The Structure of Maximum Subsets of $\{1, \ldots, n\}$ with No Solutions to $a + b = kc$

Andreas Baltz*

Mathematisches Seminar
University of Kiel, D-24098 Kiel, Germany
aba@numerik.uni-kiel.de

Peter Hegarty, Jonas Knape, Urban Larsson

Department of Mathematics
Chalmers University of Technology
Göteborg, Sweden
hegarty@math.chalmers.se {md9jonas,md0larur}@mdstud.chalmers.se

Tomasz Schoen **

Wydzial Matematyki i Informatyki
Adam Mickiewicz University
Poznań, Poland
schoen@amu.edu.pl

Submitted: Nov 9, 2004; Accepted: Apr 22, 2005; Published: Apr 28, 2005
MR Subject Classifications: 05D05, 11P99

Abstract

If k is a positive integer, we say that a set A of positive integers is k-sum-free if there do not exist a, b, c in A such that $a + b = kc$. In particular we give a precise characterization of the structure of maximum sized k-sum-free sets in $\{1, \ldots, n\}$ for $k \geq 4$ and n large.

1 Introduction

A set of positive integers is called k-sum-free if it does not contain elements a, b, c such that

$$a + b = kc,$$

*supported by DFG, Grant SR 7/9 – 2

**research partially supported by KBN Grant 2 PO3A 007 24

where k is a positive integer. Denote by $f(n,k)$ the maximum cardinality of a k-sum-free set in $\{1,\ldots,n\}$. For $k=1$ these extremal sets are well-known: Deshoulliers, Freiman, Sós, and Temkin [1] proved in particular that the maximum 1–sum-free sets in $\{1,\ldots,n\}$ are precisely the set of odd numbers and the "top half" $\{\lceil\frac{n+1}{2}\rceil,\ldots,n\}$. For $n>8$ even $\{\frac{n}{2},\ldots,n-1\}$ forms the only additional extremal set. The famous theorem of Roth [4] gives $f(n,2)=o(n)$. Chung and Goldwasser [2] solved the case $k=3$ by showing that the set of odd integers is the unique extremal set for $n>22$. For $k\geq 4$ they gave an example of a k-sum-free set [3] of cardinality $\frac{k(k-2)}{k^2-2}n+\frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)}n+\mathcal{O}(1)$, which implies $\lim_{n\to\infty}\frac{f(n,k)}{n}\geq\frac{k(k-2)}{k^2-2}+\frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)}$, and they conjectured that this lower bound is the actual value. Moreover they conjectured that extremal k-sum-free sets consist of three intervals of consecutive integers with slight modifications at the end-points if n is large.

In this paper we prove that the first conjecture is true, and we expose a structural result which is very close to the second. Our proof is elementary. In fact it is based on two simple observations:

Suppose we are given a k-sum-free set A. Then

- $kx-y\notin A$ for all $x,y\in A$
  (Otherwise we could satisfy the equation $kx=(kx-y)+y$ in A.)
- for all $y\in A$ any interval centered around $\frac{ky}{2}$ cannot share more than half of its elements with A.
  (Otherwise we would find a pair $\lfloor\frac{ky}{2}\rfloor-d$, $\lceil\frac{ky}{2}\rceil+d$ in A, giving
  $\left(\lfloor\frac{ky}{2}\rfloor-d\right)+\left(\lceil\frac{ky}{2}\rceil+d\right)=ky$.)

2 Preparations

Let $n\in\mathbb{N}$ be large and let $k\in\mathbb{N}_{\geq 4}$. We start by agreeing on some notations.

Notations

Let $A\subseteq\{1,\ldots,n\}$ be a set of positive integers. Denote by

$$s_A:=\min A \text{ and } m_A:=\max A$$

the smallest and the largest elements of A respectively.

For $l,r\in\mathbb{R}$ let

$$\begin{aligned}
(l,r] &:= \{x\in\mathbb{N}\mid l<x\leq r\}\\
[l,r) &:= \{x\in\mathbb{N}\mid l\leq x<r\}\\
(l,r) &:= \{x\in\mathbb{N}\mid l<x<r\}\\
[l,r] &:= \{x\in\mathbb{N}\mid l\leq x\leq r\}
\end{aligned}$$

abbreviate intervals of integers. Continuous intervals will be indicated by the subscript $\mathbb{R}$.

Furthermore for any $y \in \mathbb{N}$ and $d \in \mathbb{N}_0(:= \mathbb{N} \cup \{0\})$ put

$$I_y^d := \left[\frac{ky-1}{2} - d, \frac{ky+1}{2} + d\right].$$

Note that if ky is even then $I_y^d = \left\{\frac{ky}{2} - d, \frac{ky}{2} - d + 1, \ldots, \frac{ky}{2} + d\right\}$ and $|I_y^d| = 2d+1$, while if ky is odd we have $I_y^d = \left\{\frac{ky-1}{2} - d, \ldots, \frac{ky+1}{2} + d\right\}$ and $|I_y^d| = 2d + 2$.

The first Lemma restates our introductory observations.

Lemma 1 *Let $A \subseteq [1, n]$ be a k-sum-free set. If $x, y \in A$ then $kx - y \notin A$. If $y \in A$ and $d \in \mathbb{N}_0$ then $|I_y^d \setminus A| \geq d + 1$.*

Suppose A' is a k-sum-free set consisting of intervals $(l_i, r_i]$. The interval $(l_i, r_i]$ is k-sum-free if $l_i \geq \frac{2r_i}{k}$. Moreover we observe that reasonably large consecutive intervals $(l_{i+1}, r_{i+1}], (l_i, r_i]$ (where we assume $r_{i+1} < l_i$) should satisfy $kr_{i+1} \leq l_i + s_{A'}$. This leads to the following definition, describing a successive transformation of an arbitrary k-sum-free set A into a k-sum-free set of intervals.

Definition 1 *Let $n \in N$ and let $A \subseteq [1, n]$ be k-sum-free with smallest element $s := s_A$. Define sequences (r_i), (l_i), (A_i) by:*

$$\begin{aligned} A_0 &:= A, \quad r_1 := n, \\ l_i &:= \left\lfloor \frac{2r_i}{k} \right\rfloor, \quad r_{i+1} := \left\lfloor \frac{l_i + s}{k} \right\rfloor, \\ A_i &:= (A_{i-1} \setminus (r_{i+1}, l_i]) \cup (l_i, r_i] \cap [s, n] \text{ for } i \geq 1. \end{aligned}$$

The letter $t = t_A$ will be reserved to denote the least integer such that $r_{t+1} < s$. Observe that, for all $i \geq t$,

$$A_i = A_t = [\alpha, r_t] \cup \left(\bigcup_{j=1}^{t-1} (l_j, r_j]\right), \tag{1}$$

where $\alpha = \alpha_A := \max\{l_t + 1, s\}$.

3 The structure of maximum k-sum-free sets

To obtain the structural result we consider the successive transformation of an arbitrary k-sum-free set A into a set A_t of intervals as in (1). Our plan is to show that each member of the transformation sequence (A_i) is k-sum-free and has size greater than or equal to $|A|$. For n sufficiently large, depending on k, and a maximum sized k-sum-free subset A of $[1, n]$, it will turn out that A_t consists of three intervals only, i.e.: that $t = 3$. This observation will do to determine $f(n, k)$, and we conclude our proof by showing that A

could be enlarged if it did not contain (nearly) the whole interval $(l_3, r_3]$ and consequently almost all elements from $(l_2, r_2]$ and $(l_1, r_1]$, so that in fact almost nothing happens during the transformation of an extremal set.

Lemma 2 *Let $A \subseteq [1, n]$ be k-sum-free. Let $i \in \mathbb{N}$.*

a) A_i is k-sum-free.

b) $|A_i| \geq |A_{i-1}|$.

Proof. a) Clearly, it is enough to prove the claim for $i \leq t$, so we may assume that $s \leq r_i$. Suppose there are $a, b, c \in A_i$ with $a + b = kc$. A_i is of the form

$$A_i = A_{i-1} \cap [s, r_{i+1}] \cup (l_i, r_i] \cap [s, n] \cup (l_{i-1}, r_{i-1}] \cup \ldots \cup (l_1, r_1].$$

If $c \in (l_1, r_1]$, then $kc > 2n$, which is impossible. If $i \geq 2$ and $c \in (l_j, r_j]$ for some $j \in [2, i]$, then $kc \in (2r_j, l_{j-1} + s]$ and the larger one of a, b must be in $(r_j, l_{j-1}]$. But $(r_j, l_{j-1}] \cap A_i = \emptyset$ by construction. Hence $c \in A_{i-1} \cap [s, r_{i+1}]$. Now, $kc \leq kr_{i+1} \leq l_i + s$. Since $(r_{i+1}, l_i] \cap A_i = \emptyset$, both a and b have to be in $A_{i-1} \cap [s, r_{i+1}] = A \cap [s, r_{i+1}]$. But A is k-sum-free, a contradiction.

b) The inequality is trivial for $i \geq t$. For $1 \leq i < t$ we have that $l_i \geq s$ and hence

$$A_i = (A_{i-1} \cap [1, r_{i+1}]) \cup (l_i, r_i] \cup \left(\bigcup_{j=1}^{i-1} (l_j, r_j] \right).$$

Thus it suffices to prove that

$$|A_{i-1} \cap [1, r_i]| \leq |A_{i-1} \cap [1, r_{i+1}]| + \left\lceil \frac{(k-2)r_i}{k} \right\rceil.$$

Clearly, then, it suffices to prove the inequality for $i = 1$, i.e.: to prove that, for any $n > 0$, and any k-sum-free subset A of $[1, n]$ with smallest element s_A, we have

$$|A| \leq |A \cap [1, r_{2,A}]| + \left\lceil \frac{(k-2)n}{k} \right\rceil, \tag{2}$$

where

$$r_{2,A} := \left\lfloor \frac{\lfloor 2n/k \rfloor + s_A}{k} \right\rfloor.$$

The proof is by induction on n. The result is trivial for $n = 1$. So suppose it holds for all $1 \leq m < n$ and let A be a k-sum-free subset of $[1, n]$. Note that the result is again trivial if $s_A > 2n/k$, so we may assume that $s_A \leq 2n/k$, which implies that $r_{2,A} \leq n/k$, since $k \geq 4$.

First suppose that there exists $x \in A \cap (n/k, 2n/k]$. Then $1 \leq kx - n \leq n$ and the map $f : y \mapsto kx - y$ is a 1-1 mapping from the interval $[kx - n, n]$ to itself. For each y in this interval, at most one of the numbers y and $f(y)$ can lie in A, since A is k-sum-free. To simplify notation, put $w := kx - n - 1$. Then our conclusion is that

$$|A \cap (w, n]| \leq \frac{1}{2}(n - w). \tag{3}$$

If $w = 0$ or if $A \cap [1, w] = \emptyset$, then we are done (since $k \geq 4$). Put $B := A \cap [1, w]$. Then we may assume $B \neq \emptyset$, hence $s_B = s_A$. Applying the induction hypothesis to B, we find that

$$|B| = |A \cap [1, w]| \leq |B \cap [1, r_{2,B}]| + \left\lceil \frac{(k-2)w}{k} \right\rceil. \tag{4}$$

But $s_B = s_A$ implies that $r_{2,B} \leq r_{2,A}$, hence that $B \cap [1, r_{2,B}] \subseteq A \cap [1, r_{2,A}]$. Thus (3) and (4) yield the inequality

$$|A| \leq |A \cap [1, r_{2,A}]| + \left\lceil \frac{(k-2)w}{k} \right\rceil + \frac{1}{2}(n - w),$$

which in turn implies (2), since $|A|$ is an integer. Thus we are reduced to completing the induction under the assumption that $A \cap (n/k, 2n/k] = \emptyset$. Suppose $x \in A \cap (r_{2,A}, n/k]$. Then $\lfloor 2n/k \rfloor + s_A < kx \leq n$ and $kx - s_A \notin A$. In other words, we can pair off elements in $A \cap (r_{2,A}, 2n/k]$ with elements in $(2n/k, n] \backslash A$. This immediately implies (2), and the proof of Lemma 2 is complete. $\square$

We have seen so far that any k-sum-free set A can be turned into a k-sum-free set A_t having overall size at least $|A|$. The set A_t is a union of intervals, as given by (1), though note that the final interval $[\alpha, r_t]$ may consist of a single point, since $r_t = s$ is possible. The proof of the following Lemma uses a fact shown in [3] by Chung and Goldwasser, to prove that t must be equal to three if $|A|$ is maximum.

Lemma 3 *Let A be a maximum k-sum-free subset of $[1, n]$, where $n > n_0(k)$ is sufficiently large. Let $s := s_A$ and let $t := \max\{i \in \mathbb{N} \mid r_i \geq s\}$. Then $t = 3$.*

Proof. Let A_t be the set of positive integers given by (1). In a similar manner we now define a k-sum-free subset A'_t of $(0, 1]_{\mathbb{R}}$.
Put $c := s/n$ and, for $i = 1, ..., t$ define real numbers R_i, L_i as follows :

$$R_1 := 1, \quad L_i := \frac{2R_i}{k}, \quad R_{i+1} := \frac{L_i + c}{k}.$$

Then we put

$$A'_t := [\alpha', R_t)_{\mathbb{R}} \cup \left(\bigcup_{j=1}^{t-1} [L_j, R_j)_{\mathbb{R}} \right),$$

where $\alpha' := \max\{L_t, c\}$. That A'_t is k-sum-free is shown in [3]. One sees easily that

$$|A_t| \leq n \cdot \mu(A'_t) + t, \tag{5}$$

where μ denotes the Lebesgue-measure. Now suppose that $t \neq 3$. It is shown in [3] that there exists a constant $c_k > 0$, depending only on k, such that in this case

$$|\mu(A'_t)| \leq \frac{k(k-2)}{k^2-2} + \frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)} - c_k. \tag{6}$$

In fact, in the notation of page 8 of [3], an explicit value for c_k (which we will use later) is given by

$$c_k = \frac{2}{k}(R(3) - R(4)),$$

which by definition of R amounts to

$$c_k = \frac{8(k^4-4k^2-4)(k-2)}{(k^6-2k^4-4k^2-8)(k^4-2k^2-4)k}. \tag{7}$$

Now (5) and (6) would imply that

$$|A| \leq \frac{k(k-2)}{k^2-2}n + \frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)}n - c_k n + t.$$

But we have seen in the introduction that $|A| \geq \frac{k(k-2)}{k^2-2}n + \frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)}n + \mathcal{O}(1)$ and, since $t = \mathcal{O}(\log_k n)$, we thus have a contradiction for sufficiently large n. Hence t must equal three, for large enough n, as required. $\square$

Now we are nearly in a position to determine $f(n,k)$. We want to calculate the cardinality of an extremal k-sum-free set A via computing $|A_3|$. Since $|A_3|$ depends on s_A, the following lemma will be helpful :

Lemma 4 *Let $n > n_0(k)$ be sufficiently large. If A is a maximal k-sum-free subset of $[1,n]$, then $S - 2k \leq s_A \leq S + 3$, where $S := \lfloor \frac{8n}{k^5-2k^3-4k} \rfloor$.*

Proof. Set $s := s_A$. By Lemma 3, for $n > n_0(k)$ we have $r_4 < s$. Since A is maximal we have $|A| = |A_3|$. Now, for a fixed n, the cardinality of A_3 is a function of $s \in [1,n]$ only. So we need to show that $|A_3(s)|$ attains its maximum value only for some $s \in [S-2k, S+3]$. Define

$$s' := \min\{s \in [1,n] : l_3(s) < s\}.$$

A tedious computation (see the Appendix below) yields that $s' = S + 1$ if k is even and $s' = S$ or $S + 1$ if k is odd. Hence

$$s' \in [S, S+1]. \tag{8}$$

Clearly,

$$|A_3(s)| = \begin{cases} \lceil \frac{(k-2)n}{k} \rceil + r_2(s) - l_2(s) + r_3(s) - s + 1, & \text{if } s \geq s', \\ \lceil \frac{(k-2)n}{k} \rceil + r_2(s) - l_2(s) + r_3(s) - l_3(s), & \text{if } s < s'. \end{cases} \tag{9}$$

How does $|A_3(s)|$ change (ignoring its maximality for a while) if we alter s?
First suppose $s \geq s'$. If s increases by one, then $|A_3|$ will decrease by one unless either r_2 or r_3 increases. Now r_2 can only increase (by one) once in $k(\geq 4)$ times. Almost the same is true of r_3, though its dependence on l_2 makes things a little more complicated. However, it is not hard to see that we encounter an irreversible decrease in the cardinality of $|A_3|$ after at most 3 steps of increment of s. Hence $|A_3(s)| < |A_3(s')|$ if $s \geq s' + 3$.
Next suppose $s < s'$. If we decrease s, then $|A_3|$ cannot increase at all, since l_i will not decrease unless r_i does. Moreover, $|A_3|$ will become smaller if the size of any interval is diminished. So we can focus our attention on $(l_2, r_2]$. While r_2 decreases once in k times, l_2 does so no more than once in $k\lfloor \frac{k}{2} \rfloor \geq 2k$ times. Thus $|A_3(s)| < |A_3(s'-1)|$ if $s \leq s'-1-2k$.

We have now shown that, as a function of $s \in [1, n]$, the cardinality of A_3 attains its maximum only for some $s \in [s' - 2k, s' + 2]$. This, together with (8), completes the proof of the lemma. $\square$

Now we can prove the first conjecture of Chung and Goldwasser.

Theorem 1

$$\lim_{n\to\infty} \frac{f(n,k)}{n} = \frac{k(k-2)}{k^2-2} + \frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)}.$$

Proof. Let A be a maximum k-sum-free set in $[1, n]$, with n sufficiently large. From Lemma 4 we have $\frac{s_A}{n} = \frac{S^*}{n} + o(1)$, where $S^* = \frac{8n}{k^5-2k^3-4k}$. Thus we can estimate

$$\begin{aligned}
\frac{f(n,k)}{n} &= \frac{|A_3|}{n} = \frac{r_1 - l_1 + r_2 - l_2 + r_3 - S^* + 1}{n} + o(1) \\
&= \frac{1}{n}\left(n - \frac{2n}{k} + \frac{2n + kS^*}{k^2} - \frac{4n + 2kS^*}{k^3} + \frac{4n + 2kS^* + k^3 S^*}{k^4} - S^*\right) + o(1) \\
&= \frac{k^4 - 2k^3 + 2k^2 - 4k + 4}{k^4} + \frac{S^*}{nk^3}(2k^2 - 2k + 2 - k^3) + o(1) \\
&= \frac{k^4 - 2k^3 + 2k^2 - 4k + 4}{k^4} + \frac{8(2k^2 - 2k + 2 - k^3)}{(k^5 - 2k^3 - 4k)k^3} + o(1) \\
&= \frac{k^5 - 2k^4 - 4k + 8}{(k^4 - 2k^2 - 4)k} + o(1) \\
&= \frac{k(k-2)}{k^2-2} + \frac{8(k-2)}{k(k^2-2)(k^4-2k^2-4)} + o(1),
\end{aligned}$$

and the claim follows by taking the limit. $\square$

We can now show the main result.

Theorem 2 *Let $k \in \mathbb{N}_{\geq 4}$ and $n > n_1(k)$. Let S and s' be as in Lemma 4. Let $A \subseteq \{1, \ldots, n\}$ be a k-sum-free set of maximum cardinality, with smallest element $s = s_A$. Then $s \in [S, S+3]$ and $A = \mathcal{I}_3 \cup \mathcal{I}_2 \cup \mathcal{I}_1$, where*

$$\mathcal{I}_3 \in \begin{cases} \{[s, r_3], [s, r_3 + 1]\}, & \text{if } s \geq s' \\ \{[s, r_3), [s, r_3] \setminus \{r_3 - 1\}\}, & \text{if } s < s', \end{cases}$$

$$\mathcal{I}_2 \in \begin{cases} \{[l_2 + 2, r_2], [l_2 + 2, r_2 + 1]\}, & \text{if } r_3 + 1 \in A \\ \{(l_2, r_2], (l_2, r_2 + 1], [l_2, r_2), [l_2, r_2] \setminus \{r_2 - 1\}\}, & \text{if } r_3 + 1 \notin A, \end{cases}$$

$$\mathcal{I}_1 \in \begin{cases} \{[l_1 + 2, n]\}, & \text{if } r_2 + 1 \in A \\ \{[l_1, n), (l_1, n], [l_1, n] \setminus \{n - 1\}\}, & \text{if } r_2 + 1 \notin A, \end{cases}$$

If k is even, then $\mathcal{I}_i \neq [l_i, r_i] \setminus \{r_i - 1\}$ for $1 \leq i \leq 3$.

Remark. Note that Theorem 2 does not precisely determine the k-sum-free subsets of $\{1, ..., n\}$ of maximum size, for every $n > n_1(k)$. With n and k fixed, one first needs to determine for which value(s) of $s \in [S, S+3]$ the quantity $|A_3(s)|$, as given by (9), is maximized. The result will depend on n and k. Even then, for a fixed s, not all the possibilities for $\mathcal{I}_3 \cup \mathcal{I}_2 \cup \mathcal{I}_1$ need be k-sum-free. See Section 4 below for further discussion.

Proof. We have already seen that $|A_3| = |A|$. Our first aim is to show by comparing A_3 with A_2 that almost the whole interval $(l_3, r_3]$ must be in A. Having achieved this, we infer by Lemma 1 that $(r_3, l_2] \cap A$ is nearly empty. Comparing A_2 with A_1 will then reveal that most of $(l_2, r_2]$ is contained in A. Again Lemma 1 will help us to see that A cannot share many elements with $(r_2, l_1]$ and a final comparison of A_1 with A will conclude the proof.

(I) The first aim is easily reached if $s := s_A \geq l_3 + 1$. Simply note that

$$A_2 = (A \cap [s, r_3]) \cup (l_2, r_2] \cup (l_1, r_1] \subseteq [s, r_3] \cup (l_2, r_2] \cup (l_1, r_1] = A_3.$$

The maximality of $|A_2|$ gives $A_2 = A_3$ and hence $[s, r_3] \subseteq A$. Observe that $s > l_3$ together with Lemma 4 and (8) give $S \leq s \leq S + 3$.

Assume now that $s \leq l_3$. We want to show that in this case $s = l_3$. Suppose $s < l_3$ and let $B = [S - 2k, l_3] \cap A$. Define

$$C := I^1_{s_B} \cup \bigcup_{b \in B \setminus \{s_B\}} I^0_b.$$

Clearly $C \subseteq (l_3, r_3]$ for all $n \gg 0$. Then since C is the union of disjoint intervals, Lemma 1 gives that $|C \setminus A| > |B|$. Hence we get the contradiction $|A_3| = |(A_2 \setminus B) \cup (l_3, r_3]| \geq |(A_2 \setminus B) \cup (C \setminus A)| > |A_2| - |B| + |B| = |A_2|$. Therefore we are left with $s = l_3$, and this implies

$$|A_2| = |A_3| \iff |A \cap [s, r_3]| = |(l_3, r_3] \cap [s, r_3]| = |(s, r_3]|. \tag{10}$$

If $r_3 \notin A$ we can infer from (10) that

$$A \cap [s, r_3] = [s, r_3 - 1] = [l_3, r_3 - 1].$$

If $r_3 \in A$, Lemma 1 gives $kl_3 - r_3 \notin A$, so $-k+1 \le kl_3 - 2r_3 \le -1$. If $kl_3 - 2r_3 \le -2$ we get $I^1_{l_3} \subseteq (l_3, r_3]$ and $|I^1_{l_3} \setminus A| \ge 2$, which is impossible since this would imply $|A_3| > |A_2|$. Hence $kl_3 - 2r_3 = -1$ and k is odd. Using (10) one obtains

$$A \cap [s, r_3] = [l_3, r_3] \setminus \{r_3 - 1\}.$$

Suppose now that $s = l_3$ and $r_3 + 1 \in A$. Then $kl_3 - (r_3 + 1) \notin A$ and

$$r_3 - k \le kl_3 - (r_3 + 1) \le r_3 - 1.$$

This contradicts that $[s, r_3 - 2] \subseteq A$ unless $kl_3 - (r_3 + 1) = r_3 - 1$, but then $r_3 \notin A$ and $|A \cap [s, r_3]| = |A \cap [s, r_3 - 2]|$ which contradicts (10). Hence $r_3 + 1 \notin A$ if $s = l_3$.

Finally note that, if $s = l_3$ and $kl_3 \ge 2r_3 - 1$, the latter being a requirement for either of the two possibilities for $\mathcal{I}_3$ to be k-sum-free, then another computation similar to the one in the Appendix yields that $s \ge S$. Again, using Lemma 4 we obtain

$$S \le s \le S + 3, \tag{11}$$

as claimed in the statement of the theorem. This completes the first part of our proof.

(II) For the second part note that we have just shown

$$s \ge l_3. \tag{12}$$

Plugging (11) into the definition of l_3 yields (after a further tedious computation similar to that in the Appendix)

$$S - 1 \le l_3 \le S + 1, \tag{13}$$

which implies in view of (12) and (11)

$$l_3 \le s \le l_3 + 4. \tag{14}$$

Moreover we have observed that $[s, r_3 - 2] \subseteq A$. Let $\xi_1, \ldots, \xi_5 \in \{0, \ldots, k-1\}$ be constants such that

$$\begin{aligned} kl_1 &= 2r_1 - \xi_1 && (15)\\ kr_2 &= l_1 + s - \xi_2 && (16)\\ kl_2 &= 2r_2 - \xi_3 && (17)\\ kr_3 &= l_2 + s - \xi_4 && (18)\\ kl_3 &= 2r_3 - \xi_5. && (19) \end{aligned}$$

We suppose that n is sufficiently large, so we can be sure that

$$[ks-(r_3-2),k(r_3-2)-s]\cap A=\emptyset.$$

By (14) we can infer that

$$\begin{aligned}\emptyset &= [k(l_3+4)-(r_3-2),k(r_3-2)-s]\cap A\\ &= [r_3-\xi_5+4k+2,l_2-\xi_4-2k]\cap A.\end{aligned}$$

Let $J=[r_3+2,r_3-\xi_5+4k+1]\cap A$ and $K=\bigcup_{x\in J}\{kx-(s+2),kx-(s+1),kx-s\}$. Then $K\cap A=\emptyset$, $|K|=3|J|$ and by (18) and (19) we have

$$K\subseteq[l_2-\xi_4+2k-2,l_2-\xi_4-k\xi_5+4k^2+k]\subseteq(l_2+k-2,l_2+4k^2+k]\subseteq(l_2+2,r_2],$$

if $n\gg 0$. Let $B=[l_2-\xi_4-2k+1,l_2]\cap A$. If $B\cup J\subseteq\{l_2\}$ then $A\cap[r_3+2,l_2-1]=\emptyset$. Otherwise, with C as in part (I) if $|B|>1$ we can verify that $C\subseteq[r_2-\frac{3k^2-k+2}{2},r_2]\subseteq(l_2+1,r_2]$, for $n\gg 0$, and $|C\setminus A|>|B|$. Put $C:=\emptyset$ if $|B|\leq 1$. For large n, K and C are disjoint. Hence $|B\cup J|<|(C\setminus A)\cup K|$ and we get

$$|A_2|=|\,[A_1\setminus(J\cup B\cup\{r_3+1\})]\cup(l_2,r_2]|>|A_1\setminus\{r_3+1\}|.$$

Thus if $r_3+1\notin A$ we get $|A_2|>|A_1|$ so suppose $r_3+1\in A$. Then neither l_2 nor l_2+1 can be in A_1. Otherwise, since $(s-\xi_4+k),s-\xi_4+k-1\in[s,s+k]\subseteq[s,r_3-2]\subseteq A$ we get

$$k(r_3+1)=l_2+(s-\xi_4+k)=(l_2+1)+(s-\xi_4+k-1),$$

which is impossible. But $l_2+1\in A_2$, so also in this case it follows that $|A_2|>|A_1|$, since $l_2+1\notin K\cup C$ for large n. Again we conclude that $A\cap[r_3+2,l_2-1]=\emptyset$. Consequently,

$$|A_2|=|A_1|\ \Leftrightarrow\ |A\cap([l_2,r_2]\cup\{r_3+1\})|=|(l_2,r_2]|,$$

which gives $A\cap[l_2,r_2]=[l_2+2,r_2]$ if $r_3+1\in A$. If $r_3+1\notin A$ and either $l_2\notin A$ or $r_2\notin A$, we get $A\cap[l_2,r_2]=(l_2,r_2]$ or $A\cap[l_2,r_2]=[l_2,r_2)$, respectively. In case $r_3+1\notin A$ and both $l_2,r_2\in A$, we see that $kl_2-r_2=r_2-\xi_3\notin A$. If $\xi_3\geq 2$ then $I^1_{l_2}\subseteq(l_2,r_2]$ and l_2 could be profitably replaced. Hence $\xi_3=1$, $A\cap[l_2,r_2]=[l_2,r_2]\setminus\{r_2-1\}$ and k is odd.

(III) For the final interval $(l_1,r_1]$ we use Lemma 1 to conclude from

$$[s,r_3-2]\subseteq A \text{ and } [l_2+2,r_2-2]\subseteq A$$

in view of (16) and (17) that, for $n\gg 0$,

$$\begin{aligned}\emptyset &= A\cap[k(l_2+2)-(r_2-2),k(r_2-2)-(l_2+2)]\\ &= A\cap[r_2-\xi_3+2k+2,l_1+s-\xi_2-2k-l_2-2], \text{ and}\\ \emptyset &= A\cap[k(l_2+2)-(r_3-2),k(r_2-2)-s]\\ &= A\cap[2r_2-\xi_3+2k-r_3+2,l_1-\xi_2-2k]\end{aligned}$$

Let $J = [r_2 + 2, r_2 - \xi_3 + 2k + 1] \cap A$ and $K = \cup_{x \in J}\{kx - s, kx - (s+1), kx - (s+2)\}$. From (14) we have

$$K \subseteq [l_1 - \xi_2 + 2k - 2, l_1 - \xi_2 - k\xi_3 + 2k^2 + k] \subseteq (l_1 + k - 2, r_1], \quad \text{if } n \gg 0.$$

Let $B = [l_1 - \xi_2 - 2k + 1, l_1] \cap A$. If $s_B < l_1$ with C as in (I) we can verify that, for sufficiently large n,

$$C \subseteq \left[\frac{2r_1 - \xi_1 - k\xi_2 - 2k^2 + k - 5}{2}, r_1\right] \subseteq (l_1, r_1],$$

$|C \setminus A| > |B|$ and $\max K < s_C$. By analogy with part (II) we get $A \cap [r_2 + 2, l_1 - 1] = \emptyset$ and the rest of the claim follows as before. □

4 Estimates and Periodicity

We first want to estimate values of $n_i(k)$, $i = 0, 1$, for which Lemmas 3 and 4, and Theorem 2 respectively are valid. The estimates we shall arrive at can probably be improved upon. The example of a k-sum-free set A in [3], referred to in the proof of Lemma 3, satisfies

$$|A| > \frac{k(k-2)}{k^2 - 2} n + \frac{8(k-2)}{k(k^2-2)(k^4 - 2k^2 - 4)} n - 3.$$

Hence the proof of Lemma 3 goes through provided n is sufficiently large so that

$$c_k n - t_0 \geq 3, \tag{20}$$

where $t_0 = t_0(n, k)$ is the largest possible value for t in Definition 1. Now from Definition 1 we easily deduce that, if $i < t$, then $r_{i+1} \leq \left(\frac{4}{k^2}\right) r_i$, and hence that $r_t \leq \left(\frac{4}{k^2}\right)^{t-1} n$. Since $r_t \geq 1$ a priori, we can thus estimate

$$t_0 \leq \frac{1}{2} \log_{k/2} n + 1. \tag{21}$$

Since, by (7), $c_k = \mathcal{O}(\frac{1}{k^6})$, we thus deduce from (18) and (19) that one can take $n_0(k) = \mathcal{O}(k^6)$. It is then an easy and tedious exercise to go through the proof of Theorem 2 and check that one can also take $n_1(k) = \mathcal{O}(k^6)$.

Next, we explain what we mean by the word 'periodicity' in the title of this section. If $k \geq 4$ is even then, for $n > 0$, we have $s' = S + 1 = \lfloor \frac{8n}{k^5 - 2k^3 - 4k} \rfloor + 1$. Hence for a fixed k, if we regard s' as a function of n, then $s'(n) + 1 = s'(n + p_k)$, where $p_k := \frac{k^5 - 2k^3 - 4k}{8}$. For odd k, we define $p_k := k^5 - 2k^3 - 4k$ and in this case, a little more care is required to check that $s'(n) + 8 = s'(n + p_k)$.

Now for any k and n, let $\mathcal{F}(k,n)$ denote the family of maximal k-sum-free subsets of $\{1,...,n\}$. Then for n sufficiently large, as estimated above, and k even (resp. k odd), the map $s \mapsto s+1$ (resp. $s \mapsto s+8$) clearly induces a 1-1 correspondence between the sets in $\mathcal{F}(k,n)$ and $\mathcal{F}(k,n+p_k)$. This is what we mean by 'periodicity'. This observation clearly reduces, for any fixed k, the full classification of all k-sum-free subsets of $\{1,...,n\}$, for all n, to a finite computation.

As an example, we now look at $k=4$. By (7) we compute $c_4 = \frac{47}{48290}$. Then Lemma 3 is valid at least for all n satisfying

$$c_4 n - \frac{1}{2}\log_2 n - 1 \geq 3,$$

which reduces to $n \geq 11008$. One can then check that the proof of Theorem 2 also goes through for all such n. We have $p_4 = 110$. We now present the full classification of all 4-sum-free subsets of $\{1,...,n\}$, valid (at least) for all $n \geq 11008$. This was obtained with the help of a computer.
For each $s,n \in \mathbf{N}$ we define the sets $J_x(s)$, $1 \leq x \leq 13$, as follows (the l_i and r_i are functions of s and n as in Definition 1) :

$$\begin{aligned}
J_1 &= [S, r_3-1] \cup [l_2, r_2-1] \cup [l_1, n-1], \\
J_2 &= [S, r_3-1] \cup [l_2, r_2-1] \cup [l_1+1, n], \\
J_3 &= [S, r_3-1] \cup [l_2+1, r_2] \cup [l_1, n-1], \\
J_4 &= [S, r_3-1] \cup [l_2+1, r_2] \cup [l_1+1, n], \\
J_5 &= [S, r_3-1] \cup [l_2+1, r_2+1] \cup [l_1+2, n], \\
J_6(s) &= [s, r_3] \cup [l_2, r_2-1] \cup [l_1, n-1], \\
J_7(s) &= [s, r_3] \cup [l_2, r_2-1] \cup [l_1+1, n], \\
J_8(s) &= [s, r_3] \cup [l_2+1, r_2] \cup [l_1, n-1], \\
J_9(s) &= [s, r_3] \cup [l_2+1, r_2] \cup [l_1+1, n], \\
J_{10}(s) &= [s, r_3] \cup [l_2+1, r_2+1] \cup [l_1+2, n], \\
J_{11}(s) &= [s, r_3+1] \cup [l_2+2, r_2] \cup [l_1, n-1], \\
J_{12}(s) &= [s, r_3+1] \cup [l_2+2, r_2] \cup [l_1+1, n], \\
J_{13}(s) &= [s, r_3+1] \cup [l_2+2, r_2+1] \cup [l_1+2, n].
\end{aligned}$$

Note that, by Theorem 2, for a given $n \geq 11008$, every maximal 4-sum-free subset of $\{1,...,n\}$ is one of the sets $J_x(s)$, for some $s \in [S, S+3] = [s'-1, s'+2]$. By the remarks above, for each $i \in \{0,...,109\}$, there are natural 1-1 correspondences between the sets in the families $\mathcal{F}(4,n)$ for all $n \equiv i \pmod{110}$. By slight abuse of notation, we denote any such family simply by $\mathcal{F}_i$. Our computer program yielded the following result :

If $|\mathcal{F}_i| = 1$, then $i = 6,7,22,23,46,47,49,51,54,55,57,59,61,70,71,73,75,77,86,87,89$

or 91 and

$$\mathcal{F}_i \quad = \quad \{J_9(s')\},$$

or $i = 36, 37, 100$ or 101 and

$$\mathcal{F}_i \quad = \quad \{J_9(s'+1)\}.$$

If $|\mathcal{F}_i| = 2$, then $\mathcal{F}_i$ is

$$\{J_9(s'), J_9(s'+1)\} \quad \text{if} \quad i = 93, 103, 105, 107,$$

$$\{J_4, J_9(s')\} \quad if \quad i = 9, 11, 13, 25, 27,$$

$$\{J_8(s'), J_9(s')\} \quad if \quad i = 48, 50, 56, 58, 60, 72, 74, 76, 88, 90$$

$$\{J_7(s'), J_9(s')\} \quad if \quad i = 63, 65, 67, 79, 81.$$

If $|\mathcal{F}_i| = 3$:

$$\begin{aligned}
\mathcal{F}_8 = \mathcal{F}_{24} &= \{J_4, J_8(s'), J_9(s')\}, \\
\mathcal{F}_{15} &= \{J_4, J_7(s'), J_9(s')\}, \\
\mathcal{F}_{29} &= \{J_4, J_9(s'), J_9(s'+1)\}, \\
\mathcal{F}_{39} &= \{J_9(s'), J_{12}(s'), J_9(s'+1)\}, \\
\mathcal{F}_{62} = \mathcal{F}_{78} &= \{J_6(s'), J_7(s'), J_9(s')\}, \\
\mathcal{F}_{53} &= \{J_9(s'), J_{10}(s'), J_9(s'+1)\}, \\
\mathcal{F}_{83} &= \{J_7(s'), J_9(s'), J_9(s'+2)\}, \\
\mathcal{F}_{92} &= \{J_8(s'), J_9(s'), J_9(s'+1)\}, \\
\mathcal{F}_{95} = \mathcal{F}_{97} &= \{J_7(s'), J_9(s'), J_9(s'+1)\}, \\
\mathcal{F}_{102} &= \{J_9(s'), J_8(s'+1), J_9(s'+1)\}, \\
\mathcal{F}_{109} &= \{J_9(s'), J_7(s'+1), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 4$:

$$\begin{aligned}
\mathcal{F}_1 = \mathcal{F}_3 = \mathcal{F}_{17} &= \{J_2, J_4, J_7(s'), J_9(s')\}, \\
\mathcal{F}_{10} = \mathcal{F}_{12} = F_{26} &= \{J_3, J_4, J_8(s'), J_9(s')\}, \\
\mathcal{F}_{38} &= \{J_9(s'), J_{12}(s'), J_8(s'+1), J_9(s'+1)\}, \\
\mathcal{F}_{41} = \mathcal{F}_{43} &= \{J_4, J_9(s'), J_{12}(s'), J_9(s'+1)\}, \\
\mathcal{F}_{52} = &= \{J_8(s'), J_9(s'), J_{10}(s'), J_9(s'+1)\}, \\
\mathcal{F}_{64} = \mathcal{F}_{66} = \mathcal{F}_{80} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s')\}, \\
\mathcal{F}_{104} = \mathcal{F}_{106} &= \{J_8(s'), J_9(s'), J_8(s'+1), J_9(s'+1)\}, \\
\mathcal{F}_{69} &= \{J_7(s'), J_9(s'), J_{10}(s'), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 5:$

$$\begin{aligned}
\mathcal{F}_{14} &= \{J_3, J_4, J_6(s'), J_7(s'), J_9(s')\},\\
\mathcal{F}_{19} &= \{J_2, J_4, J_7(s'), J_9(s'), J_9(s'+2)\},\\
\mathcal{F}_{28} &= \{J_3, J_4, J_8(s'), J_9(s'), J_9(s'+1)\},\\
\mathcal{F}_{31} &= \{J_4, J_7(s'), J_9(s'), J_{12}(s'), J_9(s'+1)\},\\
\mathcal{F}_{82} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_9(s'+2)\},\\
\mathcal{F}_{94} &= \{J_6(s'), J_7(s'), J_9(s'), J_8(s'+1), J_9(s'+1)\},\\
\mathcal{F}_{99} &= \{J_7(s'), J_9(s'), J_9(s'+1), J_{10}(s'+1), J_9(s'+2)\},\\
\mathcal{F}_{108} &= \{J_8(s'), J_9(s'), J_6(s'+1), J_7(s'+1), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 6$:

$$\begin{aligned}
\mathcal{F}_{5} &= \{J_2, J_4, J_7(s'), J_9(s'), J_{10}(s'), J_9(s'+1)\},\\
\mathcal{F}_{33} &= \{J_2, J_4, J_7(s'), J_9(s'), J_{12}(s'), J_9(s'+1)\},\\
\mathcal{F}_{45} &= \{J_4, J_9(s'), J_{12}(s'), J_{13}(s'), J_7(s'+1), J_9(s'+1)\},\\
\mathcal{F}_{68} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{10}(s'), J_9(s'+1)\},\\
\mathcal{F}_{85} &= \{J_7(s'), J_9(s'), J_{10}(s'), J_9(s'+1), J_{12}(s'+1), J_9(s'+2)\},\\
\mathcal{F}_{96} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_8(s'+1), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 7$:

$$\begin{aligned}
\mathcal{F}_0 = \mathcal{F}_{16} &= \{J_1, J_2, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s')\},\\
\mathcal{F}_{40} &= \{J_4, J_8(s'), J_9(s'), J_{11}(s'), J_{12}(s'), J_8(s'+1), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 8$:

$$\begin{aligned}
\mathcal{F}_{2} &= \{J_1, J_2, J_3, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s')\},\\
\mathcal{F}_{21} &= \{J_2, J_4, J_7(s'), J_9(s'), J_{10}(s'), J_9(s'+1), J_{12}(s'+1), J_9(s'+2)\},\\
\mathcal{F}_{30} &= \{J_3, J_4, J_6(s'), J_7(s'), J_9(s'), J_{12}(s'), J_8(s'+1), J_9(s'+1)\},\\
\mathcal{F}_{35} &= \{J_2, J_4, J_7(s'), J_9(s'), J_{12}(s'), J_9(s'+1), J_{10}(s'+1), J_9(s'+2)\},\\
\mathcal{F}_{42} &= \{J_3, J_4, J_8(s'), J_9(s'), J_{11}(s'), J_{12}(s'), J_8(s'+1), J_9(s'+1)\},\\
\mathcal{F}_{98} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_8(s'+1), J_9(s'+1), J_{10}(s'+1), J_9(s'+2)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 9$:

$$\begin{aligned}
\mathcal{F}_{18} &= \{J_1, J_2, J_3, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_9(s'+2)\},\\
\mathcal{F}_{84} &= \{J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{10}(s'), J_9(s'+1), J_{12}(s'+1), J_8(s'+2), J_9(s'+2)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 10$:

$$\begin{aligned}
\mathcal{F}_{4} &= \{J_1, J_2, J_3, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{10}(s'), J_9(s'+1)\},\\
\mathcal{F}_{44} &= \{J_3, J_4, J_8(s'), J_9(s'), J_{11}(s'), J_{12}(s'), J_{13}(s'), J_6(s'+1), J_7(s'+1), J_9(s'+1)\}.
\end{aligned}$$

If $|\mathcal{F}_i| = 11, 13$ or 14, we get precisely one family for each size:

$$\begin{aligned}
\mathcal{F}_{32} &= \{J_1, J_2, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{11}(s'), J_{12}(s'), J_8(s'+1), J_9(s'+1)\},\\
\mathcal{F}_{20} &= \{J_1, J_2, J_3, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{10}(s'),\\
&\quad J_9(s'+1), J_{12}(s'+1), J_8(s'+2), J_9(s'+2)\},\\
\mathcal{F}_{34} &= \{J_1, J_2, J_3, J_4, J_6(s'), J_7(s'), J_8(s'), J_9(s'), J_{11}(s'), J_{12}(s'),\\
&\quad J_8(s'+1), J_9(s'+1), J_{10}(s'+1), J_9(s'+2)\}.
\end{aligned}$$

Note, in particular, that $|\mathcal{F}(4,n)| \le 14$ for all sufficiently large n. Computer simulations suggest the same may be true for any even k, with a similar result for odd k, but we leave the investigation of this possibility to a subsequent paper.

Appendix

As a prototype for a type of calculation which appears in several places in the paper, we now show, in the notation of Lemma 4, that $s' = S + 1$ when k is even.

We must investigate the condition $l_3(s) < s$. By definition of l_3 this is just

$$\begin{aligned}
\left\lfloor \frac{2r_3}{k} \right\rfloor < s \quad &\Leftrightarrow \quad \frac{2r_3}{k} < s \Leftrightarrow r_3 < \frac{ks}{2} \Leftrightarrow \left\lfloor \frac{l_2 + s}{k} \right\rfloor < \frac{ks}{2} \Leftrightarrow \frac{l_2+s}{k} < \frac{ks}{2}\\
&\Leftrightarrow \quad l_2 < \left(\frac{k^2}{2} - 1\right) s \Leftrightarrow \frac{2r_2}{k} < \left(\frac{k^2}{2} - 1\right) s \Leftrightarrow r_2 < \left(\frac{k^3}{4} - \frac{k}{2}\right) s\\
&\Leftrightarrow \quad \frac{l_1 + s}{k} < \left(\frac{k^3}{4} - \frac{k}{2}\right) s \Leftrightarrow l_1 < \left(\frac{k^4}{4} - \frac{k^2}{2} - 1\right) s\\
&\Leftrightarrow \quad \frac{2n}{k} < \left(\frac{k^4}{4} - \frac{k^2}{2} - 1\right) s \Leftrightarrow n < \left(\frac{k^5}{8} - \frac{k^3}{4} - \frac{k}{2}\right) s \Leftrightarrow s > \frac{8n}{k^5 - 2k^3 - 4k}\\
&\Leftrightarrow \quad s > S.
\end{aligned}$$

Thus $s' = S + 1$, as required.

Acknowledgments

We would like to thank the anonymous referee whose detailed comments greatly helped us improving our paper.

References

[1] J. M. Deshoulliers, G. Freiman, V. Sós, and M. Temkin, On the structure of sum-free sets II, *Asterisque* **258** (1999), 149-161.

[2] F. R. K. Chung and J. L. Goldwasser, Integer sets containing no solutions to $x+y = 3k$, in "The Mathematics of Paul Erdős", R. L. Graham and J. Nešetřil eds., Springer Verlag, Heidelberg, (1996).

[3] F. R. K. Chung and J. L. Goldwasser, Maximum subsets of $(0, 1]$ with no solutions to $x + y = kz$, *Electron. J. Combin.* **3** (1996), R1.

[4] K. Roth, On certain sets of integers, *J. London Math. Society* **28** (1953), 104-109.