Journal of Integer Sequences, Vol. 21 (2018),
Article 18.8.2

When Sets Can and Cannot Have Sum-Dominant Subsets

Hùng Việt Chu
Department of Mathematics
Washington and Lee University
Lexington, VA 24450
USA
chuh19@mail.wlu.edu

Nathan McNew
Department of Mathematics
Towson University
Towson, MD 21252
USA
nmcnew@towson.edu

Steven J. Miller
Department of Mathematics
Williams College
Williamstown, MA 01267
USA
sjm1@williams.edu
Steven.Miller.MC.96@aya.yale.edu

Victor Xu and Sean Zhang
Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213
USA
vzx@andrew.cmu.edu
xiaoronz@andrew.cmu.edu

Abstract

A finite set of integers A is a sum-dominant (also called a More Sums Than Differences or MSTD) set if $|A+A| > |A-A|$. While almost all subsets of $\{0,\dots,n\}$ are not sum-dominant, interestingly a small positive percentage are. We explore sufficient conditions on infinite sets of positive integers such that there are either no sum-dominant subsets, at most finitely many sum-dominant subsets, or infinitely many sum-dominant subsets. In particular, we prove no subset of the Fibonacci numbers is a sum-dominant set, establish conditions such that solutions to a recurrence relation have only finitely many sum-dominant subsets, and show there are infinitely many sum-dominant subsets of the primes.

1 Introduction

For any finite set of natural numbers $A \subset \mathbb{N}$, we define the sumset

$$A + A \ := \ \{a + a' : a, a' \in A\} \tag{1}$$

and the difference set

$$A - A \ := \ \{a - a' : a, a' \in A\}; \tag{2}$$

A is sum-dominant (also called a More Sums Than Differences or MSTD set) if $|A+A| > |A-A|$ (if the two cardinalities are equal it is called balanced, and otherwise difference-dominant). As addition is commutative and subtraction is not, it was natural to conjecture that sum-dominant sets are rare. Conway gave the first example of such a set, $\{0, 2, 3, 4, 7, 11, 12, 14\}$, and this is the smallest such set. Later authors constructed infinite families, culminating in the work of Martin and O'Bryant, which proved a small positive proportion of subsets of $\{0,\dots,n\}$ are sum-dominant as $n \to \infty$, and Zhao, who estimated this percentage at around $4.5 \cdot 10^{-4}$. See [3, pp. 172–174] and [6, 7, 9, 10, 15, 16, 17, 18, 19, 23] for general overviews, examples, constructions, bounds on percentages and some generalizations, [11, 13, 12, 21] for some explicit constructions of infinite families of sum-dominant sets, and [1, 2, 14, 22] for some extensions to other settings.

Much of the above work looks at finite subsets of the natural numbers, or equivalently subsets of $\{0, 1, \dots, n\}$ as $n \to \infty$. We investigate the effect of restricting the initial set on the existence of sum-dominant subsets. In particular, given an infinite set $A = \{a_k\}_{=1}^{\infty}$, when does A have no sum-dominant subsets, only finitely many sum-dominant subsets, or infinitely many sum-dominant subsets? *We assume throughout the rest of the paper that every such sequence A is strictly increasing and non-negative.*

Our first result shows that if the sequence grows sufficiently rapidly and there are no 'small' subsets which are sum-dominant, then there are no sum-dominant subsets.

Theorem 1. *Let $A = \{a_k\}_{k=1}^{\infty}$ be a strictly increasing sequence of non-negative numbers. If there exists a positive integer r such that*

1. $a_k > a_{k-1} + a_{k-r}$ *for all* $k \geq r+1$, *and*
2. A *does not contain any sum-dominant set* S *with* $|S| \leq 2r-1$,

then A *contains no sum-dominant set.*

We prove this in §2. As the smallest sum-dominant set has 8 elements (see [6]), the second condition is trivially true if $r \leq 4$. In particular, we immediately obtain the following interesting result.

Corollary 2. *No subset of the Fibonacci numbers* $\{0, 1, 2, 3, 5, 8, \dots\}$ *is a sum-dominant set.*

The proof is trivial, and follows by taking $r = 3$ and noting

$$F_k \;=\; F_{k-1} + F_{k-2} \;>\; F_{k-1} + F_{k-3} \tag{3}$$

for $k \geq 4$.

After defining a class of subsets we present a partial result on when there are at most finitely many sum-dominant subsets.

Definition 3 (Special Sum-Dominant Set)**.** For a sum-dominant set S, we call S a special sum-dominant set if $|S+S| - |S-S| \geq |S|$.

We prove sum-dominant sets exist in §3.1. Note if S is a special sum-dominant set then if $S' = S \cup \{x\}$ for any sufficiently large x then S' is also a sum-dominant set. We have the following result about a sequence having at most finitely many sum-dominant sets (see §3 for the proof).

Theorem 4. *Let* $A = \{a_k\}_{k=1}^{\infty}$ *be a strictly increasing sequence of non-negative numbers. If there exists a positive integer* s *such that the sequence* $\{a_k\}$ *satisfies*

1. $a_k > a_{k-1} + a_{k-3}$ *for all* $k \geq s$, *and*
2. $\{a_1, \dots, a_{4s+6}\}$ *has no special sum-dominant subsets,*

then A *contains at most finitely many sum-dominant sets.*

The above results concern situations where there are not many sum-dominant sets; we end with an example of the opposite behavior.

Theorem 5. *There are infinitely many sum-dominant subsets of the primes.*

We will see later that this result follows immediately from the Green-Tao Theorem [4], which asserts that the primes contain arbitrarily long progressions. We also give a conditional proof in §4. There we assume the Hardy-Littlewood conjecture (see Conjecture 14) holds. The advantage of such an approach is that we have an explicit formula for the number of the needed prime tuples up to x, which gives a sense of how many such solutions exist in a given window.

2 Subsets with no sum-dominant sets

We prove Theorem 1, establishing a sufficient condition to ensure the non-existence of sum-dominant subsets.

Proof of Theorem 1. Let $S = \{s_1, s_2, \ldots, s_k\} = \{a_{g(1)}, a_{g(2)}, \ldots, a_{g(k)}\}$ be a finite subset of A, where $g : \mathbb{Z}^+ \to \mathbb{Z}^+$ is an increasing function. We show that S is not a sum-dominant set by strong induction on $g(k)$.

We proceed by induction. We show that if A has no sum-dominant subsets of size k, then it has no sum-dominant subsets of size $k+1$; as any sum-dominant set has only finitely many elements, this completes the proof.

For the Basis Step, we know (see [6]) that all sum-dominant sets have at least 8 elements, so any subset S of A with exactly k elements is not a sum-dominant set if $k \leq 7$; in particular, S is not a sum-dominant set if $g(k) \leq 7$. Thus we may assume for $g(k) \geq 8$ that all S' of the form $\{s_1, \ldots, s_{k-1}\}$ with $s_{k-1} < a_{g(k)}$ are not sum-dominant sets. The proof is completed by showing

$$S \;=\; S' \cup \{a_{g(k)}\} \;=\; \{s_1, \ldots, s_{k-1}, a_{g(k)}\} \tag{4}$$

is not sum-dominant sets for any $a_{g(k)}$.

We now turn to the inductive step. We know that S' is not a sum-dominant set by the inductive assumption. Also, if $k \leq 2r-1$ then $|S| \leq 2r-1$ and S is not a sum-dominant set by the second assumption of the theorem. If $k \geq 2r$, consider the number of new sums and differences obtained by adding $a_{g(k)}$. As we have at most k new sums, the proof is completed by showing there are at least k new differences.

Since $k \geq 2r$, we have $k - \lfloor \frac{k+1}{2} \rfloor \geq r$. Let $t = \lfloor \frac{k+1}{2} \rfloor$. Then $t \leq k - r$, which implies $s_t \leq s_{k-r}$. The largest difference in absolute value between elements in S is $s_{k-1} - s_1$; we now show that we have added at least $k+1$ distinct differences greater than $s_{k-1} - s_1$ in absolute value, which will complete the proof. We have

$$\begin{aligned} a_{g(k)} - s_t \;&\geq\; a_{g(k)} - s_{k-r} \;=\; a_{g(k)} - a_{g(k-r)} \\ &\geq\; a_{g(k)} - a_{g(k)-r} \\ &>\; a_{g(k)-1} - a_1 \qquad\qquad \text{(by the first assumption on } \{a_n\}\text{)} \\ &\geq\; s_{k-1} - a_1 \;\geq\; s_{k-1} - s_1. \end{aligned} \tag{5}$$

Since $a_{g(k)} - s_t \geq s_{k-1} - s_1$, we know that

$$a_{g(k)} - s_t,\ \ldots,\ a_{g(k)} - s_2,\ a_{g(k)} - s_1$$

are t differences greater than the greatest difference in S'. As we could subtract in the opposite order, S contains at least

$$2t \;=\; 2\left\lfloor \frac{k+1}{2} \right\rfloor \;\geq\; k \tag{6}$$

new differences. Thus $S+S$ has at most k more sums than $S'+S'$ but $S-S$ has at least k more differences compared to $S'-S'$. Since S' is not a sum-dominant set, we see that S is not a sum-dominant set. □

Remark 6. We thank the referee for the following alternative formulation of our proof. Given any infinite increasing sequence $\{a_{g(i)}\}$ that is a subset of a set A satisfying $a_k > a_{k_1} + a_{k-r}$ for all $k > r$, let $S_k = \{a_{g(1)}, \ldots, a_{g(k)}\}$ and $\Delta_k = |S_k - S_k| - |S_k + S_k|$. Similar arguments as above show that $\{\Delta_k\}$ is increasing for $k \geq 2r$.

We immediately obtain the following.

Corollary 7. *Let $A = \{a_k\}_{k=1}^{\infty}$ be a strictly increasing sequence of non-negative numbers. If $a_k > a_{k-1} + a_{k-4}$ for all $k \geq 5$, then A contains no sum-dominant subsets.*

Proof. From [6] we know that all sum-dominant sets have at least 8 elements. When $r = 4$ the second condition of Theorem 1 holds, completing the proof. □

For another example, we consider shifted geometric progressions.

Corollary 8. *Let $A = \{a_k\}_{k=1}^{\infty}$ with $a_k = c\rho^k + d$ for all $k \geq 1$, where $0 \neq c \in \mathbb{N}$, $d \in \mathbb{N}$, and $1 < \rho \in \mathbb{N}$. Then A contains no sum-dominant subsets.*

Proof. Without loss of generality we may shift and assume $d = 0$ and $c = 1$; the result now follows immediately from simple algebra. □

Remark 9. Note that if ρ is an integer greater than the positive root of $x^4 - x^3 - 1$ (the characteristic polynomial associated to $a_k = a_{k-1} + a_{k-4}$ from Theorem 4, which is approximately 1.3803) then the above corollary holds for $\{c\rho^k + d\}$.

3 Subsets with finitely many sum-dominant sets

We start with some properties of special sum-dominant sets, and then prove Theorem 4. The arguments are similar to those used in proving Theorem 1. *In this section, in particular in all the statements of the lemmas, we assume the conditions of Theorem 4 hold.* Thus $A = \{a_k\}_{k=1}^{\infty}$ and there is an integer s such that the sequence $\{a_k\}$ satisfies

1. $a_k > a_{k-1} + a_{k-3}$ for all $k \geq s$, and
2. $\{a_1, \ldots, a_{4s+6}\}$ has no special sum-dominant subsets.

3.1 Special sum-dominant sets

Recall a sum-dominant set S is special if $|S+S|-|S-S| \geq |S|$. For any $x \geq \sum_{a\in S} a$, adding x creates $|S|+1$ new sums and $2|S|$ new differences. Let $S^* = S \cup \{x\}$. Then

$$|S^*+S^*|-|S^*-S^*| \;\geq\; |S|+(|S|+1)-2|S| \;=\; 1, \tag{7}$$

and S^* is also a sum-dominant set. Hence, from one special sum-dominant set $S \subset \{a_n\}_{n=1}^{\infty} =: A$, we can generate infinitely many sum-dominant sets by adding any large integer in A. We immediately obtain the following converse.

Lemma 10. *If a set S is not a special sum-dominant set, then $|S+S|-|S-S| < |S|$, and by adding any large $x \geq \sum_{a\in S} a$, $S \cup \{x\}$ has at least as many differences as sums. Thus only finitely many sum-dominant sets can be generated by appending one integer from A to a non-special sum-dominant set S.*

Note that special sum-dominant sets exist. We use the base expansion method (see [6]), which states that given a set A, for all m sufficiently large if

$$A_t \;=\; \left\{\sum_{i=1}^{t} a_i m^{i-1} : a_i \in A\right\} \tag{8}$$

then

$$|A_t \pm A_t| \;=\; |A \pm A|^t; \tag{9}$$

the reason is that for m large the various elements are clustered with different pairs of clusters yielding well-separated sums. To construct the desired special sum-dominant set, consider the smallest sum-dominant set $S = \{0,2,3,4,\,7,\,11,\,12,14\}$. Using the method of base expansion, taking $m = 10^{2017}$ we obtain S_3 containing $|S_3| = 8^3 = 512$ elements such that $|S_3+S_3| = |S+S|^3 = 26^3 = 17576$ and $|S_3-S_3| = |S-S|^3 = 25^3 = 15625$. Then $|S_3+S_3|-|S_3-S_3| > |S_3|$.

3.2 Finitely many sum-dominant sets on a sequence

If a sequence $A = \{a_n\}_{n=1}^{\infty}$ contains a special sum-dominant set S, then we can get infinitely many sum-dominant subsets on the sequence just by adding sufficiently large elements of A to S. Therefore for a sequence A to have at most finitely many sum-dominant subsets, it is necessary that it has no special sum-dominant sets. Using the result from the previous subsection, we can prove Theorem 4.

We establish some notation before turning to the proof in the next subsection. We can write A as the union of $A_1 = \{a_1,\dots,a_{s-1}\}$ and $A_2 = \{a_s, a_{s+1},\dots\}$. We assume this is done with an $s \geq 5$ so that we can use Corollary 7, which implies that A_2 contains no sum-dominant sets. Thus any sum-dominant set must contain some elements from A_1.

We prove a lemma about A_2.

Lemma 11. *Let $S' = \{s_1, \dots, s_{k-1}\}$ be a subset of A containing at least 3 elements $a_{r_1}, a_{r_2}, a_{r_3}$ in A_2, with $r_3 > r_2 > r_1$. Consider the index $g(k) > r_3$, and let $S = S' \cup \{a_{g(k)}\}$. Then either S is not a sum-dominant set, or S satisfies $|S - S| - |S + S| > |S' - S'| - |S' + S'|$. Thus the excess of sums to differences from S is* less than *the excess from S'.*

Proof. We follow a similar argument as in Theorem 1.

If $k \le 7$, then S is not a sum-dominant set.

If $k \ge 8$, then $k - \lfloor \frac{k+3}{2} \rfloor \ge 3$. Let $t = \lfloor \frac{k+2}{2} \rfloor$. Then $t \le k-3$, and $s_t \le s_{k-3}$, and

$$\begin{aligned} a_{g(k)} - s_t &\ge a_{g(k)} - s_{k-3} = a_{g(k)} - a_{g(k-3)} \\ &\ge a_{g(k)} - a_{g(k)-3} \\ &> a_{g(k)-1} = a_{g(k)-1} - a_1 \qquad \text{(by assumption on } a\text{)} \\ &\ge s_{k-1} - a_1 \ge s_{k-1} - s_1. \end{aligned} \tag{10}$$

In the set S', the greatest difference is $s_{k-1} - s_1$. Since $a_{g(k)} - s_t \ge s_{k-1} - s_1$, we know that $a_{g(k)} - s_t, \dots, a_{g(k)} - s_2, a_{g(k)} - s_1$ are all differences greater than the greatest difference in S'.

By a similar argument, $s_t - a_{g(k)}, \dots, s_2 - a_{g(k)}, s_1 - a_{g(k)}$ are all differences smaller than the smallest difference in S'.

So S contains at least $2t = 2\lfloor \frac{k+3}{2} \rfloor > 2 \cdot \frac{k+1}{2} = k + 1$ new differences compared to S', and S satisfies

$$|S - S| - |S + S| \ > \ |S' - S'| - |S' + S'|, \tag{11}$$

completing the proof. □

3.3 Proof of Theorem 4

Recall that we write $A = A_1 \cup A_2$ with $A_1 = \{a_1, \dots, a_{s-1}\}$, $A_2 = \{a_s, a_{s+1}, \dots\}$, and by Corollary 7 A_2 contains no sum-dominant sets (thus any sum-dominant set must contain some elements from A_1). We first prove a series of useful results which imply the main theorem.

Our first result classifies the possible sum-dominant subsets of A. Since any such set must have at least one element of A_1 in it but not necessarily any elements of A_2, we use the subscript n below to indicate how many elements of A_2 are in our sum-dominant set.

Lemma 12 (Classification of Sum-Dominant Subsets of A)**.** *Notation as above, let K_n be a sum-dominant subset of $A = A_1 \cup A_2$ with n elements in A_2. Thus we may write*

$$K_n \ = \ S \cup \{a_{r_1}, \dots, a_{r_n}\}$$

for some

$$S \ \subset \ A_1 \ = \ \{a_1, \dots, a_{s-1}\}, \quad s \le r_1 < r_2 < \cdots < r_n.$$

Set

$$d \ = \ \max_{K_3}(|K_3 + K_3| - |K_3 - K_3|, 1).$$

Then $n \leq d+3$. In other words, a sum-dominant subset of A can have at most $d+3$ elements of A_2.

Proof. Let S_m be any subset of A with m elements of A_2. Lemma 11 tells us that for any S_m with $m \geq 3$, when we add any new element $a_{r_{m+1}}$ to get S_{m+1}, either S_{m+1} is not a sum-dominant set, or

$$|S_{m+1} - S_{m+1}| - |S_{m+1} + S_{m+1}| \ \geq \ |S_m - S_m| - |S_m + S_m| + 1.$$

For an $n > d+3$, assume there exists a sum-dominant set; if so, denote it by K_n. For $3 \leq k \leq n$, define S_k as the set obtained by deleting the $(n-k)$ largest elements from K_n (equivalently, keeping only the k smallest elements from K_n which are in A_2). We prove that each S_k is sum-dominant, and then show that this forces S_n not to be sum-dominant; this contradiction proves the theorem as $K_n = S_n$.

If S_k is not a sum-dominant set for any $k \geq 3$, by Lemma 11 either S_{k+1} is not a sum-dominant set, or

$$|S_{k+1} - S_{k+1}| - |S_{k+1} + S_{k+1}| \ \geq \ |S_k - S_k| - |S_k + S_k| + 1 \ \geq \ 0,$$

in which case S_{k+1} is also not a sum-dominant set (because S_k is not sum-dominant, the set S_{k+1} generates at least as many differences as sums). As we are assuming K_n (which is just S_n) is a sum-dominant set, we find S_{n-1} is sum-dominant. Repeating the argument, we find that S_{n-2} down to S_3 must also all be sum-dominant sets, and we have

$$|S_n - S_n| - |S_n + S_n| \ \geq \ |S_3 - S_3| - |S_3 + S_3| + (n-3). \tag{12}$$

Since S_3 is one of the K_3's (i.e., it is a sum-dominant subset of A with exactly three elements of A_2), by the definition of d the right hand side above is at least $n - 3 - d$. As we are assuming $n > d+3$ we see it is positive, and hence S_n is not sum-dominant. As $S_n = K_n$ we see that K_n is not a sum-dominant set, contradicting our assumption that there is a sum-dominant set K_n with $n > d+3$, proving the theorem. □

Lemma 13. *For $n \geq 0$, let k_n denote the number of subsets $K_n \subset A$ which are sum-dominant and contain exactly n elements from A_2. We write*

$$K_n \ = \ S \cup \{a_{r_1}, \dots, a_{r_n}\} \quad \text{with } S \subset A_1. \tag{13}$$

Then

1. *k_n is finite for all $n \geq 0$, and*

2. *every K_n is not a special sum-dominant set.*

Proof. We prove each part by induction. It is easier to do both claims simultaneously as we induct on n. We break the analysis into $n \in \{0, 1, 2, 3\}$ and $n \geq 4$. The proof for $n = 0$

is immediate, while $n \in \{1, 2, 3\}$ follow by obtaining bounds on the indices permissible in a K_n, and then $n \geq 4$ follows by induction. We thus must check (1) and (2) for $n \leq 3$. While the arguments for $n \leq 3$ are all similar, it is convenient to handle each case differently so we can control the indices and use earlier results, in particular removing the largest element in A_2 yields a set which is not a special sum-dominant set.

Case $n = 0$: As A_1 is finite, it has finitely many subsets and thus k_0, which is the number of sum-dominant subsets of A_1, is finite (it is at most $2^{|A_1|}$). Further any K_0 is a subset of

$$A_1 \;=\; \{a_1, \dots, a_{s-1}\},$$

which is a subset of

$$A' \;=\; \{a_1, \dots, a_{4s+6}\}. \tag{14}$$

As we have assumed A' has no special sum-dominant set, no K_0 can be a special sum-dominant set.

Case $n = 1$: We start by obtaining upper bounds on r_1, the index of the smallest (and only) element in our set coming from A_2. Consider the index $4s$. We claim that

$$a_{4s} \;>\; \sum_{a \in A_1} a. \tag{15}$$

This is because $|A_1| < s$ and $a_k > a_{k-1} + a_{k-3}$ for all $k \geq s$, and hence

$$\begin{aligned}
\sum_{a \in A_1} a \;&<\; s \cdot a_s \\
&<\; \frac{s}{2}\left(a_s + a_{s+2}\right) \;<\; \frac{s}{2} \cdot a_{s+3} \\
&<\; \frac{s}{4}\left(a_{s+3} + a_{s+5}\right) \;<\; \frac{s}{4} \cdot a_{s+6} \dots \\
&<\; \frac{s}{2^{\lceil \log_2 s \rceil}} a_{s+3\lceil \log_2(s) \rceil} \\
&<\; a_{s+3s} \;=\; a_{4s}
\end{aligned}$$

(by doing the above $\lceil \log_2 s \rceil$ times we ensure that $s/2^{\lceil \log_2 s \rceil} < 1$, and since $s \geq 1$ we have $3s \geq 3\lceil \log_2(s) \rceil$). Therefore for all r_1 sufficiently large,

$$a_{r_1} \;>\; a_{4s} \;>\; \sum_{a \in A_1} a. \tag{16}$$

Clearly there are only finitely many sum-dominant subsets K_1 with $r_1 \leq 4s$; the analysis is completed by showing there are no sum-dominant sets with $r_1 > 4s$. Imagine there was a sum-dominant K_1 with $a_{r_1} > a_{4s}$. Then K_1 is the union of a set of elements $S = \{s_1, \dots, s_m\}$ in A_1 and a_{r_1} in A_2. As $\sum_{s \in S} s < a_{r_1}$, by Lemma 10 we find K_1 is not a sum-dominant set.

All that remains is to show none of the K_1 are special sum-dominant sets. This is immediate, as each sum-dominant K_1 is a subset of $\{a_1, \ldots, a_{4s}\}$, which is a subset of A' (defined in (14)). As we have assumed A' has no special sum-dominant set, no K_1 can be a special sum-dominant set.

Case $n = 2$: Consider the index $4s+3$. If K_2 is a sum-dominant set then it has two elements, $a_{r_1} < a_{r_2}$, that are in A_2. We show that if $r_2 \geq 4s+3$ then there can be no sum-dominant sets, and thus there are only finitely many K_2.

For all $r_2 \geq 4s+3$,

$$a_{r_2} - a_{r_2-1} \ > \ a_{r_2-3} \ \geq \ a_{4s} \ > \sum_{a \in A_1} a. \tag{17}$$

Assume there is a sum-dominant K_2 with $r_2 \geq 4s+3$. It contains some elements $S = \{s_1, \ldots, s_m\}$ in A_1 and a_{r_1}, a_{r_2} in A_2. We have

$$a_{r_2} - a_{r_1} \ \geq \ a_{r_2} - a_{r_2-1} \ > \sum_{a \in S} a.$$

Therefore $a_{r_2} > \left(\sum_{a \in S} a\right) + a_{r_1}$, and $S \cup \{a_{r_1}\}$ is not a special sum-dominant set by the $n = 1$ case[1]. Hence, by Lemma 10 we find $K_2 = (S \cup \{a_{r_1}\}) \cup \{a_{r_2}\}$ is not a sum-dominant set.

Finally, as K_2 is a subset of $\{a_1, \ldots, a_{4s+1}\}$, which is a subset of A', by assumption K_2 is not a special sum-dominant set.

Case $n = 3$: Let K_3 be a sum-dominant set with three elements from A_2. We show that if $r_3 \geq 4s+6$ then there are no such K_3; as there are only finitely many sum-dominant sets with $r_3 < 4s+6$, this completes the counting proof in this case.

Consider the index $4s+6$. For all $r_3 \geq 4s+6$,

$$a_{r_3-3} - a_{r_3-4} \ > \ a_{r_3-6} \ \geq \ a_{4s} \ > \sum_{a \in A_1} a. \tag{18}$$

Consider any K_3 with $r_3 \geq 4s+6$. We write K_3 as $S \cup \{a_{r_1}, a_{r_2}, a_{r_3}\}$ and $S \subset A_1$. If $|S| < 5$, we know that $|K_3| < 8$, and K_3 is not a sum-dominant set as such a set has at least 8 elements. We can therefore assume that $|S| \geq 5$. We have two cases.

Subcase 1: $r_2 \leq r_3 - 3$: Thus

$$a_{r_3} - a_{r_2} - a_{r_1} \ \geq \ a_{r_3} - a_{r_3-3} - a_{r_3-4} \ \geq \ a_{r_3-1} - a_{r_3-4} \ \geq \ a_{r_3-2} > a_{r_3-6} \ > \sum_{a \in S} a.$$

[1] If $S' = S \cup \{a_{r_1}\}$ is sum-dominant then it is not special, while if it is not sum-dominant then clearly it is not a special sum-dominant set.

As $S \cup \{a_{r_1}, a_{r_2}\}$ is not a special sum-dominant set by the $n = 2$ case[2], adding a_{r_3} with

$$a_{r_3} \ > \ \left(\sum_{s \in S} s\right) + a_{r_1} + a_{r_2}$$

creates a non-sum-dominant set by Lemma 10.

Subcase 2: $r_2 > r_3 - 3$: Using (18) we find

$$a_{r_3} - a_{r_2} \ \geq \ a_{r_3} - a_{r_3-1} \ > \ \sum_{a \in S} a$$

and

$$a_{r_2} - a_{r_1} \ > \ a_{r_3-2} - a_{r_3-3} \ > \ \sum_{a \in S} a.$$

Therefore the differences between a_{r_1}, a_{r_2}, a_{r_3} are large relative to the sum of the elements in S, and our new sums and new differences are well-separated from the old sums and differences. Explicitly, $K_3 + K_3$ consists of $S + S$, $a_{r_1} + S$, $a_{r_2} + S$, $a_{r_3} + S$, plus at most 6 more elements (from the sums of the a_r's), while $K_3 - K_3$ consists of $S - S$, $\pm(a_{r_1} - S)$, $\pm(a_{r_2} - S)$, $\pm(a_{r_3} - S)$, plus possibly some differences from the differences of the a_r's.

As S is not a special sum-dominant set, we know $|S + S| - |S - S| < |S|$ (if S is not sum-dominant the claim holds trivially, while if it is sum-dominant it holds because S is not special). Thus for K_3 to be sum-dominant, we must have

$$\begin{aligned} 0 \ &< \ |K_3 + K_3| - |K_3 - K_3| \\ &\leq \ (|S + S| + 3|S| + 6) - (|S - S| + 6|S|) \\ &< \ 6 - 2|S|; \end{aligned}$$

as $|S| \geq 5$ this is impossible, and thus K_3 cannot be sum-dominant.

Finally, as again K_3 is a subset of $A' = \{a_1, \dots, a_{4s+6}\}$, no K_3 is a special sum-dominant set.

Case $n \geq 4$ (inductive step): We proceed by induction. We may assume that k_n is finite for some $n \geq 3$, and must show that k_{n+1} is finite. By the earlier cases we know there is an integer t_n such that if K_n is a sum-dominant subset of A with exactly n elements of A_2, then the largest index r_n of an $a_i \in K_n$ is less than t_n.

We claim that if K_{n+1} is a sum-dominant subset of A then each index is less than t_{n+1}, where t_{n+1} is the smallest index such that if $r_{n+1} \geq t_{n+1}$ then

$$a_{r_{n+1}} \ > \ \sum_{i < r_n} a_i. \tag{19}$$

[2]As before, if it is sum-dominant it is not special, while if it is not sum-dominant it cannot be sum-dominant special; thus we have the needed inequalities concerning the sizes of the sets.

We write

$$K_{n+1} \ = \ S \cup \{a_{r_1}, \ldots, a_{r_n}, a_{r_{n+1}}\}, \quad S \subset A_1, \quad \{a_{r_1}, \ldots, a_{r_n}\} \subset A_2.$$

We show that if $r_{n+1} \geq t_{n+1}$ then K_{n+1} is not sum-dominant. Let $S_n = K_{n+1} \setminus \{a_{r_{n+1}}\}$. We have two cases.

- If $r_n < t_n$, then by the inductive hypothesis S_n is not a special sum-dominant set. So adding $a_{r_{n+1}} > \sum_{x \in S_n} x$ to S_n gives a non-sum-dominant set by Lemma 10.

- If $r_n \geq t_n$, then by the inductive hypothesis S_n is not a sum-dominant set. So $|S_n - S_n| - |S_n + S_n| \geq 0$. Since $n \geq 3$, we can apply Lemma 11, and either $K_{n+1} = S_n \cup \{a_{r_{n+1}}\}$ is not a sum-dominant set, or

$$|K_{n+1} - K_{n+1}| - |K_{n+1} + K_{n+1}| \ > \ |K_n - K_n| - |K_n + K_n| > 0,$$

in which case S_{n+1} is still not a sum-dominant set.

We conclude that for all sum-dominant sets S_{n+1}, we must have $r_{n+1} < t_{n+1}$. So k_{n+1} is finite.

Consider any sum-dominant set $K_{n+1} = S_n \cup \{a_{r_{n+1}}\}$. Applying lemma 11 again, we have $|K_{n+1} - K_{n+1}| - |K_{n+1} + K_{n+1}| > |S_n - S_n| - |S_n + S_n|$. We know, from inductive hypothesis, that S_n is not a special sum-dominant set. Therefore all possible K_{n+1} are not special sum-dominant sets.

By induction, k_n is finite for all $n \geq 0$, and all K_n are not special sum-dominant sets. $\square$

Proof of Theorem 4. By Lemma 12 every sum-dominant subset of A is of the form K_0, K_1, K_2, ..., K_{d+3} where the K_n are as in (13). By Lemma 13 there are only finitely many sets of the form K_n for $n \leq d+3$, and thus there are only finitely many sum-dominant subsets of A. $\square$

4 Sum-dominant subsets of the prime numbers

We now investigate sum-dominant subsets of the primes. While Theorem 5 follows immediately from the Green-Tao theorem, we first conditionally prove there are infinitely many sum-dominant subsets of the primes as this argument gives a better sense of what the 'truth' should be (i.e., how far we must go before we find sum-dominant subsets).

4.1 Admissible prime tuples and prime constellations

We first consider the idea of prime m-tuples. A prime m-tuple $(b_1, b_2, \ldots, b_m)$ represents a pattern of differences between prime numbers. An integer n matches this pattern if $(b_1 + n, b_2 + n, \ldots, b_m + n)$ are all primes.

A prime m-tuple $(b_1, b_2, \ldots, b_m)$ is called admissible if for all integers $k \geq 2$, $\{b_1, b_2, \ldots, b_m\}$ does not cover all values modulo k. If a prime m-tuple is not admissible, whenever $n > k$ then at least one of $b_1 + n, b_2 + n, \ldots, b_m + n$ is divisible by k and greater than k, so this cannot be an m-tuple of prime numbers (in this case the only n which can lead to an m-tuple of primes are $n \leq k$, and there are only finitely many of these).

It is conjectured in [5] that all admissible m-tuples are matched by infinitely many integers.

Conjecture 14 (Hardy-Littlewood [5])**.** Let $b_1, b_2, \ldots, b_m$ be m distinct integers, $v_p(b) = v(p; b_1, b_2, \ldots, b_m)$ the number of distinct residues of $b_1, b_2, \ldots b_m$ to the modulus p, and $P(x; b_1, b_2, \ldots, b_m)$ the number of integers $1 \leq n \leq x$ such that every element in $\{n + b_1, n + b_2, \ldots, n + b_m\}$ is prime. Assume $(b_1, b_2, \ldots, b_m)$ is admissible (thus $v_p(b) \neq p$ for all p). Then

$$P(x) \ \sim \ \mathfrak{S}(b_1, b_2, \ldots, b_m) \int_2^x \frac{du}{(\log u)^m} \tag{20}$$

when $x \to \infty$, where

$$\mathfrak{S}(b_1, b_2, \ldots, b_m) \ = \ \prod_{p \geq 2} \left(\left(\frac{p}{p-1} \right)^{m-1} \frac{p - v_p(b)}{p-1} \right) \ \neq \ 0.$$

As $(b_1, b_2, \cdots, b_m)$ is an admissible m-tuple, $v(p; b_1, b_2, \ldots, b_m)$ is never equal to p and equals m for $p > \max\{|b_i - b_j|\}$. The product $\mathfrak{S}(b_1, b_2, \ldots, b_m)$ thus converges to a positive number as each factor is non-zero and is $1 + O_m(1/p^2)$. Therefore this conjecture implies that every admissible m-tuple is matched by infinitely many integers.

4.2 Infinitude of sum-dominant subsets of the primes

We now show the Hardy-Littlewood conjecture implies there are infinitely many subsets of the primes which are sum-dominant sets.

Theorem 15. *If the Hardy-Littlewood conjecture holds for all admissible m-tuples then the primes have infinitely many sum-dominant subsets.*

Proof. Consider the smallest sum-dominant set $S = \{0, 2, 3, 4, 7, 11, 12, 14\}$. We know that $\{p, p+2s, p+3s, p+4s, p+7s, p+11s, p+12s, p+14s\}$ is a sum-dominant set for all positive integers p, s. Set $s = 30$ and let $T = (0, 60, 90, 120, 210, 330, 360, 420)$. We deduce that if there are infinitely many n such that $n + T = (n, n+60, n+90, n+120, n+210, n+330, n+360, n+420)$ is an 8-tuple of prime numbers, then there are infinitely many sum-dominant sets of prime numbers.

We check that T is an admissible prime 8-tuple. When $m > 8$, the eight numbers in T clearly don't cover all values modulo m. When $m \leq 8$, one sees by straightforward computation that T does not cover all values modulo m.

By Conjecture 14, there are infinitely many integers p such that every element of $\{p, p+60, p+90, p+120, p+210, p+330, p+360, p+420\}$ is prime. These are all sum-dominant sets, so there are infinitely many sum-dominant sets on primes. $\square$

Of course, all we need is that the Hardy-Littlewood conjecture holds for one admissible m-tuple which has a sum-dominant subset. We may take $p = 19$, which gives an explicit sum-dominant subset of the primes: $\{19, 79, 109, 139, 229, 349, 379, 439\}$ (a natural question is which sum-dominant subset of the primes has the smallest diameter). If one wishes, one can use the conjecture to get some lower bounds on the number of sum-dominant subsets of the primes at most x. The proof of Theorem 5 follows similarly.

Proof of Theorem 5. By the Green-Tao theorem, the primes contain arbitrarily long arithmetic progressions. Thus for each $N \geq 14$ there are infinitely many pairs (p, d) such that

$$\{p, p+d, p+2d, \dots, p+Nd\} \tag{21}$$

are all prime. We can then take subsets as in the proof of Theorem 15. $\square$

5 Future work

We list some natural topics for further research.

- Can the conditions in Theorem 1 or 4 be weakened?
- What is the smallest special sum-dominant set by diameter, and by cardinality?
- What is the smallest, in terms of its largest element, set of primes that is sum-dominant?

6 Acknowledgments

The first named author was supported by Washington & Lee University, and the third named author was partially supported by NSF grants DMS1265673 and DMS1561945. We thank the students from the Math 21-499 Spring '16 research class at Carnegie Mellon and the participants from CANT 2016 for many helpful conversations, Angel Kumchev for comments on an earlier draft, and the referee for many suggestions which improved and clarified the exposition, as well as suggestions on good topics for future research.

References

[1] T. Do, A. Kulkarni, S. J. Miller, D. Moon, and J. Wellens, Sums and differences of correlated random sets, *J. Number Theory* **147** (2015), 44–68.

[2] T. Do, A. Kulkarni, S. J. Miller, D. Moon, J. Wellens, and J. Wilcox, Sets characterized by missing sums and differences in dilating polytopes, *J. Number Theory* **157** (2015), 123–153.

[3] G. A. Freiman and V. P. Pigarev, *Number Theoretic Studies in the Markov Spectrum and in the Structural Theory of Set Addition* (Russian), Kalinin. Gos. Univ., Moscow, 1973.

[4] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Ann. of Math.* **167** (2008), 481–547.

[5] G. H. Hardy and J. Littlewood, Some problems of 'partitio numerorum'; III: on the expression of a number as a sum of primes, *Acta Math.* **44** (1923), 1–70.

[6] P. V. Hegarty, Some explicit constructions of sets with more sums than differences, *Acta Arith.* **130** (2007), 61–77.

[7] P. V. Hegarty and S. J. Miller, When almost all sets are difference dominated, *Random Structures Algorithms* **35** (2009), 118–136.

[8] G. Iyer, O. Lazarev, S. J. Miller, and L. Zhang, Generalized more sums than differences sets, *J. Number Theory* **132** (2012), 1054–1073.

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

[10] 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., 2007, pp. 287–305.

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

[12] S. J. Miller and D. Scheinerman, Explicit constructions of infinite families of MSTD sets, in *Additive Number Theory*, Springer, 2010, pp. 229–248.

[13] S. J. Miller, S. Pegado, and L. Robinson, Explicit constructions of large families of generalized more sums than differences sets, *Integers* **12** (2012), #A30.

[14] S. J. Miller and K. Vissuet, Most subsets are balanced in finite groups, in M. B. Nathanson, ed., *Combinatorial and Additive Number Theory — CANT 2011 and 2012*, Springer Proceedings in Mathematics & Statistics, Vol. 101, 2014, pp. 147–157.

[15] M. B. Nathanson, Sums of finite sets of integers, *Amer. Math. Monthly* **79** (1972), pp. 1010–1012.

[16] M. B. Nathanson, Problems in additive number theory. I, in *Additive Combinatorics*, CRM Proc. Lecture Notes, Vol. 43, *Amer. Math. Soc.*, 2007, pp. 263–270.

[17] M. B. Nathanson, Sets with more sums than differences, *Integers* **7** (2007), Paper #A5.

[18] I. Z. Ruzsa, On the cardinality of $A + A$ and $A - A$, in *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, *Coll. Math. Soc. J. Bolyai.*, Vol. 18, 1978, pp. 933–938.

[19] I. Z. Ruzsa, Sets of sums and differences, in *Seminar on Number Theory, Paris 1982–83*, Progr. Math., Vol. 51, Birkhäuser, 1984, pp. 267–273.

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

[21] Y. Zhao, Constructing MSTD sets using bidirectional ballot sequences, *J. Number Theory* **130** (2010), 1212–1220.

[22] Y. Zhao, Counting MSTD sets in finite abelian groups, *J. Number Theory* **130** (2010), 2308–2322.

[23] Y. Zhao, Sets characterized by missing sums and differences, *J. Number Theory* **131** (2011), 2107–2134.

2010 *Mathematics Subject Classification*: Primary 11P99; Secondary 11K99.
Keywords: MSTD set.

Received June 1 2018; revised versions received August 21 2018; November 13 2018; November 16 2018. Published in *Journal of Integer Sequences*, November 23 2018.

Return to Journal of Integer Sequences home page.