= \frac{1}{2} \varphi^{|G|} \sum_{d \in G} \left(1 - \varphi^{-2\text{ord}(d)} \right)^{|G|/\text{ord}(d)}, \end{aligned} \quad (10)$$

where the second line follows from $\text{ord}(d) = \text{ord}(-d)$. The second step is not an equality due to the negligible $d = 0$ term. The last step uses the assumption that $|G|$ is odd, so that $\text{ord}(d)$ is odd. It remains to determine the asymptotics for the RHS of (10). We will need the following lemma.

Lemma 4.1. *Let $n \geq k > 0$, then*

$$1 - \left(1 - \varphi^{-2k} \right)^{n/k} \leq \min \left\{ 1, \frac{n}{k} \varphi^{-2k} \right\}.$$

Proof. The inequality $1 - \left(1 - \varphi^{-2k} \right)^{n/k} \leq 1$ is obvious. The inequality

$$1 - \left(1 - \varphi^{-2k} \right)^{n/k} \leq \frac{n}{k} \varphi^{-2k}$$

follows directly from Bernoulli's inequality, which states that $(1+x)^a \geq 1+ax$ whenever $x > -1$ and $a \geq 1$ (Bernoulli's inequality can be proved by checking the first derivative in x). \square

Let $\mathcal{E}(G)$ denote the subset of G consisting of elements whose orders are less than $\log_\varphi |G|$. So the hypotheses of Theorem 1.1 imply that $|\mathcal{E}(G)| = o(|G|)$.

Lemma 4.2. *Let G be an abelian group of odd order. Then*

$$|G| - |\mathcal{E}(G)| - 1 \leq \sum_{d \in G} \left(1 - \varphi^{-2 \operatorname{ord}(d)}\right)^{|G|/\operatorname{ord}(d)} \leq |G|.$$

Proof. The RHS inequality is clear since each term in the sum is at most 1. For the LHS, we need to prove that

$$\sum_{d \in G} \left(1 - \left(1 - \varphi^{-2 \operatorname{ord}(d)}\right)^{|G|/\operatorname{ord}(d)}\right) \leq |\mathcal{E}(G)| + 1.$$

We separate the terms with small order from those with large order. We have

$$\sum_{d \in \mathcal{E}(G)} \left(1 - \left(1 - \varphi^{-2 \operatorname{ord}(d)}\right)^{|G|/\operatorname{ord}(d)}\right) \leq \sum_{d \in \mathcal{E}(G)} 1 = |\mathcal{E}(G)|,$$

and from Lemma 4.1 we get

$$\begin{aligned} \sum_{d \notin \mathcal{E}(G)} \left(1 - \left(1 - \varphi^{-2 \operatorname{ord}(d)}\right)^{|G|/\operatorname{ord}(d)}\right) &\leq \sum_{d \notin \mathcal{E}(G)} \frac{|G|}{\operatorname{ord}(d)} \varphi^{-2 \operatorname{ord}(d)} \\ &\leq \sum_{d \notin \mathcal{E}(G)} \frac{|G|}{\log_\varphi |G|} \varphi^{-2 \log_\varphi |G|} \\ &= (|G| - |\mathcal{E}(G)|) \cdot \frac{|G|}{|G|^2 \log_\varphi |G|} \\ &\leq 1 \end{aligned}$$

This completes the proof of the lemma. □

Therefore, as $|G| \rightarrow \infty$, if $|\mathcal{E}(G)| = o(|G|)$, then

$$|\operatorname{MSTD}(G)| \sim \frac{1}{2} \varphi^{|G|} \sum_{d \in G} \left(1 - \varphi^{-2 \operatorname{ord}(d)}\right)^{|G|/\operatorname{ord}(d)} \sim \frac{1}{2} \varphi^{|G|} |G|.$$

This completes the odd case of Theorem 1.1.

Recall the *rank* of a finite abelian group is the smallest number of cyclic groups whose direct product gives the group. The following proposition shows that Theorem 1.1 can be applied when $\{G_n\}$ has uniformly bounded rank (this is already clear in the even case).

Proposition 4.3. *Let r and t be positive integers. Suppose that the finite abelian group G has rank at most r , then the number of elements of G with order less than t is less than t^{r+1} .*

Proof. In general, the number of elements of $\mathbb{Z}/a\mathbb{Z}$ with order dividing m is $\gcd(a, m)$. If an element of $\mathbb{Z}/a_1\mathbb{Z} \times \cdots \times \mathbb{Z}/a_r\mathbb{Z}$ has order dividing m , then each component has order dividing

m , and the number of such elements is $\prod_i \gcd(a_i, m) \leq m^r$. Therefore, the number of elements of order dividing m in G is at most m^r . Summing over all $m < t$, we find that the number of elements with order less than t is at most $1^r + 2^r + \dots + (t-1)^r < t^{r+1}$. \square

If the rank of $\{G_n\}$ is uniformly bounded by r , then $|\mathcal{E}(G_n)| < (\log_\varphi |G_n|)^{r+1} = o(|G_n|)$, and thus the hypotheses of Theorem 1.1 in the odd case are satisfied.

5 Discussion and further directions

We can ask for the rate of convergence in the asymptotics in Theorem 1.1. A closer look at the proofs indicate that, in the even case, the ratio of the two quantities converges to 1 exponentially fast with respect to the size of the group. The odd case does not converge as quickly—our proofs show that the rate of convergence is at most linear in the proportion of elements with order less than $\log_\varphi |G_n|$. We should be able to obtain a more precise asymptotic result for the odd case by including the number of elements with “small” order as a parameter. However, in the interest of keeping this paper short, we did not attempt such analysis.

We offer some ideas for generalizations and variations. The author [21] recently studied the number of subsets of $\{1, \dots, n\}$ which miss a particular number of sums and a particular number of differences. It would be nice to investigate similar questions for the case of finite abelian groups: given nonnegative integers s and d , how many subsets $A \subset G$ satisfies $|A + A| = |G| - s$ and $|A - A| = |G| - d$? We also ask for constructions of families of MSTD subsets of $\mathbb{Z}/n\mathbb{Z}$, similar to the constructions of families of MSTD subsets of $\{1, \dots, n\}$ in the integers given recently by Miller, Orosz, and Scheinerman [10] and the author [22]. There could be generalizations to other linear forms such as $A + A - A$; see [10, Sec. 4] for work in this direction in the integers case. Finally, we can consider variations where we choose a random subset of G based on some other probability model, for instance, so that the number of expected elements in the chosen subset is not $|G|/2$ but $O(\sqrt{|G|})$; see Hegarty and Miller [4] for this analysis in the case of integers.

A Computing the Fibonacci indices of certain graphs

In this appendix we prove the formula stated in Section 3.2 for the number of independent sets in the path, cycle, ladder, and prism graphs.

A.1 Path graph P_n

Let v be the first node in the path. If we do not include v in the independent set, then rest of the independent set can be chosen as any independent set of P_{n-1} . On the other hand, if we include v in the independent set, then the neighbor of v cannot be included, and the rest of the independent set can be chosen as any independent set of P_{n-2} . Thus $i(P_n) = i(P_{n-1}) + i(P_{n-2})$. Checking the initial values, we find that $i(P_n) = F_{n+2}$, where F_n is the n -th Fibonacci number ($F_1 = F_2 = 1$, $F_{n+2} = F_{n+1} + F_n$).

A.2 Cycle graph C_n

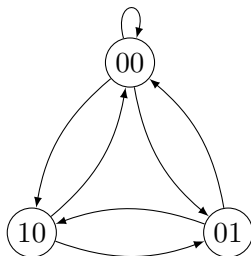
An argument analogous to the one above shows that

$$i(C_n) = i(P_{n-1}) + i(P_{n-3}) = F_{n+1} + F_{n-1} = L_n = \varphi^n + (-\varphi)^{-n}$$

where L_n is n -th Lucas number ($L_0 = 2, L_1 = 1, L_{n+2} = L_{n+1} + L_n$).

A.3 Ladder graph $P_n \times P_2$

We can compute $i(P_n \times P_2)$ using the transfer matrix method [19, Sec. 4.7]. The number of independent sets in $P_n \times P_2$ is equal to the number of walks of $n - 1$ steps (starting at any vertex) in the following graph. Indeed, every such walk corresponds to a labeling of the vertices of $P_n \times P_2$ with 0 and 1 such that no two 1's are adjacent.



The transfer matrix (i.e., adjacency matrix) is

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

The sum of all the entries of A^n equals to the number of walks of n steps in the above graph, which equals to $i(P_{n+1} \times P_2)$. We have

$$\sum_{n \geq 0} A^n x^n = (1 - Ax)^{-1} = \frac{1}{(1+x)(1-2x-x^2)} \begin{pmatrix} 1-x^2 & x+x^2 & x+x^2 \\ x+x^2 & 1-x-x^2 & x \\ x+x^2 & x & 1-x-x^2 \end{pmatrix}.$$

By taking the sum of all the entries of the matrix, we see that

$$\sum_{n \geq 0} i(P_{n+1} \times P_2) x^n = \frac{3+4x+x^2}{(1+x)(1-2x-x^2)} = \frac{3+x}{1-2x-x^2} = \frac{1}{2x} \left(\frac{1+\sqrt{2}}{1-(1+\sqrt{2})x} + \frac{1-\sqrt{2}}{1-(1-\sqrt{2})x} - 2 \right)$$

By comparing the coefficient of x^n , we find that

$$i(P_n \times P_2) = \frac{1}{2} \left((1+\sqrt{2})^{n+1} + (1-\sqrt{2})^{n+1} \right).$$

A.4 Prism graph $C_n \times P_2$

We can reuse the transfer matrix from the previous computation. Independent sets in $C_n \times P_2$ correspond to closed walks of length n , and so there are $\text{trace}(A^n)$ of them. This can be computed as the number of sum of n -th powers of the eigenvalues of A . Hence,

$$i(C_n \times P_2) = (1 + \sqrt{2})^n + (1 - \sqrt{2})^n + (-1)^n.$$

Acknowledgments

This research was carried out at the University of Minnesota Duluth under the supervision of Joseph Gallian with the financial support of the National Science Foundation and the Department of Defense (grant number DMS 0754106), the National Security Agency (grant number H98230-06-1-0013), and the MIT Department of Mathematics. The author would like to thank Joseph Gallian for his encouragement and support. The author would also like to thank Nathan Kaplan for reading the paper and making valuable suggestions.

References

- [1] N. Alon, *Independent sets in regular graphs and sum-free subsets of finite groups*, Israel J. Math. **73** (1991), no. 2, 247–256.
- [2] D. Galvin and Y. Zhao, *The number of independent sets in a graph with small maximum degree*, Graphs Combin. (to appear).
- [3] P. Hegarty, *Some explicit constructions of sets with more sums than differences*, Acta Arith. **130** (2007), no. 1, 61–77.
- [4] P. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures Algorithms **35** (2009), no. 1, 118–136.
- [5] F. Hennecart, G. Robert, and A. Yudin, *On the number of sums and differences*, Astérisque (1999), no. 258, xiii, 173–178, Structure theory of set addition.
- [6] J. Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combin. Probab. Comput. **10** (2001), no. 3, 219–237.
- [7] J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.
- [8] G. Martin and K. O’Bryant, *Many sets have more sums than differences*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287–305.
- [9] R.E. Merrifield and H.E. Simmons, *Topological methods in chemistry*, Wiley, 1989.

- [10] S. J. Miller, B. Orosz, and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, *J. Number Theory* **130** (2010), no. 5, 1221–1233.
- [11] M. B. Nathanson, *Problems in additive number theory. I*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 263–270.
- [12] ———, *Sets with more sums than differences*, *Integers* **7** (2007), A5, 24 pp. (electronic).
- [13] V. P. Pigarev and G. A. Freĭman, *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, pp. 172–174.
- [14] H. Prodinger and R. F. Tichy, *Fibonacci numbers of graphs*, *Fibonacci Quart.* **20** (1982), no. 1, 16–21.
- [15] F. Roesler, *A mean value density theorem of additive number theory*, *Acta Arith.* **96** (2000), no. 2, 121–138.
- [16] I. Z. Ruzsa, *On the cardinality of $A + A$ and $A - A$* , *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, Vol. II, *Colloq. Math. Soc. János Bolyai*, vol. 18, North-Holland, Amsterdam, 1978, pp. 933–938.
- [17] ———, *Sets of sums and differences*, *Seminar on number theory, Paris 1982–83 (Paris, 1982/1983)*, *Progr. Math.*, vol. 51, Birkhäuser Boston, Boston, MA, 1984, pp. 267–273.
- [18] ———, *On the number of sums and differences*, *Acta Math. Hungar.* **59** (1992), no. 3-4, 439–447.
- [19] R.P. Stanley, *Enumerative combinatorics. Vol. 1*, *Cambridge Studies in Advanced Mathematics*, vol. 49, Cambridge University Press, Cambridge, 1997.
- [20] E. W. Weisstein, *Independence polynomial*, from MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/IndependencePolynomial.html>.
- [21] Y. Zhao, *Sets characterized by missing sums and differences*, preprint at <http://arxiv.org/abs/0911.2292>.
- [22] ———, *Constructing MSTD sets using bidirectional ballot sequences*, *J. Number Theory* **130** (2010), no. 5, 1212–1220.
- [23] ———, *The number of independent sets in a regular graph*, *Combin. Probab. Comput.* **19** (2010), no. 2, 315–320.