GENERALIZED NONAVERAGING INTEGER SEQUENCES

Dennis Tseng
Massachusetts Institute of Technology, Cambridge, MA 02139, United States
dtseng@mit.edu

Abstract

Let the sequence S_m of nonnegative integers be generated by the following conditions: Set the first term $a_0 = 0$, and for all $k \geq 0$, let a_{k+1} be the least integer greater than a_k such that no element of $\{a_0, \ldots, a_{k+1}\}$ is the average of $m-1$ *distinct* other elements. Szekeres gave a closed-form description of S_3 in 1936, and Layman provided a similar description for S_4 in 1999. We first find closed forms for some similar greedy sequences that avoid averages in terms not *all the same*. Then, we extend the closed-form description of S_m from the known cases when $m = 3$ and $m = 4$ to any integer $m \geq 3$. With the help of a computer, we also generalize this to sequences that avoid solutions to specific weighted averages in distinct terms. Finally, from the closed forms of these sequences, we find bounds for their growth rates.

1. Introduction

Often in combinatorial number theory, we wish to find the maximum number of integers that can be chosen from $\{0, 1, \ldots, n-1\}$ without creating a solution to some linear equation in the chosen integers. Ruzsa initiated a systematic study of this problem over all linear equations [7, 8], and the problem has also been extended to systems of linear equations [4, 9]. A couple well-studied examples include constructing sets of integers without three-term arithmetic progressions, which corresponds to avoiding solutions to $x_1+x_2-2x_3 = 0$, and constructing Sidon sets, which are defined by having no nontrivial solutions to $x_1 + x_2 - x_3 - x_4 = 0$. One way to approach this problem is through the use of a greedy algorithm.

Given an integer $m \geq 3$, define the sequence S_m of nonnegative integers by the following conditions:

(i) $a_0 = 0$

(ii) Having chosen $a_0, a_1, \ldots, a_k$, let a_{k+1} be the least integer greater than a_k such that there are no *distinct* $x_1, x_2, \ldots, x_m \in \{a_0, a_1, \ldots, a_{k+1}\}$ with

$$x_1 + \cdots + x_{m-1} = (m-1)x_m.$$

The sequence S_m constructs a sequence of integers that avoids solutions to $x_1 + \cdots + x_{m-1} = (m-1)x_m$ using a greedy algorithm. Generating S_3, which avoids three-term arithmetic progressions, we obtain

$0, 1, 3, 4, 9, 10, 12, 13, 27, 28, 30, 31, 36, 37, 39, 40, 81 \ldots$

There is an alternative definition for S_3. An integer is in S_3 if and only if there is no 2 in its representation in base 3. This follows from a more general result, as Erdős and Turán [3] wrote that Szekeres showed the use of the greedy algorithm to avoid m-term arithmetic progressions, for m prime, results in a sequence that contains the integers that do not contain the digit $m-1$ when expressed in base m.

The nice closed-form description suggests that we can extend this to more general averages. The sequence S_4 has a similar closed-form description as S_3. The following theorem is due to Layman [5].

Theorem 1. *An integer is in S_4 if and only if it can be written in the form $M+r$, where the base 4 representation of M has only 3's and 0's and ends with a 0, and r is any integer from 0 to 4 inclusive.*

Extending this generalization will form the basis of the rest of our investigation. In Section 2, we present the closed forms of some related sequences that avoid solutions to weighted averages in terms *not all the same*. Then in Section 3, we prove a result that can be used to find the closed forms of S_m for all $m \geq 3$ and the closed forms of sequences that avoid solutions to specific weighted averages. Finally in Section 4, given the closed forms, we can derive bounds that allows us to show how efficient the greedy algorithm is asymptotically.

1.1. Definitions

We make some definitions to simplify the notation for the rest of the paper. Unless otherwise stated, for an ordered tuple $E = (d_1, \ldots, d_{m-1})$, we will assume throughout the paper that $1 \le d_1 \le d_2 \ldots \le d_{m-1}$, i.e. the components are arranged in nondecreasing order. Let the ordered tuple $E_m = (1, 1, \ldots, 1)$, where there are $m - 1$ components in the tuple.

Definition 1. Given an ordered tuple $E = (d_1, \ldots, d_{m-1})$, let $d(E) = d_1 + \cdots + d_{m-1}$. When the choice of E is obvious, we will simply denote $d(E)$ as d.

Definition 2. Call an ordered tuple of positive integers $E = (d_1, \ldots, d_{m-1})$ *valid* if and only if the following conditions are satisfied:
(i) $1 = d_1$.
(ii) $d_2 \le d_1, d_3 \le d_1 + d_2, \ldots, d_{m-1} \le d_1 + \ldots + d_{m-2}$.
In particular, this implies that $d_1 = d_2 = 1$.

For example $E_3 = (1, 1)$. Also, E_m, for all $m \ge 3$, and $(1, 1, 2, 4, 8)$ are valid ordered tuples, while $(1, 1, 3)$ is not a valid ordered tuple.

1.2. Definition of Sequences

In this paper, we will focus on finding closed forms for the following sequences.

Definition 3. Given an ordered tuple $E = (d_1, \ldots, d_{m-1})$, define the sequence A_E of nonnegative integers by the following conditions:
(i) $a_0 = 0$
(ii) Having chosen $a_0, a_1, \ldots, a_k$, let a_{k+1} be the least integer greater than a_k such that there are no terms $x_1, x_2, \ldots, x_m \in \{a_0, \ldots, a_{k+1}\}$, *not all the same*, that satisfy $d_1x_1 + \cdots + d_{m-1}x_{m-1} = dx_m$.

Definition 4. Given an ordered tuple $E = (d_1, \ldots, d_{m-1})$, define the sequence S_E of nonnegative integers by the following conditions:
(i) $a_0 = 0$
(ii) Having chosen $a_0, a_1, \ldots, a_k$, let a_{k+1} be the least integer greater than a_k such that there are no *distinct* terms $x_1, x_2, \ldots, x_m \in \{a_0, \ldots, a_{k+1}\}$ that satisfy $d_1x_1 + \cdots + d_{m-1}x_{m-1} = dx_m$.

To simplify notation, we will refer to the sequences S_{E_m} and A_{E_m} for integer $m \ge 3$ as simply S_m and A_m respectively.

2. Analysis of the Sequences A_E

2.1. A Property of Valid Ordered Tuples

We will prove a property of valid ordered tuples that we will use throughout the paper.

Proposition 1. *An ordered tuple* $E = (d_1, \ldots, d_{m-1})$ *is valid if and only if for every integer* $0 \leq j \leq d-1$*, there exists a subset* H_j *of* $\{2, \ldots, m-1\}$ *such that* $\sum_{k \in H_j} d_k = j$.

Proof. We will show that, given a valid ordered tuple $E = (d_1, \ldots, d_{m-1})$, there exists a subset $H_j \subset \{2, \ldots, m-1\}$ for every integer $0 \leq j \leq d-1$ by induction. For the base case, $H_0 = \{\}$ and $H_1 = \{2\}$. Now, assume that for some integer $3 \leq l \leq m-1$, we have found a subset H_j of $\{2, \ldots, l-1\}$ for all $0 \leq j \leq \sum_{k=2}^{l-1} d_k$.

Let j be an integer with $1 + \sum_{k=2}^{l-1} d_k \leq j \leq \sum_{k=2}^{l} d_k$.

Then, let $H_j = H_{j-d_l} \cup \{l\}$. Since $d_l \leq \sum_{k=1}^{l-1} d_k \leq j \leq \sum_{k=2}^{l} d_k$, $0 \leq j - d_l \leq \sum_{k=2}^{l-1} d_k$ and H_{j-d_l} must exist by induction. Our induction is complete.

Now to prove the other direction, let $E = (d_1, \ldots, d_{m-1})$ be any ordered tuple of positive integers such that for every $0 \leq j \leq d-1$, there exists a subset H_j of $\{2, \ldots, m-1\}$ such that $\sum_{k \in H_j} d_k = j$. In order for H_1 to exist, $d_2 = 1$, which means $d_1 = 1$. Now, assume for the sake of contradiction that there is some integer $3 \leq l \leq m-1$ such that $d_l > \sum_{k=1}^{l-1} d_k$. Then, we cannot create the subset $H_{d_l - 1}$, because the subset H_{d_l-1} cannot contain any integers greater than $l-1$, or else $\sum_{k \in H_{d_l-1}} d_k > d_l - 1$. Also, by assumption, $\sum_{k=2}^{l-1} d_k < d_l - 1$, so the subset H_{d_l-1} cannot contain only integers less than or equal to $l-1$, which is a contradiction. □

2.2. Closed Form of A_E

Theorem 2. *Given a valid ordered tuple* E*, an integer is in* A_E *if and only if it contains only 0's and 1's in its base* $d+1$ *representation.*

Proof. Let the sequence B_E be the nonnegative integers with only 0's and 1's in their base $d+1$ representation in increasing order. We show that B_E is the same as A_E. Let $E = (d_1, \ldots, d_{m-1})$.

Lemma 2.1. *It is impossible to choose m integers $x_1, x_2, \ldots, x_m$, not all the same, that are terms of the sequence B_E such that*

$$d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m. \tag{1}$$

Proof. Assume for the sake of contradiction that there are $x_1, \ldots, x_m$, not all equal, that satisfy equation (1). Let $t_{0,k}, t_{1,k}, \ldots$ be the digits of x_k in base $d+1$, i.e. $x_k = \sum_{i=0}^{\infty} t_{i,k}(d+1)^i$ for all $1 \leq k \leq m$. From equation (1),

$$\sum_{k=1}^{m-1} \sum_{i=0}^{\infty} d_k t_{i,k}(d+1)^i = d \sum_{i=0}^{\infty} t_{i,m}(d+1)^i.$$

There is no carrying in base $d+1$ when we add $\sum_{k=1}^{m-1} d_k x_k$ because x_k contains only 0's and 1's in its base $d+1$ representation for all $1 \leq k \leq m$ and $\sum_{k=1}^{m-1} d_k < d+1$. Therefore, if $t_{i,m} = 0$, then $t_{i,k} = 0$ for all $1 \leq k \leq m-1$. If $t_{i,m} = 1$, then $t_{i,k} = 1$ for all $1 \leq k \leq m-1$. But then $x_1 = x_2 = \ldots = x_m$ contradicting the condition that $x_1, x_2, \ldots, x_m$ cannot all be the same. □

Now we show it is impossible to insert terms into B_E, which means B_E satisfies the "greedy" condition of A_E.

Lemma 2.2. *Given any integer x_1 that is not in B_E, we can find terms $x_2, x_3, \ldots, x_m$ of B_E, each less than x_1 such that $d_1 x_1 + d_2 x_2 + \cdots + d_{m-1} x_{m-1} = d x_m$.*

Proof. Since E is a valid ordered tuple, by Proposition 1, for every $0 \leq j \leq d-1$, there exists a set $H_j \subset \{2, \ldots, m-1\}$ such that $\sum_{k \in H_j} d_k = j$. Let $t_{0,k}, t_{1,k}, \ldots$ be the digits of x_k in base $d+1$, i.e. $x_k = \sum_{i=0}^{\infty} t_{i,k}(d+1)^i$ for all $1 \leq k \leq m$. For every $i \geq 0$, if $t_{i,1} = 0$, then let $t_{i,k} = 0$ for all $2 \leq k \leq m-1$. If $t_{i,1} > 0$, let $t_{i,k} = 1$ for all $k \in H_{d-t_{i,1}}$ and $t_{i,k} = 0$ for all $k \notin H_{d-t_{i,1}}$ so that $\sum_{k=1}^{m-1} d_k t_{i,k} = d$. Then, the sum $\sum_{k=1}^{m-1} d_k x_k$ has only 0's and d's when written in base $d+1$. When we divide the sum $\sum_{k=1}^{m-1} d_k x_k$ by d, we obtain an integer that has only 0's and 1's when written in base $d+1$, which is in B_E. Note that $t_{i,1}$ must be greater than 1 for some $i = i_0$ as x_1 is not a term of B_E. Then, $t_{i_0,1} > t_{i_0,k}$ for all $2 \leq k \leq m$. Since $t_{i,1} \geq t_{i,k}$ for all $2 \leq k \leq m$ and $i \geq 0$, $x_1 > x_2, \ldots, x_m$ as desired. □

Since we have proven no m terms in B_E satisfy the equation $d_1x_1 + \cdots + d_{m-1}x_{m-1} = dx_m$ and no terms can be inserted into B_E without creating a solution to the equation, B_E is the same sequence as A_E. □

2.3. A Property of the Sequence A_E

By Theorem 2, the term a_n of A_E can be found by writing n in binary and reading it in base $d+1$. Then, the following result quickly follows.

Proposition 2. *The number of 1's in the base 2 representation of n is congruent modulo d to the n^{th} term of A_E.*

Proof. Write $n = \sum_{i=0}^{\infty} t_i 2^i$, with $t_0, t_1, \ldots$ as its digits in base 2. Then, $a_n = \sum_{i=0}^{\infty} t_i(d+1)^i \equiv \sum_{i=0}^{\infty} t_i \pmod{d}$. □

Corollary 1. *The terms of A_3 modulo 2 is the* Thue-Morse sequence*, where the n^{th} term is a 0 if n has an even number of 1's in its binary expansion and a 1 otherwise by Proposition 1 in [1].*

3. Analysis of the Sequences S_E

We first give an alternative way to represent the nonnegative integers.

Proposition 3. *Given positive integers $M \geq 2$ and c, every nonnegative integer x can be expressed in the form $x = c\sum_{i=0}^{\infty} t_i M^i + r$ in exactly one way, with integer $0 \leq r < c$ and sequence $t_0, t_1, \ldots$ such that $t_i \in \{0, \ldots, M-1\}$ for all $i \geq 0$.*

Proof. Given a positive integer x, let r_0 and m_0 be the remainder and quotient when x is divided by c. So $x = r_0 + cm_0$ and r_0 and m_0 are uniquely defined. Then $r = r_0$, and the digits of m_0 in base M is the sequence $t_0, t_1, \ldots$, which also must be uniquely defined. □

We now present our main result, which can be used to find closed forms of the sequences S_E for specific choices of E.

Theorem 3. *For some positive integer z and some sequence S_E for valid ordered tuple E, let the set R_E be $\{a_0, \ldots, a_z\}$ and the constant $c_E = a_{z+1}$.*
Let $\max(R_E)$ denote the maximum element a_z. Suppose the following conditions (i) and (ii) are satisfied:

(i) $c_E = 1 + d\max(R_E) - \sum_{k=2}^{m-1} d_k(m-k-1)$.

(ii) For every integer $0 \le r_1 \le c_E - 1$ *and every integer* $0 \le j \le d-2$*, there exists a subset* H_j *of* $\{2, \ldots, m-1\}$ *and terms* $r_2, \ldots, r_m \in R_E$ *such that* $\sum_{k \in H_j} d_k = j$, $\sum_{k=1}^{m-1} d_k r_k = d r_m$*, all elements of* $\{r_k : k \in H_j\} \cup \{r_m\}$ *are distinct, and all elements of* $\{r_k : k \notin H_j \cup \{1, m\}\}$ *are distinct.*
Then all terms in the sequence S_E *can be expressed in the form*

$$c_E \sum_{i=0}^{\infty} t_i (d+1)^i + r, \tag{2}$$

such that $t_i = 0$ *or 1 for all* i *and* $r \in R_E$.

We make a few notes before presenting the proof. First, in order to simply notation, we will drop the subscripts on c_E and R_E when the choice of E is obvious. Also, we will denote c_{E_m} and R_{E_m} for all integer $m \ge 3$ as simply c_m and R_m.

Next, Theorem 1 is a special case of Theorem 3. As we will show in Section 3.1, if $E = (1,1,1)$, then we can have $c_4 = 12$ and $R_4 = \{0,1,2,3,4\}$. If $N = \sum_{i=0}^{\infty} t_i 4^i$ is a nonnegative integer with 0's and 1's as digits when expressed in base 4, then $c_4 N$ has 0's and 3's as digits and ends in a 0 in base 4. As N ranges over all nonnegative integers with 0's and 1's as digits when expressed in base 4 and r ranges over all elements of R_4, $c_4 N + r$ ranges over exactly the same values as described by Layman in Theorem 1.

Also, given E, the choice of c and R is not unique. Using the example where $E = (1,1,1)$ above, we could also let $c_4 = 48$ and $R_4 = \{0,1,2,3,4,12,13,14,15,16\}$, where Theorem 3 would still predict the same terms for the sequence S_4. Therefore, given E, we will use the minimum value of c that satisfies Theorem 3.

Proof. Let $\mathcal{B}_E$ be the sequence of all integers that can be expressed in the form $c\sum_{i=0}^{\infty} t_i (d+1)^i + r$, *with* $t_i = 0$ *or* 1 *for all* $i \ge 0$ *and* $r \in R$, arranged in increasing order. We prove that $\mathcal{B}_E$ is the same sequence as S_E.

Lemma 3.1. *There are not distinct terms* $x_1, x_2, \ldots, x_m$ *in* $\mathcal{B}_E$ *such that* $d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m$.

Proof. We prove this by contradiction. Assume there are m distinct numbers $x_1, x_2, \ldots, x_m$ in $\mathcal{B}_E$ such that

$$d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m. \tag{3}$$

Because $x_1, x_2, \ldots, x_m$ are in $\mathcal{B}_E$, we can express $x_k = c\sum_{i=0}^{\infty} t_{i,k}(d+1)^i + r_k$, with $t_i = 0$ or 1 for all $i \geq 0$ and $r \in R$, for all $1 \leq k \leq m$. Let $X = dx_m$ and express X as $c\sum_{i=0}^{\infty} T_i(d+1)^i + \mathcal{R}$ such that $T_i = dt_{i,m}$ for all $i \geq 0$ and $\mathcal{R} = dr_m$. Because of equation (3),

$$c\sum_{i=0}^{\infty} T_i(d+1)^i + \mathcal{R} = c\left(\sum_{k=1}^{m-1}\sum_{i=0}^{\infty} d_k t_{i,k}(d+1)^i\right) + \sum_{k=1}^{m-1} d_k r_k. \tag{4}$$

If $\mathcal{R} \neq \sum_{k=1}^{m-1} d_k r_k$, then $\mathcal{R} - \sum_{k=1}^{m-1} d_k r_k$ is a multiple of c or equation (4) cannot be satisfied. Since both $\mathcal{R}$ and $\sum_{k=1}^{m-1} d_k r_k$ are bounded above and below by $d\max(R)$ and 0, the difference between $\mathcal{R}$ and $\sum_{k=1}^{m-1} d_k r_k$ is at most $d\max(R)$.

We show that $d\max(R) < 2c$. By condition (i), $2c > 2d\max(R) - 2\sum_{k=2}^{m-1} d_k(m-k-1)$. Then, since $\max(R) \geq m-2$ and $\left(\sum_{k=2}^{m-1} d_k\right)\left(\frac{0+m-3}{2}\right) \geq \sum_{k=2}^{m-1} d_k(m-k-1)$ by the rearrangement inequality,

$$2c > 2d\max(R_E) - 2\sum_{k=2}^{m-1} d_k(m-k-1)$$
$$2c > d\max(R_E) + d(m-2) - 2\left(\sum_{k=2}^{m-1} d_k\right)\left(\frac{0+m-3}{2}\right)$$
$$2c > d\max(R_E).$$

Since $d\max(R) < 2c$, $\mathcal{R}$ and $\sum_{k=1}^{m-1} d_k r_k$ can differ only by c.

Therefore, we have 3 cases to consider.

Case 1: $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$

If $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$, then we have $\sum_{i=0}^{\infty} T_i(d+1)^i = \sum_{k=1}^{m-1}\sum_{i=0}^{\infty} d_k t_{i,k}(d+1)^i$, which means $T_i = \sum_{k=1}^{m-1} d_k t_{i,k}$ for all i by the same argument we used in Lemma 2.1. If

$T_i = 0$, then $t_{i,k} = 0$ for all $1 \le k \le m-1$. If $T_i = d$, then $t_{i,k} = 1$ for all $1 \le k \le m-1$. Then, for $x_1, x_2, \ldots, x_m$ to be distinct, there must be m distinct values $r_1, r_2, \ldots, r_m$ that satisfy equation $\mathcal{R} = dr_m = \sum_{k=1}^{m-1} d_k r_k$. However, this is impossible because $r_1, r_2, \ldots, r_m$ are terms of S_E.

Case 2: $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k + c$

Let i_0 be the minimum nonnegative integer such that $T_{i_0} = 0$. Subtract c from $\mathcal{R}$, add 1 to T_{i_0} and set $T_i = 0$ for all $i < i_0$ so that $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$ and the value of X is unchanged. This process is similar to the process of carrying digits upon addition. Therefore, $T_{i_0} = 1$ and T_i is 0 or d for all $i \neq i_0$. Since $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$, $T_i = \sum_{k=1}^{m-1} d_k t_{i,k}$ for all i. For all $i \neq i_0$, if T_i is 0, then $t_{i,k} = 0$ for all $1 \le k \le m-1$. If $T_i = d$, then $t_{i,k} = 1$ for all $1 \le k \le m-1$. Finally, $t_{i_0,k} = 0$ for all $1 \le k \le m-1$ except when $k = k_0$ for some k_0, where $d_{k_0} = 1$ and $t_{i_0,k_0} = 1$.

Since $d_{k_0} = 1$, without loss of generality, we can let $k_0 = 1$. Then, $r_2, \ldots, r_{m-1}$ must be distinct for $x_2, \ldots, x_{m-1}$ to be distinct. So by the rearrangement inequality, the minimum value of $\sum_{k=1}^{m-1} d_k r_k$ is $0 \cdot d_1 + \sum_{k=2}^{m-1} d_k(m-k-1)$. Also, since $\mathcal{R} \le d\max(R)$ and we subtracted c from $\mathcal{R}$, $\mathcal{R} \le d\max(R) - c$. Since $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$, that means $d\max(R) - c \ge \sum_{k=2}^{m-1} d_k(m-k-1)$. However, this contradicts condition (i).

Case 3: $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k - c$

Let T_{i_0} be the minimum nonnegative integer such that $T_{i_0} = d$. Add c to $\mathcal{R}$, subtract 1 from T_{i_0} and set $T_i = d$ for all $i < i_0$ so that $R_E = \sum_{k=1}^{m-1} d_k r_k$ and the value of X is unchanged. This process is similar to carrying digits upon subtraction. So $T_{i_0} = d-1$ and T_i is 0 or d for all $i \neq i_0$. Since $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$, $T_i = \sum_{k=1}^{m-1} d_k t_{i,k}$ for all i. For all $i \neq i_0$, if T_i is 0, then $t_{i,k} = 0$ for all k. If $T_i = d$, then $t_{i,k} = 1$ for all k. Also $t_{i_0,k} = 1$ for all $1 \le k \le m-1$ except when $k = k_0$ for some k_0 where $d_{k_0} = 1$ and $t_{i_0,k_0} = 0$.

Since $d_{k_0} = 1$, without loss of generality, we can let $k_0 = 1$. Then, $r_2, \ldots, x_{m-1}$ must be distinct so that $x_2, \ldots, r_{m-1}$ are distinct. So by the rearrangement inequality, the value of $\sum_{k=1}^{m-1} d_k r_k$ is less than or equal to $d_1 \max(R) + \sum_{k=2}^{m-1} d_k(\max(R) - m + 1 + k)$. Also, since $\mathcal{R} \geq 0$ and we added c to $\mathcal{R}$, $\mathcal{R} \geq c$. Since $\mathcal{R} = \sum_{k=1}^{m-1} d_k r_k$, that means

$$c \leq d_1 \max(R) + \sum_{k=2}^{m-1} d_k(\max(R) - m + 1 + k)$$
$$c \leq d \max(R) - \sum_{k=2}^{m-1} d_k(m - k - 1),$$

which contradicts condition (i). □

To finish the proof of Theorem 3, we need to show that no additional elements can be inserted into $\mathcal{B}_E$.

Lemma 3.2. *Given any value y_1 not a term of $\mathcal{B}_E$, there are distinct terms $y_2, y_3, \ldots, y_m$ of $\mathcal{B}_E$, each less than y_1, such that there is a permutation $x_1, \ldots, x_m$ of $y_1, \ldots, y_m$ such that $d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m$.*

Proof. By Proposition 3, we can express y_1 in the form $y_1 = c \sum_{i=0}^{\infty} t_{i,1}(d+1)^i + r_1$, where $t_{i,1}$ is an integer between 0 and d inclusive for all $i \geq 0$ and r_1 is an integer between 0 and $c - 1$ inclusive.

Express y_k, for all $2 \leq k \leq m$, as $c \sum_{i=0}^{\infty} t_{i,k}(d+1)^i + r_k$, where $t_{i,k}$ is 0 or 1 and $r_k \in R$ for all i.

If $t_{i,1}$ is 0 or 1 for all $i \geq 0$, let $t_{i,1} = \cdots = t_{i,m}$ for all $i \geq 0$. Then, $r_1 \notin R$ or else y_1 is a term of $\mathcal{B}_E$. Therefore, we can find distinct $r_2, \ldots, r_m$, all less than r_1, such that there exists a permutation $s_1, \ldots, s_m$ of $r_1, \ldots, r_m$ that satisfies $d_1 s_1 + \cdots + d_{m-1} s_{m-1} = d s_m$. Finally, we can let $y_k = r_k + c \sum_{i=0}^{\infty} t_{i,1}(d+1)^i$ and $x_k = s_k + c \sum_{i=0}^{\infty} t_{i,1}(d+1)^i$ for all $1 \leq k \leq m$.

Now we consider the case when $t_{i,1} > 1$ for some $i \geq 0$. Let $x_k = y_k$ for all $1 \leq k \leq m$. For all $i \geq 0$, let $t_{i,k} = 0$ for all $2 \leq k \leq m$ if $t_{i,1} = 0$. If $t_{i,1} \geq 1$, then let $t_{i,k} = 1$ for all $k \in H_{d-t_{i,1}} \cup \{m\}$ and $t_{i,k} = 0$ otherwise, where $H_{d-t_{i,1}}$ is a subset of $\{2, \ldots, m-1\}$ such that $\sum_{k \in H_{d-t_{i,1}}} d_k = d - t_{i,1}$. Pick any i_0 for which $t_{i_0,1} > 1$.

Let $j = d - t_{i_0,1}$. By condition (ii), we can find a set H_j and terms $r_2, \ldots, r_m \in R$ such that $\sum_{k \in H_j} d_k = j$, $\sum_{k=1}^{m-1} d_k r_k = d r_m$, all elements of $\{r_k : k \in H_j\} \cup \{r_m\}$ are distinct, and all elements of $\{r_k : k \notin H_j \cup \{1, m\}\}$ are distinct.

Then, $x_p \neq x_q$ if $p \in H_j \cup \{m\}$ and $q \notin H_j \cup \{1, m\}$ because $t_{i_0,j} \neq t_{i_0,k}$. Also, all elements of $\{x_k : k \in H_j\} \cup \{x_m\}$ are distinct, and all elements of $\{x_k : k \notin H_j \cup \{1, m\}\}$ are distinct. Therefore, $x_2, \ldots, x_m$ are distinct.

Finally, since for all i, $t_{i,1} \geq t_{i,k}$ and $t_{i_0,1} > t_{i_0,k}$ for all $2 \leq k \leq m$, $x_1 > x_k$ for all $2 \leq k \leq m$.

Then, since $\sum_{k=1}^{m-1} d_k t_{i,k} = d t_{i,m}$ for all i and $\sum_{k=1}^{m-1} d_k r_k = d r_m$, $d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m$. □

Since no m terms in $\mathcal{B}_E$ satisfy the equation $d_1 x_1 + \cdots + d_{m-1} x_{m-1} = d x_m$ and no additional terms can be inserted without creating a solution to the equation, $\mathcal{B}_E$ is the same as S_E and our proof of Theorem 3 is complete. □

This suggests a connection between the sequences A_E and S_E.

Corollary 2. *Given a valid ordered tuple E, let $\mathcal{A}_E$ be the set of the integers in the sequence A_E. Then, if the sequence S_E of terms $a_0, a_1, \ldots$ satisfies conditions (i) and (ii) of Theorem 3 for some z, the set $\{ca + r : a \in \mathcal{A}_E, r \in R\}$ contains the integers in S_E, where $c = a_{z+1}$ and $R = \{a_k : 0 \leq k \leq z\}$.*

3.1. Closed form for S_m

Definition 5. Let $N = \{0, 1, \ldots, 2n-1\} \cup \{2n+1\}$. For every integer $m \geq 3$, Table 1 gives the set of integers R_m and the integer c_m.

R_m	c_m	m
$\{0\}$	1	3
$\{0,1,2,3,5,7,13,26,27,28,29,31\}$	122	5
$\{0,1,2,3,4,5,7,10,33,34,35,36,37,38\}$	219	7
$\{0,1,\ldots,2n\}$	$2n^2+3n-2$	$2n$, $n>1$
$N \cup \{3n+1\} \cup \{c+2n^2+5n : c \in N\}$	$4n^3+12n^2+5n$	$2n+1$, $n>3$

Table 1: Definition of R_m and c_m

Theorem 4. *An integer is in the sequence S_m if and only if it can be expressed in the form*

$$c_m \sum_{i=0}^{\infty} t_i m^i + r, \tag{5}$$

where t_i can be either 0 *or* 1 *for all* $i \geq 0$ *and* $r \in R_m$.

Proof. We need to show that conditions (i) and (ii) of Theorem 3 are satisfied.

Lemma 4.1. *The set* $S_m \cap [0, c_m - 1]$ *is the same as* R_m.

Proof. In the appendix, we prove the case $m = 2n$ in Lemma 4.3 and the case $m = 2n + 1$, with integer $n > 3$, in Lemma 4.5. The cases for when $m = 3, 5, 7$ are brute forced with a computer. □

Since $S_m \cap [0, c_m - 1] = R_m$, we can easily check that condition (i) is satisfied.

Now, we show that condition (ii) is satisfied. We want to show that for every integer $0 \leq r_1 \leq c_m - 1$ and every integer $0 \leq j \leq m - 3$, there exists a subset H_j of $\{2, \ldots, m-1\}$ and terms $r_2, \ldots, r_m \in R_m$ such that $|H_j| = j$, $\sum_{k=1}^{m-1} r_k = (m-1)r_m$, all elements of $\{r_k : k \in H_j \cup \{m\}\}$ are distinct, and all elements of $\{r_k : k \notin H_j \cup \{1, m\}\}$ are distinct.

Let j be any integer between 0 and $m - 3$ inclusive. First, we consider the case when $r_1 \notin R_M$. Since E_m is a valid ordered tuple, we can find a subset H_j of $\{2, \ldots, m-1\}$ such that $|H_j| = j$. Also, by the definition of the sequence S_m, for every $r_1 \notin R_m$, we can find distinct $r_2, \ldots, r_m < r_1$ such that $\sum_{k=1}^{m-1} r_k = (m-1)r_m$, so that condition (ii) is satisfied. Now we consider the case for when $r_1 \in R_m$.

Lemma 4.2. *Given any* $r_1 \in R_m$, *we can find* $r_2, r_3, \ldots, r_m \in R_m$ *such that* $r_2, \ldots, r_{m-1}$ *are distinct and* $\sum_{k=1}^{m-1} r_k = (m-1)r_m$.

Proof. The result follows immediately from Lemmas 4.6 and 4.8 in the Appendix, where we prove the cases when m is even and m is odd separately. □

Let r_1 be an element of R_m. By Lemma 4.2, let $r_2, \ldots, r_m \in R_m$ be chosen such that $\sum_{k=1}^{m-1} r_k = (m-1)r_m$ and $r_2, \ldots, r_{m-1}$ are distinct. If there is some value $2 \leq k_0 \leq m - 1$ for which $r_{k_0} = r_m$, then let $k_0 \notin H_j$. Otherwise, we can let any j integers between 2 and $m - 1$ to be in H_j.

Since both conditions (i) and (ii) are satisfied, the proof is complete. □

3.2. Closed forms for particular S_E

With a computer program, we tested the valid ordered tuples $E = (d_1, \ldots, d_{m-1})$ for when $4 \leq m \leq 7$ until the terms exceeded 80,000 to identify closed forms for S_E

E	Closed Form	R_E
(1, 1, 1)	$12\sum_{i=0}^{\infty} t_i 4^i + r$	$r \in \{0, 1, 2, 3, 4\}$
(1, 1, 2)	$16\sum_{i=0}^{\infty} t_i 5^i + r$	$r \in \{0, 1, 2, 3, 4\}$
(1, 1, 1, 1)	$122\sum_{i=0}^{\infty} t_i 5^i + r$	$r \in \{0, 1, 2, 3, 5, 7, 13, 26, 27, 28, 29, 31\}$
(1, 1, 1, 2)	$103\sum_{i=0}^{\infty} t_i 6^i + r$	$r \in \{0, 1, 2, 3, 4, 14, 18, 19, 20, 21\}$
(1, 1, 2, 3)	$81\sum_{i=0}^{\infty} t_i 8^i + r$	$r \in \{0, 1, 2, 3, 4, 14, 17, 31, 130, 131, 132, 133, 134, 144, 147\}$
(1, 1, 2, 4)	$29\sum_{i=0}^{\infty} t_i 9^i + r$	$r \in \{0, 1, 2, 3, 4\}$
(1, 1, 1, 1, 1)	$25\sum_{i=0}^{\infty} t_i 6^i + r$	$r \in \{0, 1, 2, 3, 4, 5, 6\}$
(1, 1, 1, 1, 2)	$31\sum_{i=0}^{\infty} t_i 7^i + r$	$r \in \{0, 1, 2, 3, 4, 5, 6\}$
(1, 1, 1, 1, 3)	$30\sum_{i=0}^{\infty} t_i 8^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 1, 1, 4)	$51\sum_{i=0}^{\infty} t_i 9^i + r$	$r \in \{0, 1, 2, 3, 4, 6, 7\}$
(1, 1, 1, 2, 2)	$106\sum_{i=0}^{\infty} t_i 8^i + r$	$r \in \{0, 1, 2, 3, 4, 14, 15, 16\}$
(1, 1, 1, 2, 3)	$1170\sum_{i=0}^{\infty} t_i 9^i + r$	$r \in \{0, 1, 2, 3, 4, 14, 17, 31, 130, 131, 132, 133, 134, 144, 147\}$
(1, 1, 1, 3, 3)	$38\sum_{i=0}^{\infty} t_i 10^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 1, 3, 4)	$43\sum_{i=0}^{\infty} t_i 11^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 1, 3, 5)	$48\sum_{i=0}^{\infty} t_i 12^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 1, 3, 6)	$653\sum_{i=0}^{\infty} t_i 13^i + r$	$r \in \{0, 1, 2, 3, 4, 12, 34, 42, 48, 55\}$
(1, 1, 2, 2, 2)	$32\sum_{i=0}^{\infty} t_i 9^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 2, 2, 3)	$208\sum_{i=0}^{\infty} t_i 10^i + r$	$r \in \{0, 1, 2, 3, 4, 18, 19, 20, 24\}$
(1, 1, 2, 2, 5)	$3622\sum_{i=0}^{\infty} t_i 12^i + r$	$r \in \{0, 1, 2, 3, 4, 19, 22, 28, 50, 300, 301, 302, 303, 304, 319, 322, 330\}$
(1, 1, 2, 2, 6)	$52\sum_{i=0}^{\infty} t_i 13^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 2, 3, 3)	$401\sum_{i=0}^{\infty} t_i 11^i + r$	$r \in \{0, 1, 2, 3, 4, 8, 37, 38, 39, 40, 41\}$
(1, 1, 2, 3, 4)	$420\sum_{i=0}^{\infty} t_i 12^i + r$	$r \in \{0, 1, 2, 3, 4, 23, 35, 37, 39\}$
(1, 1, 2, 3, 7)	$61\sum_{i=0}^{\infty} t_i 15^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 2, 4, 4)	$50\sum_{i=0}^{\infty} t_i 13^i + r$	$r \in \{0, 1, 2, 3, 4, 5\}$
(1, 1, 2, 4, 7)	$80\sum_{i=0}^{\infty} t_i 16^i + r$	$r \in \{0, 1, 2, 3, 4, 5, 6\}$

Table 2: Closed Forms for S_E

for 129 choices of E. The 25 tuples the computer found when $4 \leq m \leq 6$ are given in Table 2, where $t_i = 0$ or 1 for $i \geq 0$ for each of the closed forms.

4. Asymptotics

Let $g(n)$ be the number of terms of A_E that are less than n, for some positive real n and valid ordered tuple E. Similarly, let $h(n)$ be the number of terms of S_E that are less than n. We will derive bounds for $g(n)$ and $h(n)$ and growth rates of A_E and S_E.

For any valid ordered tuple E and nonnegative integer i_0, $g((d+1)^{i_0}) = 2^{i_0}$ because there are 2^{i_0} numbers that, when expressed in base $d+1$, have at most i_0 digits and only 0's and 1's as digits. Therefore for a nonnegative integer n, we have

$$2^{\lfloor \log_{d+1}(n) \rfloor} \leq g(n) \leq 2^{\lceil \log_{d+1}(n) \rceil},$$
$$\frac{1}{2} \cdot 2^{\log_{d+1}(n)} \leq g(n) \leq 2 \cdot 2^{\log_{d+1}(n)},$$
$$\frac{1}{2} n^{\log_{d+1}(2)} \leq g(n) \leq 2n^{\log_{d+1}(2)}.$$

From these bounds, $g(n) = \Theta(n^{\log_{d+1}(2)})$. Also, from these bounds, we can derive bounds for the growth rate of A_E. Let the terms of A_E be $a_0, a_1, \ldots$. Since $g(a_n) = n$,

$$\frac{1}{2} a_n^{\log_{d+1}(2)} \leq n \leq 2a_n^{\log_{d+1}(2)},$$
$$2^{\log_2(d+1)} n^{\log_2(d+1)} \geq a_n \geq 2^{-\log_2(d+1)} n^{\log_2(d+1)}.$$

Therefore, $a_n = \Theta(n^{\log_2(d+1)})$.

Now, we bound $h(n)$. Suppose that all terms of S_E can be expressed in the form

$$r + c \sum_{i=0}^{\infty} t_i (d+1)^i, \tag{6}$$

where t_i is 0 or 1 for all $i \geq 0$, c is a constant, and $r \in R$ for a set R that contains nonnegative integers that are all less than c. Then, for any positive integer multiple $k_0 c$ of c, we have $h(k_0 c) = |R|\, g(k_0)$ because there are $g(k_0)$ ways to choose the sequence $t_0, t_1, \ldots$ and $|R|$ ways to choose r. Therefore, for a nonnegative integer

n, we have

$$|R|\,g(\left\lfloor \frac{n}{c} \right\rfloor) \leq h(n) \leq |R|\,g(\left\lceil \frac{n}{c} \right\rceil),$$
$$|R|\,g(\frac{n}{c}-1) \leq h(n) \leq |R|\,g(\frac{n}{c}+1),$$
$$|R|\,\frac{1}{2}(\frac{n}{c}-1)^{\log_{d+1}(2)} \leq h(n) \leq |R|\,2(\frac{n}{c}+1)^{\log_{(d+1)}(2)},$$
$$\frac{1}{2}\,|R|\,c^{-\log_{d+1}(2)}(n-c)^{\log_{d+1}(2)} \leq h(n) \leq 2\,|R|\,c^{-\log_{d+1}(2)}(n+c)^{\log_{d+1}(2)}.$$

From these bounds, we get $h(n) = \Theta(n^{\log_{d+1}(2)})$. Also, from these bounds, we can derive a bound for the growth rate of S_E. Let the terms of S_E be $a_0, a_1, \ldots$. Since $h(a_n) = n$,

$$|R|\,\frac{1}{2}(\frac{a_n}{c}-1)^{\log_{d+1}(2)} \leq n \leq |R|\,2(\frac{a_n}{c}+1)^{\log_{d+1}(2)},$$
$$c(\frac{2}{|R|})^{\log_2(d+1)} n^{\log_2(d+1)} + c \geq a_n \geq c(\frac{1}{2\,|R|})^{\log_2(d+1)} n^{\log_2(d+1)} - c.$$

Therefore, $a_n = \Theta(n^{\log_2(d+1)})$.

Given a valid ordered tuple $E = (d_1, \ldots, d_{m-1})$, let $f(n)$ be the maximum cardinality over all subsets of $\{0, \ldots, n-1\}$ that do not contain a solution to $d_1x_1 + \cdots + d_{m-1}x_{m-1} = dx_m$ in elements not all the same. Milenkovic, Kashyap, and Leyba [6] showed that Behrend's construction [2] can be modified to show that $f(n) \geq \gamma_1 n e^{-\gamma_2\sqrt{\ln(n)} - \frac{1}{2}\ln(\ln(n))}(1+o(1))$ for $n > d^2$, where $\gamma_1 = d^2\sqrt{\frac{1}{2}\ln(d)}$, $\gamma_2 = 2\sqrt{2\ln(d)}$, and $o(1)$ vanishes as $n \to \infty$. Since $f(n)$ is asymptotically greater than $g(n)$, for all valid ordered tuples E, and $h(n)$, for all tuples E for which we have a closed form of S_E, we have shown that the greedy algorithm is not optimal in these cases.

However, it should be noted that Behrend's construction, while much stronger asymptotically, is less efficient for small values of n. For example, if we let $E = E_4$ and $n = 10^{10}$, the bound obtained by Milenkovic, Kashyap, and Leyba shows that $f(10^{10}) \geq 3187$. The bounds obtained by the greedy algorithm show $h(10^{10}) \geq \left\lceil |R|\,\frac{1}{2}(\frac{10^{10}}{c}-1)^{\log_{d+1}(2)} \right\rceil = 15360$ and $f(10^{10}) \geq g(10^{10}) \geq \left\lceil \frac{1}{2}(10^{10})^{\log_{d+1}(2)} \right\rceil = 10133$.

5. Conclusion

We have found the closed forms of all sequences A_E, given any valid ordered tuple E. Also, we have found the closed forms of S_E for specific choices of E, including E_m for all $m \geq 3$. Possible future work include simplifying the condition needed to be satisfied in Theorem 3 or extending Theorem 3 to cover more tuples E for when S_E has a closed form. Also, generating the sequences and plotting them suggests that, in general, there are sequences that cannot be described in a similar way to our closed forms. Further research can also be done include in bounding the rates of growth of these sequences. For example, given an ordered tuple of positive integers $E = (d_1, \ldots, d_{m-1})$, it appears that S_E grows at least as fast asymptotically as S_m.

6. Acknowledgments

I would like to thank the Center for Excellence in Education, the Research Science Institute, and Akamai for funding me in the summer. I would also like to thank Nan Li for mentoring me, Professor Richard Stanley for the project idea, and my tutor Dr. John Rickert. I am also grateful for the guidance of Professor Jake Wildstrom, Professor John Layman, and Dr. Tanya Khovanova during the research process, for the help of an anonymous referee, Scott Kominers, Wei Lue and Dr. Johnothon Sauer during the publication process, and for the advice of Travis Hance, in improving the algorithm of my computer program.

A. Appendix

We present proofs of the Lemmas that were omitted in the main paper.

Definition 6. Let the set $S_m(k)$ contain the terms of S_m that are less than or equal to k.

A.1. Method

We present a method that will be used repeatedly in the proofs following Lemmas.

Given integers α and z, we need to determine whether there exist $x_2, x_3, \ldots, x_m \in S_m(z)$ such that $\alpha + \sum_{k=2}^{m-1} x_k = (m-1)x_m$.

Let the set $W = \{w_1, w_2 \ldots, w_s\}$ be the set $S_m(z) \backslash \{x_2, x_3, \ldots, x_{m-1}\}$. Then, $\alpha + \sum_{k=2}^{m-1} x_k = (m-1)x_m$ is equivalent to

$$\alpha + \sum_{k=0}^{|S_m(z)|-1} a_k - \sum_{k=1}^{s} w_k = (m-1)x_m \tag{7}$$

A.2. Proofs

Lemma 4.3. *If $m = 2n$ for any integer $n > 1$, the the only terms of S_m less than $2n^2 + 3n - 2$ is in $\{0, 1, \ldots, 2n\}$.*

Proof. To prove Lemma 4.3, we prove two claims.

Claim 1: The first $2n+1$ terms of S_{2n} are the integers from 0 to $2n$ inclusive.

The first $2n-1$ terms of S_{2n} are the integers from 0 to $2n-2$ inclusive because there are not enough distinct terms less than $2n-1$ to satisfy the equation $\sum_{k=1}^{2n-1} x_k = (2n-1)x_{2n}$.

If we substitute $\alpha = 2n-1$, $m = 2n$, $z = 2n-2$, and $W = \{x_{2n}\}$ into equation (7), we obtain $x_{2n} = \frac{2n-1}{2}$, which is not an integer.

If we substitute $\alpha = 2n$, $m = 2n$, $z = 2n-1$, and $W = \{x_{2n}, x\}$ into equation (7), we obtain $n(2n+1-2x_{2n}) = x$. The value of x is between 0 and $2n$ only if $0 \leq 2n+1-2x_{2n} \leq 2$. But since $2n+1-2x_{2n}$ is odd, $x = n = x_{2n}$ which is a contradiction.

Claim 2: For every value of $2n+1 \leq \alpha \leq 2n^2+3n-3$, there are distinct $x_2, x_3, \ldots, x_{2n} \in S_{2n}(2n)$ such that $\alpha + \sum_{k=2}^{2n-1} x_k = (2n-1)x_{2n}$.

We find an explicit construction for all $2n+1 \leq \alpha \leq 2n^2+3n-3$. If $2n+1 \leq \alpha \leq 2n^2+n-1$, we let $\alpha = pn+q$, where $0 \leq q \leq n-1$. Plug in $m = 2n$, $z = \alpha - 1$, $W = \{a, b, x_{2n}\}$, and $x_{2n} = n+c$ in equation (7), we obtain $(p+1)n+q = 2nc+a+b$. Let $a = 0$ if p is odd and $a = n$ if p is even. Let $b = q$ and $c = \lfloor (p+1)/2 \rfloor$ so that $(p+1)n = 2nc + a$ and $q = b$, which satisfies $(p+1)n + q = 2nc + a + b$.

Now we make sure that x_{2n}, a and b are distinct.

If p is even, then $a = n > q = b$ and $x_{2n} = n = a$ only if $c = 0$. But $p \geq 2$ so $c = \lfloor (p+1)/2 \rfloor > 0$.

If p is odd, then $x_{2n} = n + c > q = b$. Also $a = b = 0$ only if $q = 0$, in which case we need to redefine our values of x_{2n}, a and b to ensure their distinctness. If p is odd and $q = 0$, let $a = 2n$, $b = 0$ and $c = \lfloor (p+1)/2 \rfloor - 1$. Then, $a > x_{2n} > b$.

If $2n^2+n \leq \alpha \leq 2n^2+3n-3$, let $x_{2n} = 2n$, $b = 2n-1$ and $a = \alpha - (2n^2+n-1)$. Since $a < b < x_{2n}$, a, b and x_{2n} are distinct.

From Claim 1 and Claim 2, we have proven that the integers from 0 to $2n$ inclusive are in S_{2n} and that the integers between $2n+1$ and $2n^2+3n-3$ inclusive are not, finishing the proof for Lemma 4.3. □

To help prove Lemma 4.5, we prove Lemma 4.4.

Lemma 4.4. *Given the $2 \leq k \leq 2n-2$ consecutive integers $y_1 < y_2 < \cdots < y_k$ between $2n-k-1$ and $2n-2$ and an integer p, we can find a set of k integers that does not contain p and is a subset of $S_{2n+1}(2n+1)$ such that the sum of its elements equal to the sum of the original k consecutive integers.*

Proof. If p is not one of the integers $y_1, \ldots, y_k$, we are done. If not, let y_{i_0} be the median of $\{y_1, \ldots, y_k\}$. If $p < y_{i_0}$, decrement the $p - y_1 + 1$ smallest integers and increment the $p - y_1 + 1$ largest integers in $\{y_1, \ldots, y_k\}$.

If $p > y_{i_0}$, increment the $y_k - p + 1$ largest integers and decrement the $y_k - p + 1$ smallest integers in $\{y_1, \ldots, y_k\}$.

If $p = y_{i_0}$, then k must be odd, which means $k < 2n-2$ and $y_1 \geq 2$. Then, decrement the $i_0 - 1$ smallest integers and increment the i_0 largest integers in $\{y_1, \ldots, y_k\}$. Then decrement the smallest integer y_1 again so $\{y_1, \ldots, y_k\} \subset S_{2n+1}(2n+1)$. □

Lemma 4.5. *If $n > 3$, then $S_{2n+1}(4n^3+12n^2+5n-1) = N \cup \{3n+1\} \cup \{c+2n^2+5n : c \in N\}$, where $N = \{0, 1, \ldots, 2n-1\} \cup \{2n+1\}$.*

Proof. We start with $a_0 = 0$ and generate the terms to show they are the terms listed in Lemma 4.5.

The integers $0, 1, \ldots, 2n-1$ must be in the S_{2n+1} because there are not $2n+1$ distinct terms in the sequence, which means there cannot be distinct terms $x_1, x_2, \ldots, x_{2n+1}$ that satisfy

$$\sum_{k=1}^{2n} x_k = 2nx_{2n+1}. \tag{8}$$

Now we show $2n$ cannot be a term of S_{2n+1}. If $2n$ were a term of S_{2n+1}, we can find a solution for equation (8) by letting $x_{2n+1} = n$ and $x_k = k-1$ if $k \le n$ and $x_k = k$ if $k \ge n+1$.

We use contradiction to prove that $2n+1$ is the next term. If we let $\alpha = 2n+1$, $m = 2n+1$, $z = 2n$ and $W = \{x_{2n+1}\}$ in equation (7), we obtain $x_{2n+1} = n + 1/(2n+1)$, which is not an integer.

We show $3n+1$ is the next term in S_{2n+1}. In (7), let $\alpha = 2n+x$ where $2 \le x \le n$, $m = 2n+1$, $z = \alpha - 1$ and $W = \{x, x_{2n+1}\}$. Then, we obtain $x_{2n+1} = n+1$, which is in $S_{2n+1}(\alpha - 1)$.

To prove that $3n+1$ is the next term, we again use contradiction. In equation (7), let $\alpha = 3n+1$, $m = 2n+1$, $z = 3n$ and $W\{x, x_{2n+1}\}$. Then, we obtain $n+1+(n+1-x)/(2n+1) = x_{2n+1}$.

Since $0 \le x < 2n+1$, the only way $n+1-x$ can be a multiple of $2n+1$ is if $x = n+1$. But then $x_{2n+1} = n+1 = x$, which is a contradiction.

Now, we show that given any $3n+2 \le \alpha < 2n^2+5n-1$, we can find distinct $x_2, x_3, \ldots, x_{2n+1} \in S_{2n+1}(3n+1)$ such that $\alpha + \sum_{k=2}^{2n} x_k = (2n)x_{2n+1}$.

In equation (7), let $m = 2n+1$, $z = \alpha - 1$, and $W = \{x, y, x_{2n+1}\}$. Then we obtain $n+1+\frac{n+1+\alpha-x-y}{2n+1} = x_{2n+1}$. For every $3n+2 \le \alpha \le 2n^2-2n-2$, let $\alpha = (2n+1)A+B$, where $2 \le A \le n-2$ and $-n \le B \le n$. We present the solutions for x_{2n+1}, x and y given α in Table 3. For every $2n^2-2n+1 \le \alpha \le 2n^2+5n-1$,

α	x_{2n+1}	x	y	
$A(2n+1)+B$	$n+1+A$	0	$n+1+B$	$B \le n-2,\ B \ne A$
$A(2n+1)+B$	$n+1+A$	1	$n+B$	$B = A,\ B \le n-2$
$A(2n+1)+B$	$n+1+A$	$B-n+2$	$2n-1$	$n-1 \le B \le n,\ A < n-2$
$A(2n+1)+B$	$n+1+A$	$B-n+3$	$2n-2$	$n-1 \le B \le n,\ A = n-2$

Table 3: If $3n+2 \le \alpha \le 2n^2-2n-2$

let $\alpha = 2n^2 + C$, where $-2n+1 \le C \le 5n-1$. We present the solutions in Table 4.

We show that $2n^2+5n, 2n^2+5n+1, \ldots, 2n^2+7n-1, 2n^2+7n+1$ are the next terms in S_{2n+1} by contradiction. Let $\alpha \in \{2n^2+5n+c : 0 \le c \le 2n-1\} \cup \{2n^2+7n+1\}$. Assume that there are terms $x_2, x_3, \ldots, x_{2n+1}$ in the sequence, each less than α such that

$$\alpha + \sum_{k=2}^{2n} x_k = (2n)x_{2n+1}. \tag{9}$$

We prove that $x_{2n+1} \ge 2n^2+5n$, also by contradiction. Assume that $x_{2n+1} < 2n^2+5n$. If α is the only integer among α, $x_2, x_3, \ldots, x_{2n}$ that is greater than or

α	x_{2n+1}	x	y	
$2n^2+C$	$2n-1$	$C+2n+5$	$2n-2$	$-2n-1 \leq C \leq -8$
$2n^2+C$	$2n-1$	$2n-5$	$2n+1$	$C=-7$
$2n^2+C$	$2n-1$	$C+n+2$	$3n+1$	$-6 \leq C \leq n-4$
$2n^2+C$	$2n+1$	0	$C+1$	$n-3 \leq C \leq 2n-2$
$2n^2+C$	$2n+1$	$C-2n+2$	$2n-1$	$2n-1 \leq C \leq 4n-4$
$2n^2+C$	$2n+1$	$C-3n$	$3n+1$	$4n-3 \leq C \leq 5n-1$

Table 4: If $2n^2-2n-1 \leq \alpha \leq 2n^2+5n-1$

equal to $2n^2+5n$, then the minimum value for x_{2n+1} is $x_{2n+1} \geq \frac{2n^2+5n+\sum_{k=0}^{2n-2} k}{2n} = 2n+1+\frac{1}{2n}$, which is greater than $2n+1$. So x_{2n+1} can only be $3n+1$. But since $x_2, x_3, \ldots, x_{2n}$ cannot be $3n+1$, by equation (9), $3n+1 \leq \frac{(2n^2+7n+1)+(2n+1)+\sum_{k=2}^{2n-1} k}{2n} = 2n+4+\frac{1}{2n}$, which is a contradiction because $n>3$. If at least one of the integers $x_2, x_3, \ldots, x_{2n}$ are greater than or equal to $2n^2+5n$, then by equation (9) $x_{2n+1} \geq \frac{(2n^2+5n)+(2n^2+5n+1)+\sum_{k=0}^{2n-3} k}{2n} = 3n+2+\frac{n+4}{2n}$, which cannot occur because there are no terms between $3n+2$ and $2n^2+5n-1$ inclusive. Therefore $x_{2n+1} \geq 2n^2+5n$.

Let $\alpha = M+r$ such M is $2n^2+5n$ and $r \in S_{2n+1}(2n+1)$ and $x_i = M_i + r_i$, where M_i is 0 or $2n^2+5n$ and $r_i \in S_{2n+1}(3n+1)$. Also, r_i can be $3n+1$ only if $M_i = 0$. Then,

$$M + r + \sum_{k=2}^{2n} M_k + \sum_{k=2}^{2n} r_k = (2n)M_{2n+1} + (2n)r_{2n+1}.$$

Since $x_{2n+1} \geq 2n^2+5n$, $M_{2n+1} = 2n^2+5n$. The maximum value of $r + \sum_{k=2}^{2n} r_k - 2nr_{2n+1}$ is less or equal to than twice the sum of the n largest elements of $S_{2n+1}(3n+1)$, since the minimum value of $2nr_{2n+1}$ is 0 and no three elements of $\{r, r_2, \ldots, r_{2n}\}$ can be pairwise equal. Otherwise, two elements of $\{\alpha, x_2, \ldots, x_{2n}\}$ must be equal.

So the maximum value of the difference is $2\left((3n+1)+(2n+1)+\sum_{k=n+2}^{2n-1} k\right) - 2n \cdot 0 = 3n^2+5n+2$.

Since $3n^2+5n+2 < 2(2n^2+5n)$, at most one of elements of $\{M_k : 2 \leq k \leq 2n\}$ can be 0, or else the difference $r + \sum_{k=2}^{2n} r_k - 2nr_{2n+1}$ is less than $2nM_{2n+1} - M - \sum_{k=2}^{2n} M_k$. If $\alpha < 2n^2+7n-1$, there are not $2n-1$ distinct integers between $2n^2+5n$ and $\alpha - 1$ inclusive, which means α is in S_{2n+1}. If $\alpha = 2n^2+7n-1$, and not all $\{M_k : 2 \leq k \leq 2n\}$ are equal to $2n^2+5n$, then the maximum value for x_{2n+1} would be $x_{2n+1} \leq \frac{\sum_{k=1}^{2n-1} 2n^2+5n+k+(3n+1)}{2n} = 2n^2+5n-\frac{3n-1}{2n}$, which is less than $2n^2+5n$, contradicting the assumption that $x_{2n+1} \geq 2n^2+5n$.

The integer $2n^2+7n$ is not in S_{2n+1} because equation (8) is satisfied if we let $x_{2n+1} = 2n^2+6n$ and $x_k = 2n^2+5n+k-1$ if $k \leq n$ and $x_k = 2n^2+5n+k$ if $k \geq n+1$.

If $\alpha = 2n^2+7n+1$ and not all elements of $\{M_k : 2 \leq k \leq 2n\}$ are $2n^2+5n$, then the maximum value for x_{2n+1} is $x_{2n+1} \leq \frac{(2n^2+7n+1)+\left(\sum_{k=2}^{2n-1} 2n^2+5n+k\right)+(3n+1)}{2n} = 2n^2+5n-\frac{n-1}{2n}$, which is less than $2n^2+5n$, contradicting $x_{2n+1} \geq 2n^2+5n$. If all $M_2, M_3, \ldots, M_{2n} \in \{2n^2+5n\}$, then by assumption (9),

$$M + r + \sum_{k=2}^{2n} M_k + \sum_{k=2}^{2n} r_k = (2n)M_{2n+1} + (2n)r_{2n+1}$$

$$r + \sum_{k=2}^{2n} r_k = (2n)r_{2n+1}. \tag{10}$$

But r, $r_2, r_3, \ldots, r_{2n+1}$ are distinct elements of $S_{2n+1}(2n+1)$, so equation (10) has no solutions and $2n^2+7n+1$ is in the sequence.

We now show that $\{c : 2n^2+7n+2 \leq c \leq 4n^3+12n^2+5n-1\} \cap S_{2n+1}(4n^3+12n^2+5n-1) = \emptyset$. So given any $2n^2+7n+2 \leq \alpha \leq 4n^3+12n^2+5n-1$, we show that there are distinct $x_2, x_3, \ldots, x_{2n+1}$ in the sequence such that $\alpha + \sum_{k=2}^{2n} x_k = (2n)x_{2n+1}$.

For ease of notation, we represent the integers $x_2, x_3, \ldots, x_{2n}$ with the two sets $U = \{u_1, u_2, \ldots, u_p\}$ and $V = \{v_1, v_2, \ldots, v_q\}$. The set U contains the elements of $\{x_2, x_3, \ldots, x_{2n}\}$ that are greater than or equal to $2n^2+5n$, with $2n^2+5n$ subtracted from each those integers. The set V contains the elements of $\{x_2, x_3, \ldots, x_{2n}\}$ that are less than $2n^2+5n$. All elements in set U must be in $S_{2n+1}(2n+1)$ and all elements in set V must be in $S_{2n+1}(3n+1)$. We can express

$$\alpha + \sum_{k=2}^{2n} x_k = \alpha + \sum_{k=1}^{p} u_k + \sum_{k=1}^{q} v_k + |U|(2n^2+5n), \tag{11}$$

which implies that

$$x_{2n+1} = \frac{\alpha + \sum_{k=1}^{p} u_k + \sum_{k=1}^{q} v_k + |U|(2n^2+5n)}{2n}.$$

The solutions for $\{x_k : 2 \leq k \leq 2n+1\}$ for all $2n^2+7n+2 \leq \alpha \leq 2n^2+11n-1$ are displayed in Table 5. By Lemma 4.4, we can define $G(k,p)$ as a subset of k elements of $S_{2n+1}(2n+1)$ that has the same sum as the consecutive integers between $2n-k-1$ and $2n-2$ inclusive and does not contain the integer p. Since Lemma 4.4 only applies to when $2 \leq k \leq 2n-2$, we need to define $G(k,p)$ for when $k = 0$ or 1. Let $G(0,p) = \{\}$, $G(1,p) = 2n-2$ for all $p \neq 2n-2$, and $G(1, 2n-2)$ be undefined.

α	U	V	x_{2n+1}	
$2n^2+7n+2$	$S_{2n+1}(2n+1)\backslash\{0,1,3\}$	$\{2n+1\}$	$2n^2+5n$	
$2n^2+7n+3$	$S_{2n+1}(2n+1)\backslash\{0,1,2\}$	$\{2n-1\}$	$2n^2+5n$	
$2n^2+7n+4$	$S_{2n+1}(2n+1)\backslash\{0,1,3\}$	$\{2n-1\}$	$2n^2+5n$	
$2n^2+C$	$S_{2n+1}(2n+1)\backslash$ $\{0,2,C-7n-2\}$	$\{2n-1\}$	$2n^2+5n$	$7n+5 \leq C$ $\leq 9n+1$
$2n^2+C$	$S_{2n+1}(2n+1)\backslash$ $\{0,C-9n-1,2n+1\}$	$\{2n-1\}$	$2n^2+5n$	$9n+2 \leq C$ $\leq 11n-1$

Table 5: If $2n^2+7n+2 \leq \alpha \leq 2n^2+11n-1$

Also, if $T = \{t_i : 0 \leq i \leq j\}$ is a set of distinct nonnegative integers arranged in increasing order, we define $H(T)$ to take the smallest value of $t_i > i$ and decrement it. Let $H^{(k)}(T)$ denote applying the function H to T k times and $[n]$ be the set containing the integers from 1 to n inclusive.

The solutions for $\{x_k : 2 \leq k \leq 2n+1\}$ for all $2n^2+11n \leq \alpha \leq 4n^3+12n^2+n-1$ are displayed in Table 6. There may be multiple ways to express α as $2n^2+11n+(2n^2+5n)A+2nB+C$, in which case there are multiple solutions shown. Notice that we cannot have $A = 2n-3$ and $B = 2n-2$ at the same time, as $G(1, 2n-2)$ is undefined. To correct this, we let $A = 2n-2$, $B = n-4$, and let $V = H^{(C+n)}(\{2, \ldots, 2n-1\} \cup \{2n+1\})$.

α	U	V	x_{2n+1}	
$2n^2+11n$ $+(2n^2+5n)A$ $+2nB+C$	$G(2n-2-A$ $,B)$	$H^{(C)}([A]\cup$ $\{2n-1\})$	$2n^2+5n$ $+B$	$0 \leq B \leq 2n-1,$ $0 \leq C \leq 2n-1,$ $0 \leq A \leq 2n-2$

Table 6: If $2n^2+11n \leq \alpha \leq 4n^3+12n^2+n-1$ and $A = 2n-3$ and $B = 2n-2$ are not true at the same time

We now present Table 7 giving a solution for every $4n^3+12n^2+n \leq \alpha \leq 4n^3+12n^2+5n-1$.

α	U	V	x_{2n+1}	
$4n^3+12n^2+n+C$	$\emptyset$	$S_{2n+1}(2n+1)\backslash\{0,C+1\}$	$2n^2+7n+1$	$0 \leq C \leq 2n-2$
$4n^3+12n^2+n+C$	$\emptyset$	$S_{2n+1}(2n+1)\backslash\{1,2n-1\}$	$2n^2+7n+1$	$C = 2n-1$
$4n^3+12n^2+n+C$	$\emptyset$	$S_{2n+1}(2n+1)\backslash\{C-2n,2n+1\}$	$2n^2+7n+1$	$2n \leq C \leq 4n-1$

Table 7: If $4n^3+12n^2+n \leq \alpha \leq 4n^3+12n^2+5n-1$

Since we have worked from 0 to $4n^3+12n^2+5n-1$ and tested if each integer in that range is in S_{2n+1} and found that the results match the statement in Lemma 4.5, our proof is complete. □

To prove Lemma 4.2, we need to prove Lemma 4.6 and Lemma 4.8.

Lemma 4.6. *If $m = 2n$, given any $\alpha \in \{0, 1, \dots, 2n\}$, we can find distinct $x_2, x_3, \dots, x_{2n} \in S_{2n}(2n)$ such that $\alpha + \sum_{k=2}^{2n-1} x_k = (2n-1)x_{2n}$.*

Proof. In equation (7), let $m = 2n$, $z = 2n$, and $W = \{a, b\}$. Then, we obtain $\alpha = n(2x_{2n} - 2n - 1) + a + b$.

We display the solutions for $0 \leq \alpha \leq 2n$ in Table 8.

a	b	x_{2n}	
n	$2n$	$n-1$	$\alpha = 0$
0	$n+\alpha$	n	$1 \leq \alpha \leq n$
0	$\alpha - n$	$n+1$	$n+1 \leq \alpha \leq 2n$

Table 8: If $0 \leq \alpha \leq 2n$

Since we have covered all the values for α from 0 to $2n$ inclusive, we are done with the proof of Lemma 4.6. □

To prove Lemma 4.8, we use of the following result.

Lemma 4.7. *Given any $\alpha \in S_{2n+1}(2n+1)$, we can find $x_2, x_3, \dots, x_{2n+1} \in S_{2n+1}(2n+1)$ such that $x_2, x_3, \dots, x_{2n}$ are distinct and $\alpha + \sum_{k=2}^{2n} x_k = 2nx_{2n+1}$.*

Proof. In equation (7), let $m = 2n+1$, $z = 2n+1$ and $W = \{a, b\}$. Then, we obtain $\alpha = 2nx_{2n+1} + a + b - 2n^2 - n - 1$.

Notice that x_{2n+1} does not necessarily have to be distinct from a and b. We display the solutions for $0 \leq \alpha \leq n-2$ in Table 9. Since we have covered all the

a	b	x_{2n+1}	
0	$n+1+\alpha$	n	$0 \leq \alpha \leq n-2$
1	$2n-1$	n	$\alpha = n-1$
$\alpha - n$	$2n+1$	n	$n \leq \alpha \leq 2n-1$
$n+1$	$2n+1$	n	$\alpha = 2n+1$

Table 9: If $\alpha \in S_{2n+1}(2n+1)$

cases when $r \in \{0, 1, \dots, 2n-1\} \cup \{2n+1\}$, the proof for Lemma 4.7 is complete. □

Now we prove Lemma 4.2 for the case when m is odd.

Lemma 4.8. *Given any $\alpha \in R_{2n+1}$, we can find $x_2, x_3, \ldots, x_{2n+1} \in R_{2n+1}$ such that $x_2, x_3, \ldots, x_{2n}$ are distinct and $\alpha + \sum_{k=2}^{2n} x_k = 2nx_{2n+1}$.*

Proof. First we prove this for when $n > 3$ and then deal with the special cases when $n \leq 3$.

If $n > 3$ and $\alpha \in S_{2n+1}(2n+1)$, by Lemma 4.7, we can select $x_2, x_3, \ldots, x_{2n+1} \in S_{2n+1}(2n+1)$ to satisfy the lemma.

Similarly, if $\alpha \in \{2n^2 + 5n + c : 0 \leq c \leq 2n-1\} \cup \{2n^2 + 7n + 1\}$, we see this is the same set as $S_{2n+1}(2n+1)$ with $2n^2 + 5n$ added to each element. Therefore, also by Lemma 4.7,we can select $x_2, x_3, \ldots, x_{2n+1}$ from the set $\alpha \in \{2n^2 + 5n + c : 0 \leq c \leq 2n-1\} \cup \{2n^2 + 7n + 1\}$ to satisfy the lemma.

If $\alpha = 3n+1$, then let $x_2 = 0$, $x_k = k-1$ for all $3 \leq k \leq 2n$ and $x_{2n+1} = n+1$. Then, $\alpha + \sum_{k=3}^{2n} x_k = 2nx_{2n+1}$.

If $n = 1$, set $x_2 = x_3 = 0$.

If $n = 2$, then the cases for when $\alpha \in \{0, 1, 2, 3, 5\}$ and $\alpha \in \{26, 27, 28, 29, 31\}$ are covered in Lemma 4.7. If $\alpha = 3n + 1 = 7$, $7 + 0 + 2 + 3 = 4 \cdot 3$. If $\alpha = 13$, $13 + 3 + 5 + 7 = 4 \cdot 7$.

If $n = 3$, then the only difference between R_7 and the general definition for R_{2n+1} when $n > 3$ is the the missing 40, so we only need to consider if $\alpha \in \{33, 34, 35, 36, 37, 38\}$. The case for when $\alpha \in \{0, 1, 2, 3, 4, 5, 7\}$ is covered in Lemma 4.7 and if $\alpha = 10$, $10 + 0 + 2 + 3 + 4 + 5 = 6 \cdot 4$.

The solutions for when $33 \leq \alpha \leq 38$ are presented in table 10 below.

α	$\{x_2, x_3, x_4, x_5, x_6\}$	x_7
33	$\{10, 7, 5, 4, 1\}$	10
34	$\{10, 7, 5, 4, 0\}$	10
35	$\{10, 7, 5, 3, 0\}$	10
36	$\{10, 7, 5, 2, 0\}$	10
37	$\{10, 7, 5, 1, 0\}$	10
38	$\{10, 7, 4, 1, 0\}$	10

Table 10: If $33 \leq \alpha \leq 38$ and $n = 3$

Since we have covered all the cases when $m = 2n+1$ is odd, the proof for Lemma 4.8 is complete. □

References

[1] J.-P. Allouche and J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, in Sequences and their applications (Singapore, 1998), Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, pp. 1–16.

[2] F. A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A., 32 (1946), pp. 331–332.

[3] P. Erdős and P. Turán, *On some sequences of integers*, J. London Math. Soc., 11 (1936), pp. 261–264.

[4] P. H. Koester, *An extension of Behrend's theorem*, Online J. Anal. Comb., (2008), pp. Art. 4, 8.

[5] J. W. Layman, *Some properties of a certain nonaveraging sequence*, J. Integer Seq., 2 (1999), p. Article 99.1.3 (HTML document) (electronic).

[6] O. Milenkovic, N. Kashyap, and D. Leyba, *Shortened array codes of large girth*, IEEE Trans. Inform. Theory, 52 (2006), pp. 3707–3722.

[7] I. Z. Ruzsa, *Solving a linear equation in a set of integers. I*, Acta Arith., 65 (1993), pp. 259–282.

[8] ——, *Solving a linear equation in a set of integers. II*, Acta Arith., 72 (1995), pp. 385–397.

[9] A. Shapira, *Behrend-type constructions for sets of linear equations*, Acta Arith., 122 (2006), pp. 17–33.