On sum-free and solution-free sets of integers

Andrew Treglown

University of Birmingham

October 23rd 2017

Includes joint work with József Balogh, Hong Liu, Maryam Sharifzadeh; Robert Hancock; and Kitty Meeks.

Introduction

- A set $S \subseteq \mathbb{Z}$ is sum-free if no solutions to $x + y = z$ in S.
- Often we will be working in $[n] := \{1, \dots, n\}$.

Examples:

- $\{1, 2, 4\}$ is not sum-free.
- Set of odds is sum-free.
- $\{n/2 + 1, n/2 + 2, \dots, n\}$ is sum-free.

What do sum-free subsets of $[n]$ look like?

- Every sum-free subset of $[n]$ has size at most $\lceil n/2 \rceil$.

Theorem (Deshouillers, Freiman, Sós and Temkin 1999)

If $S \subseteq [n]$ is sum-free then at least one of the following holds:

(i) $|S| \leq 2n/5 + 1$;

(ii) S *consists of odds;*

(iii) $|S| \leq min(S)$.

Very recently, Tuan Tran refined this result.

What do sum-free subsets of $[n]$ look like?

- Every sum-free subset of $[n]$ has size at most $\lceil n/2 \rceil$.

Theorem (Deshouillers, Freiman, Sós and Temkin 1999)

If $S \subseteq [n]$ is sum-free then at least one of the following holds:

(i) $|S| \leq 2n/5 + 1$;

(ii) *S consists of odds;*

(iii) $|S| \leq min(S)$.

Very recently, Tuan Tran refined this result.

What do sum-free subsets of $[n]$ look like?

- Every sum-free subset of $[n]$ has size at most $\lceil n/2 \rceil$.

Theorem (Deshouillers, Freiman, Sós and Temkin 1999)

If $S \subseteq [n]$ is sum-free then at least one of the following holds:

(i) $|S| \leq 2n/5 + 1$;

(ii) *S consists of odds;*

(iii) $|S| \leq min(S)$.

Very recently, Tuan Tran refined this result.

Examples of sum-free sets:

- Set of odds is sum-free.
- $\{n/2+1, n/2+2, \ldots, n\}$ is sum-free.

These two examples show there are at least $2^{n/2}$ sum-free subsets of $[n]$.

Conjecture (Cameron-Erdős 1990)

The number of sum-free subsets of $[n]$ is $\Theta(2^{n/2})$.

The number of sum-free sets

Examples of sum-free sets:

- Set of odds is sum-free.
- $\{n/2+1, n/2+2, \dots, n\}$ is sum-free.

These two examples show there are at least $2^{n/2}$ sum-free subsets of $[n]$.

Conjecture (Cameron-Erdős 1990)

The number of sum-free subsets of $[n]$ is $\Theta(2^{n/2})$.

The number of sum-free sets

Theorem (Green, Sapozhenko 2003)

There are constants c_e and c_o, s.t. the number of sum-free subsets of $[n]$ is

$$(1+o(1))c_e 2^{n/2}, \text{ or } (1+o(1))c_o 2^{n/2}$$

depending on the parity of n.

- This result doesn't tell us anything about the distribution of the sum-free sets in $[n]$.
- In particular, recall that $2^{n/2}$ sum-free subsets of $[n]$ lie in a single maximal sum-free subset of $[n]$.

The number of maximal sum-free sets

Conjecture (Cameron-Erdős 1999)

There is an absolute constant $c > 0$, s.t. the number of maximal sum-free subsets of $[n]$ is $O(2^{n/2-cn})$.

They also showed there are at least $2^{\lfloor n/4 \rfloor}$ maximal sum-free subsets of $[n]$.

Lower bound construction

There are at least $2^{\lfloor n/4 \rfloor}$ maximal sum-free subsets of $[n]$.

- Suppose n is even. Let S consist of n together with precisely one number from each pair $\{x, n-x\}$ for odd $x < n/2$.
- Notice distinct S lie in distinct maximal sum-free subsets of $[n]$.
- Roughly $2^{n/4}$ choices for S.

The number of maximal sum-free sets

Denote by $f_{\max}(n)$ the number of maximal sum-free subsets in $[n]$. Recall that $f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}$.

Conjecture (Cameron-Erdős 1999)

$$\exists c > 0, \quad f_{\max}(n) = O(2^{n/2-cn}).$$

Theorem (Łuczak-Schoen 2001)

$$f_{\max}(n) \leq 2^{n/2-2^{-28}n} \quad \textit{for large} \quad n$$

Theorem (Wolfovitz 2009)

$$f_{\max}(n) \leq 2^{3n/8+o(n)}.$$

Theorem (Balogh-Liu-Sharifzadeh-T. 2015)

$$f_{\max}(n) = 2^{n/4+o(n)}.$$

The number of maximal sum-free sets

Theorem (Balogh-Liu-Sharifzadeh-T. 2015)

$$f_{\max}(n) = 2^{n/4+o(n)}.$$

Theorem (Balogh-Liu-Sharifzadeh-T. 2015+)

For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.

The number of maximal sum-free sets

Theorem (Balogh-Liu-Sharifzadeh-T. 2015)

$$f_{\max}(n) = 2^{n/4+o(n)}.$$

Theorem (Balogh-Liu-Sharifzadeh-T. 2015+)

For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.

Tools

From additive number theory:

- Container lemma of Green.
- Removal lemma of Green.
- Structure of sum-free sets by Deshouillers, Freiman, Sós and Temkin.

From extremal graph theory: upper bound on the number of maximal independent sets for

- all graphs by Moon and Moser.
- triangle-free graphs by Hujter and Tuza.
- Not too sparse and almost regular graphs.

Sketch of the proof

Theorem (Balogh-Liu-Sharifzadeh-T. 2015)

$$f_{\max}(n) = 2^{n/4+o(n)}.$$

Lemma (Container Lemma, Green)

There exists $\mathcal{F} \subseteq 2^{[n]}$, *s.t.*

(i) $|\mathcal{F}| = 2^{o(n)}$;

(ii) $\forall S \subseteq [n]$ *sum-free*, $\exists F \in \mathcal{F}$, *s.t.* $S \subseteq F$;

(iii) $\forall F \in \mathcal{F}$, $|F| \leq (1/2 + o(1))n$ *and the number of Schur triples in* F *is* $o(n^2)$.

By (i) and (ii), it suffices to show that for every container $A \in \mathcal{F}$,

$$f_{\max}(A) \leq 2^{n/4+o(n)}.$$

Theorem (Balogh-Liu-Sharifzadeh-T. 2015)

$$f_{\max}(n) = 2^{n/4+o(n)}.$$

Lemma (Container Lemma, Green)

There exists $\mathcal{F} \subseteq 2^{[n]}$*, s.t.*

(i) $|\mathcal{F}| = 2^{o(n)}$;

(ii) $\forall S \subseteq [n]$ *sum-free,* $\exists F \in \mathcal{F}$*, s.t.* $S \subseteq F$;

(iii) $\forall F \in \mathcal{F}$, $|F| \leq (1/2 + o(1))n$ *and the number of Schur triples in* F *is* $o(n^2)$.

By (i) and (ii), it suffices to show that for every container $A \in \mathcal{F}$,

$$f_{\max}(A) \leq 2^{n/4+o(n)}.$$

Constructing maximal sum-free sets

Removal+Structural lemmas $\Rightarrow$ classify containers $A \in \mathcal{F}$:

- Case 1: small container, $|A| \leq 0.45n$;
- Case 2: 'interval' container, 'most' of A in $[n/2+1, n]$.
- Case 3: 'odd' container, $|A \setminus O| = o(n)$.

Moreover, in all cases $A = B \cup C$ where B is sum-free and $|C| = o(n)$.

Crucial observation

Every maximal sum-free subset in A can be built in two steps:
(1) Choose a sum-free set S in C;
(2) Extend S in B to a maximal one.

Constructing maximal sum-free sets

Removal+Structural lemmas $\Rightarrow$ classify containers $A \in \mathcal{F}$:

- Case 1: small container, $|A| \leq 0.45n$;
- Case 2: 'interval' container, 'most' of A in $[n/2+1, n]$.
- Case 3: 'odd' container, $|A \setminus O| = o(n)$.

Moreover, in all cases $A = B \cup C$ where B is sum-free and $|C| = o(n)$.

Crucial observation

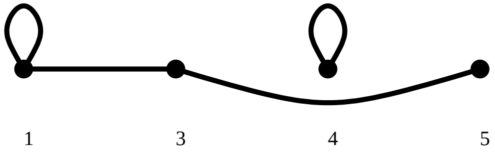
Every maximal sum-free subset in A can be built in two steps:
(1) Choose a sum-free set S in C;
(2) Extend S in B to a maximal one.

Maximal sum-free sets $\Rightarrow$ maximal independent sets

Definition

Given $S, B \subseteq [n]$, the link graph of S on B is $L_S[B]$, where $V = B$ and $x \sim y$ iff $\exists z \in S$ s.t. $\{x, y, z\}$ is a Schur triple.

$L_2[1, 3, 4, 5]$



Maximal sum-free sets $\Rightarrow$ maximal independent sets

Definition

Given $S, B \subseteq [n]$, the link graph of S on B is $L_S[B]$, where $V = B$ and $x \sim y$ iff $\exists z \in S$ s.t. $\{x, y, z\}$ is a Schur triple.

Lemma

Given $S, B \subseteq [n]$ sum-free and $I \subseteq B$, if $S \cup I$ is a maximal sum-free subset of $[n]$, then I is a maximal independent set in $L_S[B]$.

Case 1: small container, $|A| \leq 0.45n$.

Recall $A = B \cup C$, B sum-free, $|C| = o(n)$.

Crucial observation

Every maximal sum-free subset in A can be built in two steps:
(1) Choose a sum-free set S in C;
(2) Extend S in B to a maximal one.

- Fix a sum-free $S \subseteq C$ (at most $2^{|C|} = 2^{o(n)}$ choices).
- Consider link graph $L_S[B]$.
- Moon-Moser: $\forall$ graphs G, $MIS(G) \leq 3^{|G|/3}$.
- So # extensions in (2) is at most $MIS(L_S[B])$,

$$MIS(L_S[B]) \leq 3^{|B|/3} \leq 3^{0.45n/3} \ll 2^{0.249n}.$$

- In total, A contains at most $2^{o(n)} \times 2^{0.249n} \ll 2^{n/4}$ maximal sum-free sets.

Case 1: small container, $|A| \leq 0.45n$.

Recall $A = B \cup C$, B sum-free, $|C| = o(n)$.

Crucial observation

Every maximal sum-free subset in A can be built in two steps:
(1) Choose a sum-free set S in C;
(2) Extend S in B to a maximal one.

- Fix a sum-free $S \subseteq C$ (at most $2^{|C|} = 2^{o(n)}$ choices).
- Consider link graph $L_S[B]$.
- Moon-Moser: $\forall$ graphs G, $MIS(G) \leq 3^{|G|/3}$.
- So # extensions in (2) is at most $MIS(L_S[B])$,

$$MIS(L_S[B]) \leq 3^{|B|/3} \leq 3^{0.45n/3} \ll 2^{0.249n}.$$

- In total, A contains at most $2^{o(n)} \times 2^{0.249n} \ll 2^{n/4}$ maximal sum-free sets.

Cases 2 and 3.

- Now container A could be bigger than $0.45n$.
- This means crude Moon-Moser bound doesn't give accurate bound on $f_{\max}(A)$.
- Instead we obtain more structural information about the link graphs.

- For example, when A 'close' to interval $[n/2+1, n]$ link graphs are triangle-free
- Hujta-Tuza: $MIS(G) \leq 2^{|G|/2}$ for all triangle-free graphs G.
- Gives better bound on $f_{\max}(A)$.

Cases 2 and 3.

- Now container A could be bigger than $0.45n$.
- This means crude Moon-Moser bound doesn't give accurate bound on $f_{\max}(A)$.
- Instead we obtain more structural information about the link graphs.

- For example, when A 'close' to interval $[n/2+1, n]$ link graphs are triangle-free
- Hujta-Tuza: $MIS(G) \leq 2^{|G|/2}$ for all triangle-free graphs G.
- Gives better bound on $f_{\max}(A)$.

Theorem (Balogh-Liu-Sharifzadeh-T. 2015+)

For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.

(i) Count by hand the maximal sum-free sets S that are 'extremal':
   - S that contain precisely one even number.
   - S where $\min(S) \approx n/4$, $\min_2(S) \approx n/2$.

(ii) Count remaining maximal sum-free sets using the container method.

Let $\mathcal{L}$ denote the equation $a_1x_1 + \cdots + a_kx_k = b$ where $a_1, \ldots, a_k, b \in \mathbb{Z}$.

Definitions:

1. $\mathcal{L}$ is translation-invariant if $\sum a_i = b = 0$.
2. A subset $A \subseteq [n]$ is $\mathcal{L}$-free if it does not contain any 'non-trivial' solutions to $\mathcal{L}$.
3. A subset $A \subseteq [n]$ is a maximal $\mathcal{L}$-free set if it is $\mathcal{L}$-free, and if the addition of any further $x \in [n] \setminus A$ would make it no longer $\mathcal{L}$-free.

Fundamental Questions

- **Q1:** What is the size of the largest $\mathcal{L}$-free subset of $[n]$?
- **Q2:** How many $\mathcal{L}$-free subsets of $[n]$ are there?
- **Q3:** How many maximal $\mathcal{L}$-free subsets of $[n]$ are there?

Let $\mu_{\mathcal{L}}(n)$ be the size of the largest $\mathcal{L}$-free subset of $[n]$.

$\mathcal{L}$	$\mu_{\mathcal{L}}(n)$	Comment
$x+y=z$	$\lceil n/2 \rceil$	odds or interval
$x+y=2z$	$o(n)$	Roth's theorem (1953)
$p(x+y)=rz$, $r>2p$	$n-\lfloor 2n/r \rfloor$	union (Hegarty 2007)

In general...

$\mathcal{L}$	$\mu_{\mathcal{L}}(n)$
translation-invariant	$o(n)$
not translation-invariant	$\Omega(n)$

Let $\mu_{\mathcal{L}}(n)$ be the size of the largest $\mathcal{L}$-free subset of $[n]$.

$\mathcal{L}$	$\mu_{\mathcal{L}}(n)$	Comment
$x+y=z$	$\lceil n/2 \rceil$	odds or interval
$x+y=2z$	$o(n)$	Roth's theorem (1953)
$p(x+y)=rz$, $r>2p$	$n-\lfloor 2n/r \rfloor$	union (Hegarty 2007)

In general...

$\mathcal{L}$	$\mu_{\mathcal{L}}(n)$
translation-invariant	$o(n)$
not translation-invariant	$\Omega(n)$

Theorem (Hancock–T. 2017)

Let $\mathcal{L}$ be $px + qy = z$ where $p \geq q$ and $p \geq 2, p, q \in \mathbb{N}$. If n is sufficiently large then $\mu_{\mathcal{L}}(n) = n - \lfloor n/(p+q) \rfloor$.

- We have also determined $\mu_{\mathcal{L}}(n)$ for a range of different equations $\mathcal{L}$ of the form $px + qy = rz$ where $p \geq q \geq r$.
- In each case, the extremal examples are 'intervals' or unions of 'congruency classes'.

Q2: How many $\mathcal{L}$-free subsets of $[n]$?

Let $f(n, \mathcal{L})$ be the number of $\mathcal{L}$-free subsets of $[n]$.
Clearly for any $\mathcal{L}$, we have $f(n, \mathcal{L}) \geq 2^{\mu_{\mathcal{L}}(n)}$.

Theorem (Green, Sapozhenko 2003)

There are constants c_e and c_o, s.t. the number of sum-free subsets of $[n]$ is

$$(1 + o(1))c_e 2^{n/2}, \text{ or } (1 + o(1))c_o 2^{n/2}$$

depending on the parity of n.

Theorem (Green 2005)

Let $\mathcal{L}$ be $a_1x_1 + \cdots + a_kx_k = 0$ where $a_1, \ldots, a_k \in \mathbb{Z}$.
Then $f(n, \mathcal{L}) = 2^{\mu_{\mathcal{L}}(n)+o(n)}$ (where $o(n)$ depends on $\mathcal{L}$).

Together with Hancock (2017) we replaced the term $o(n)$ with a constant in *some* cases.

Q2: How many $\mathcal{L}$-free subsets of $[n]$?

Let $f(n, \mathcal{L})$ be the number of $\mathcal{L}$-free subsets of $[n]$.
Clearly for any $\mathcal{L}$, we have $f(n, \mathcal{L}) \geq 2^{\mu_{\mathcal{L}}(n)}$.

Theorem (Green, Sapozhenko 2003)

There are constants c_e and c_o, s.t. the number of sum-free subsets of $[n]$ is

$$(1 + o(1))c_e 2^{n/2}, \text{ or } (1 + o(1))c_o 2^{n/2}$$

depending on the parity of n.

Theorem (Green 2005)

*Let $\mathcal{L}$ be $a_1 x_1 + \cdots + a_k x_k = 0$ where $a_1, \ldots, a_k \in \mathbb{Z}$.
Then $f(n, \mathcal{L}) = 2^{\mu_{\mathcal{L}}(n) + o(n)}$ (where $o(n)$ depends on $\mathcal{L}$).*

Together with Hancock (2017) we replaced the term $o(n)$ with a constant in *some* cases.

Let $f_{\max}(n, \mathcal{L})$ be the number of maximal $\mathcal{L}$-free subsets of $[n]$.
We have already seen:

Theorem (Balogh-Liu-Sharifzadeh-T. 2015+)

For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.

Q3: How many maximal $\mathcal{L}$-free subsets?

Theorem (Hancock–T. 2017)

Let $\mathcal{L}$ be $px + qy = rz$ where $p, q, r \in \mathbb{Z}$.
Then $f_{\max}(n, \mathcal{L}) \leq 3^{\mu_{\mathcal{L}}(n)/3+o(n)}$.

Bound close to best possible for some equations $\mathcal{L}$. For others way off:

Theorem (Hancock–T. 2017)

Let $\mathcal{L}$ be $qx + qy = z$ where $q \geq 2$ is an integer.
Then $f_{\max}(n, \mathcal{L}) = 2^{n/2q+o(n)}$.

Some complexity

SUM-FREE SUBSET
Input: A finite set $A \subseteq \mathbb{Z}$ and $k \in \mathbb{N}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| = k$?

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is NP-complete.*

Proof extends to all equations $\mathcal{L}$ of form $a_1 x_1 + \cdots + a_\ell x_\ell = by$ where each $a_i \in \mathbb{N}$ and $b \in \mathbb{N}$ are fixed and $\ell \geq 2$

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is APX-hard.*

i.e. there is only a polynomial-time approximation scheme if P=NP

SUM-FREE SUBSET
Input: A finite set $A \subseteq \mathbb{Z}$ and $k \in \mathbb{N}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| = k$?

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is NP-complete.*

Proof extends to all equations $\mathcal{L}$ of form $a_1x_1 + \cdots + a_\ell x_\ell = by$ where each $a_i \in \mathbb{N}$ and $b \in \mathbb{N}$ are fixed and $\ell \geq 2$

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is APX-hard.*

i.e. there is only a polynomial-time approximation scheme if P=NP

Some complexity

SUM-FREE SUBSET
Input: A finite set $A \subseteq \mathbb{Z}$ and $k \in \mathbb{N}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| = k$?

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is NP-complete.*

Proof extends to all equations $\mathcal{L}$ of form $a_1x_1 + \cdots + a_\ell x_\ell = by$ where each $a_i \in \mathbb{N}$ and $b \in \mathbb{N}$ are fixed and $\ell \geq 2$

Theorem (Meeks and T. 2017+)

SUM-FREE SUBSET *is APX-hard.*

i.e. there is only a polynomial-time approximation scheme if P=NP

Some complexity

- Erdős (1965) proved that any set of n non-zero integers contains a sum-free subset of size at least $n/3$.
- Improved to $(n+1)/3$ by Alon and Kleitman (1990) and $(n+2)/3$ by Bourgain (1997)
- Eberhard, Green and Manners (2014) proved that there are sets of n integers whose largest sum-free subset has size $n/3 + o(n)$.

- Erdős (1965) proved that any set of n non-zero integers contains a sum-free subset of size at least $n/3$.
- Improved to $(n+1)/3$ by Alon and Kleitman (1990) and $(n+2)/3$ by Bourgain (1997)
- Eberhard, Green and Manners (2014) proved that there are sets of n integers whose largest sum-free subset has size $n/3 + o(n)$.

ε-Sum-Free Subset
Input: A finite set $A \subseteq \mathbb{Z}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| \geq \varepsilon|A|$?

Theorem (Meeks and T. 2017+)

Given any rational $1/3 < \varepsilon < 1$, ε-Sum-Free Subset is NP-complete.

Question

Is there an fpt-algorithm (parameterised by k) s.t. given a set $A \subset \mathbb{Z}$ it determines whether A has a sum-free set of size $n/3 + k$?

ε-SUM-FREE SUBSET
Input: A finite set $A \subseteq \mathbb{Z}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| \geq \varepsilon |A|$?

Theorem (Meeks and T. 2017+)

Given any rational $1/3 < \varepsilon < 1$, ε-SUM-FREE SUBSET *is NP-complete.*

Question

Is there an fpt-algorithm (parameterised by k*) s.t. given a set* $A \subset \mathbb{Z}$ *it determines whether* A *has a sum-free set of size* $n/3 + k$*?*

Some complexity

ε-Sum-Free Subset
Input: A finite set $A \subseteq \mathbb{Z}$.
Question: Does there exist a sum-free subset $A' \subseteq A$ such that $|A'| \geq \varepsilon|A|$?

Theorem (Meeks and T. 2017+)

Given any rational $1/3 < \varepsilon < 1$, ε-Sum-Free Subset *is NP-complete.*

Question

Is there an fpt-algorithm (parameterised by k) s.t. given a set $A \subset \mathbb{Z}$ it determines whether A has a sum-free set of size $n/3 + k$?

Open problem

Given an abelian group G let $\mu(G)$ denote the size of the largest sum-free subset of G.

Theorem (Green–Ruzsa 2005)

There are $2^{\mu(G)+o(|G|)}$ sum-free subsets of G.

Conjecture (Balogh-Liu-Sharifzadeh-T.)

There are at most $2^{\mu(G)/2+o(|G|)}$ maximal sum-free subsets of G.

- Easy to prove $3^{\mu(G)/3+o(|G|)}$ as an upper bound.