ENUMERATING SOLUTION-FREE SETS IN THE INTEGERS

ROBERT HANCOCK AND ANDREW TREGLOWN

Abstract. Given a linear equation $\mathcal{L}$, a set $A \subseteq [n]$ is $\mathcal{L}$-free if A does not contain any 'non-trivial' solutions to $\mathcal{L}$. In this paper we consider the following three general questions:

(i) What is the size of the largest $\mathcal{L}$-free subset of $[n]$?
(ii) How many $\mathcal{L}$-free subsets of $[n]$ are there?
(iii) How many maximal $\mathcal{L}$-free subsets of $[n]$ are there?

We completely resolve (i) in the case when $\mathcal{L}$ is the equation $px + qy = z$ for fixed $p, q \in \mathbb{N}$ where $p \geq 2$. Further, up to a multiplicative constant, we answer (ii) for a wide class of such equations $\mathcal{L}$, thereby refining a special case of a result of Green [15]. We also give various bounds on the number of maximal $\mathcal{L}$-free subsets of $[n]$ for three-variable homogeneous linear equations $\mathcal{L}$. For this, we make use of container and removal lemmas of Green [15].

1. Introduction

Let $[n] := \{1, \dots, n\}$ and consider a fixed linear equation $\mathcal{L}$ of the form

$$a_1x_1 + \cdots + a_kx_k = b \tag{1.1}$$

where $a_1, \dots, a_k, b \in \mathbb{Z}$. If $b = 0$ we say that $\mathcal{L}$ is *homogeneous*. If

$$\sum_{i \in [k]} a_i = b = 0$$

then we say that $\mathcal{L}$ is *translation-invariant*. Notice that if $\mathcal{L}$ is translation-invariant then $(x, \dots, x)$ is a 'trivial' solution of (1.1) for any x. More generally, a solution $(x_1, \dots, x_k)$ to $\mathcal{L}$ is said to be *trivial* if $\mathcal{L}$ is translation-invariant and if there exists a partition $P_1, \dots, P_\ell$ of $[k]$ so that:

(i) $x_i = x_j$ for every i, j in the same partition class P_r;
(ii) For each $r \in [\ell]$, $\sum_{i \in P_r} a_i = 0$.

A set $A \subseteq [n]$ is *$\mathcal{L}$-free* if A does not contain any non-trivial solutions to $\mathcal{L}$. If the equation $\mathcal{L}$ is clear from the context, then we simply say A is *solution-free*.

The notion of an $\mathcal{L}$-free set encapsulates many fundamental topics in combinatorial number theory. Indeed, in the case when $\mathcal{L}$ is $x_1 + x_2 = x_3$ we call an $\mathcal{L}$-free set a *sum-free set*. This is a notion that dates back to 1916 when Schur [31] proved that, if n is sufficiently large, any r-colouring of $[n]$ yields a monochromatic triple x, y, z such that $x + y = z$. *Sidon sets* (when $\mathcal{L}$ is $x_1 + x_2 = x_3 + x_4$) have also been extensively studied. For example, a classical result of Erdős and Turán [13] asserts that the largest Sidon set in $[n]$ has size $(1 + o(1))\sqrt{n}$. In the case when $\mathcal{L}$ is $x_1 + x_2 = 2x_3$ an $\mathcal{L}$-free set is simply a *progression-free* set. Roth's theorem [24] states that the largest progression-free subset of $[n]$ has size $o(n)$.

In [17] we prove a number of results concerning $\mathcal{L}$-free subsets of $[n]$ where $\mathcal{L}$ is a homogeneous linear equation in *three variables.* In particular, our work is motivated by the following general questions:

(i) What is the size of the largest $\mathcal{L}$-free subset of $[n]$?
(ii) How many $\mathcal{L}$-free subsets of $[n]$ are there?
(iii) How many *maximal* $\mathcal{L}$-free subsets of $[n]$ are there?

We make progress on all three of these questions. For each question we use tools from graph theory; for (i) and (ii) our methods are somewhat elementary. For (iii) our method is more involved and utilises container and removal lemmas of Green [15].

1.1. **The size of the largest solution-free set.** As highlighted above, a central question in the study of $\mathcal{L}$-free sets is to establish the size $\mu_{\mathcal{L}}(n)$ of the largest $\mathcal{L}$-free subset of $[n]$. It is not difficult to see that the largest sum-free subset of $[n]$ has size $\lceil n/2 \rceil$, and this bound is attained by the set of odd numbers in $[n]$ and by the interval $[\lfloor n/2 \rfloor + 1, n]$.

When $\mathcal{L}$ is $x_1 + x_2 = 2x_3$, $\mu_{\mathcal{L}}(n) = o(n)$ by Roth's theorem. In fact, Sanders [27] proved that there is a constant C such that every set $A \subseteq [n]$ with $|A| \geq Cn(\log \log n)^5 / \log n$ contains a three-term arithmetic progression. On the other hand, Behrend [5] showed that there is a constant $c > 0$ so that $\mu_{\mathcal{L}}(n) \geq n \exp(-c\sqrt{\log n})$. See [12, 16] for the best known lower bound on $\mu_{\mathcal{L}}(n)$ in this case.

More generally, it is known that $\mu_{\mathcal{L}}(n) = o(n)$ if $\mathcal{L}$ is translation-invariant and $\mu_{\mathcal{L}}(n) = \Omega(n)$ otherwise (see [25]). For other (exact) bounds on $\mu_{\mathcal{L}}(n)$ for various linear equations $\mathcal{L}$ see, for example, [25, 26, 4, 11, 19].

In [17] we mainly focus on $\mathcal{L}$-free subsets of $[n]$ for linear equations $\mathcal{L}$ of the form $px + qy = z$ where $p \geq 2$ and $q \geq 1$ are fixed integers. Notice that for such a linear equation $\mathcal{L}$, the interval $[\lfloor n/(p+q) \rfloor + 1, n]$ is an $\mathcal{L}$-free set. Our first result implies that this is the largest such $\mathcal{L}$-free subset of $[n]$. Let $\min(S)$ denote the smallest element in a finite set $S \subseteq \mathbb{N}$.

Theorem 1.1. [17] *Let $\mathcal{L}$ denote the equation $px + qy = z$ where $p \geq q$ and $p \geq 2$, $p, q \in \mathbb{N}$. Let n be sufficiently large. Suppose S is an $\mathcal{L}$-free subset of $[n]$, and let $\min(S) = \lfloor \frac{n}{p+q} \rfloor - t$ where t is a non-negative integer.*

(i) *If $0 \leq t < (\frac{p+q-1}{p+q+p/q}) \lfloor \frac{n}{p+q} \rfloor$ then $|S| \leq \lceil \frac{(p+q-1)n}{p+q} \rceil - \lfloor \frac{p}{q} t \rfloor$.*
(ii) *If $t \geq (\frac{p+q-1}{p+q+p/q}) \lfloor \frac{n}{p+q} \rfloor$ then $|S| \leq \frac{(q^2+1)n}{q^2+q+1}$.*

Corollary 1.2. *Let $\mathcal{L}$ denote the equation $px + qy = z$ where $p \geq q$ and $p \geq 2$, $p, q \in \mathbb{N}$. If n is sufficiently large then $\mu_{\mathcal{L}}(n) = n - \lfloor \frac{n}{p+q} \rfloor$.*

Roughly, Theorem 1.1 implies that every $\mathcal{L}$-free subset of $[n]$ is 'interval like' or 'small'. In the case of sum-free subsets (i.e. when $p = q = 1$), a result of Deshouillers, Freiman, Sós and Temkin [10] provides very precise structural information on the sum-free subsets of $[n]$. Loosely speaking, they showed that a sum-free subset of $[n]$ is 'interval like', 'small' or consists entirely of odd numbers.

In the case when $p = q$, Corollary 1.2 was proven by Hegarty [19] (without a lower bound on n).

Very recently we have obtained the exact value of $\mu_{\mathcal{L}}(n)$ for many other linear equations $\mathcal{L}$. Indeed, we determine $\mu_{\mathcal{L}}(n)$ for a wide class of equations of the form $px + qy = rz$ where $p \geq q \geq r$ are fixed natural numbers, as well as for some equations $\mathcal{L}$ in more than three variables. This is work in progress [18].

1.2. **The number of solution-free sets.** Write $f(n,\mathcal{L})$ for the number of $\mathcal{L}$-free subsets of $[n]$. In the case when $\mathcal{L}$ is $x+y=z$, define $f(n) := f(n,\mathcal{L})$.

By considering all possible subsets of $[n]$ consisting of odd numbers, one observes that there are at least $2^{n/2}$ sum-free subsets of $[n]$. Cameron and Erdős [8] conjectured that in fact $f(n)=\Theta(2^{n/2})$. This conjecture was proven independently by Green [14] and Sapozhenko [28]. In fact, they showed that there are constants C_1 and C_2 such that $f(n)=(C_i+o(1))2^{n/2}$ for all $n\equiv i \mod 2$.

Results from [21, 29] imply that there are between $2^{(1.16+o(1))\sqrt{n}}$ and $2^{(6.45+o(1))\sqrt{n}}$ Sidon sets in $[n]$. There are also several results concerning the number of so-called (k,ℓ)-sum-free subsets of $[n]$ (see, e.g., [6, 7, 30]).

More generally, given a linear equation $\mathcal{L}$, there are at least $2^{\mu_\mathcal{L}(n)}$ $\mathcal{L}$-free subsets of $[n]$. In light of the situation for sum-free sets one may ask whether, in general, $f(n,\mathcal{L})=\Theta(2^{\mu_\mathcal{L}(n)})$. However, Cameron and Erdős [8] observed that this is false for translation-invariant $\mathcal{L}$.

Green [15] though showed that given a homogeneous linear equation $\mathcal{L}$, $f(n,\mathcal{L})=2^{\mu_\mathcal{L}(n)+o(n)}$ (where here the $o(n)$ may depend on $\mathcal{L}$). Our next result implies that one can omit the term $o(n)$ in the exponent for certain types of linear equation $\mathcal{L}$.

Theorem 1.3. [17] *Fix $p,q\in\mathbb{N}$ where (i) $q\geq 2$ and $p>q(3q-2)/(2q-2)$ or (ii) $q=1$ and $p\geq 3$. Let $\mathcal{L}$ denote the equation $px+qy=z$. Then*

$$f(n,\mathcal{L})=\Theta(2^{\mu_\mathcal{L}(n)}).$$

1.3. **The number of maximal solution-free sets.** Given a linear equation $\mathcal{L}$, we say that $S\subseteq[n]$ is a *maximal $\mathcal{L}$-free subset of* $[n]$ if it is $\mathcal{L}$-free and it is not properly contained in another $\mathcal{L}$-free subset of $[n]$. Write $f_{\max}(n,\mathcal{L})$ for the number of maximal $\mathcal{L}$-free subsets of $[n]$. In the case when $\mathcal{L}$ is $x+y=z$, define $f_{\max}(n):=f_{\max}(n,\mathcal{L})$.

A significant proportion of the sum-free subsets of $[n]$ lie in just two maximal sum-free sets, namely the set of odd numbers in $[n]$ and the interval $[\lfloor n/2\rfloor+1,n]$. This led Cameron and Erdős [9] to ask whether $f_{\max}(n)=o(f(n))$ or even $f_{\max}(n)\leq f(n)/2^{\varepsilon n}$ for some constant $\varepsilon>0$. Łuczak and Schoen [22] answered this question in the affirmative, showing that $f_{\max}(n)\leq 2^{n/2-2^{-28}n}$ for sufficiently large n. Later, Wolfovitz [32] proved that $f_{\max}(n)\leq 2^{3n/8+o(n)}$. Recently, Balogh, Liu, Sharifzadeh and Treglown [1, 2] proved the following: For each $1\leq i\leq 4$, there is a constant C_i such that, given any $n\equiv i \mod 4$, $f_{\max}(n)=(C_i+o(1))2^{n/4}$.

Except for sum-free sets, the problem of determining the number of maximal solution-free subsets of $[n]$ remains wide open. In [17] we give a number of bounds on $f_{\max}(n,\mathcal{L})$ for homogeneous linear equations $\mathcal{L}$ in three variables. The next result gives a general upper bound for such $\mathcal{L}$. Given a three-variable linear equation $\mathcal{L}$, an *$\mathcal{L}$-triple* is a multiset $\{x,y,z\}$ which forms a solution to $\mathcal{L}$. Let $\mu^*_\mathcal{L}(n)$ denote the number of elements $x\in[n]$ that do not lie in *any* $\mathcal{L}$-triple in $[n]$.

Theorem 1.4. [17] *Let $\mathcal{L}$ be a fixed homogenous three-variable linear equation. Then*

$$f_{\max}(n,\mathcal{L})\leq 3^{(\mu_\mathcal{L}(n)-\mu^*_\mathcal{L}(n))/3+o(n)}.$$

Theorem 1.4 together with the aforementioned result of Green shows that $f_{\max}(n,\mathcal{L})$ is significantly smaller than $f(n,\mathcal{L})$ for all homogeneous three-variable linear equations $\mathcal{L}$ that are not translation-invariant. So in this sense it can be viewed as a generalisation of the result of Łuczak and Schoen. The proof of Theorem 1.4 is a simple application of container and removal lemmas of Green [15]. The same idea was used to prove results in [3, 1, 2]. Although at first sight the bound in Theorem 1.4 may seem crude, perhaps surprisingly there are equations $\mathcal{L}$ where the value of $f_{\max}(n,\mathcal{L})$ is close to this bound (see Proposition 22 in [17]).

On the other hand, the following result shows that there are linear equations where the bound in Theorem 1.4 is far from tight.

Theorem 1.5. [17] *Let $\mathcal{L}$ denote the equation $px+qy=z$ where $p \geq q \geq 2$ are integers so that $p \leq q^2-q$ and $\gcd(p,q)=q$. Then*

$$f_{\max}(n,\mathcal{L}) \leq 2^{(\mu_{\mathcal{L}}(n)-\mu^*_{\mathcal{L}}(n))/2+o(n)}.$$

In the case when $\mathcal{L}$ is the equation $2x+2y=z$ we provide a matching lower bound (see Proposition 26 in [17]). However, for many other linear equations $\mathcal{L}$, the bound in Theorem 1.5 is far from tight. Indeed, this is shown by the following very recent exact result.

Theorem 1.6. [18] *Let $\mathcal{L}$ denote the equation $px+py=z$ where $p \geq 2$ is an integer. Then*

$$f_{\max}(n,\mathcal{L}) = 2^{n/2p+o(n)}.$$

Note that when $p=q=2$, the upper bound in Theorem 1.5 is the same as the bound in Theorem 1.6 (but this is the only case when the bounds meet). The proofs of Theorems 1.5 and 1.6 apply the container and removal lemmas of Green [15]. The former also utilises Theorem 1.1.

Our results suggest that, in contrast to the case of $f(n,\mathcal{L})$, it is unlikely there is a 'simple' general asymptotic formula for $f_{\max}(n,\mathcal{L})$ for all homogeneous linear equations $\mathcal{L}$. It would be extremely interesting to make further progress on this problem.

2. Overview of the proof techniques

In this section we give a brief overview of some of the ideas used to prove the results from [17]. Throughout this section, $\mathcal{L}$ will denote the equation $px+qy=z$ where $p \geq q$ and $p \geq 2$, $p,q \in \mathbb{N}$.

2.1. **The proof of Theorem 1.1.** Suppose that S is an $\mathcal{L}$-free subset of $[n]$ where $m=\min(S)$. We define the graph G_m to have vertex set $[m,n]$ and edges between c and $pm+qc$ for all $c \in [m,n]$ such that $pm+qc \leq n$. Note that S is a subset of the vertex set of G_m. Moreover, as S is $\mathcal{L}$-free it must be an independent set in G_m. (Notice though that we may have independent sets in G_m that *do not* correspond to $\mathcal{L}$-free sets.) Thus to give an upper on $|S|$ it suffices to give such a bound on the size of the largest independent set in G_m.

We observe that the components of G_m are paths. With care one can quantify precisely how many of these 'path components' have a given length. We then use this structural information to bound the size of the largest independent set in G_m.

2.2. **The proof of Theorem 1.3.** Notice that the number of $\mathcal{L}$-free subsets S of $[n]$ with $\min(S)=m$ is bounded from above by the number of independent sets in G_m. Thus, once we have obtained structural information about G_m, Theorem 1.3 follows. Indeed, we give a bound on $f(n,\mathcal{L})$ by summing up the total number of independent sets in the graphs $G_1,\dots,G_n$.

2.3. **Upper bounds on** $f_{\max}(n, \mathcal{L})$**.** To prove Theorems 1.4–1.6 we apply the following *container* result of Green [15].

Lemma 2.1. [15] *Fix a three-variable homogeneous linear equation $\mathcal{L}$. There exists a family $\mathcal{F}$ of subsets of $[n]$ with the following properties:*

(i) *Every $F \in \mathcal{F}$ has at most $o(n^2)$ $\mathcal{L}$-triples.*
(ii) *If $S \subseteq [n]$ is $\mathcal{L}$-free, then S is a subset of some $F \in \mathcal{F}$.*
(iii) $|\mathcal{F}| = 2^{o(n)}$.
(iv) *Every $F \in \mathcal{F}$ has size at most $\mu_{\mathcal{L}}(n) + o(n)$.*

We refer to the elements of $\mathcal{F}$ as *containers*. By Lemma 2.1(ii)–(iii), to prove Theorem 1.4 for example, it suffices to show that in each container $F \in \mathcal{F}$ there are at most $3^{(\mu_{\mathcal{L}}(n) - \mu^*_{\mathcal{L}}(n))/3 + o(n)}$ maximal $\mathcal{L}$-free subsets of $[n]$. To achieve this goal we again translate the problem into one about counting (maximal) independent sets in graphs.

The definition of the auxiliary graphs we use for this is more subtle than the definition of G_m, so we do not define them here. The different upper bounds in Theorems 1.4 and 1.5 arise because in Theorem 1.5 we are able to obtain structural information about these auxiliary graphs; specifically, we show these graphs are triangle-free and then apply a result of Hujter and Tuza [20] that states that any triangle-free graph on N vertices contains at most $2^{N/2}$ maximal independent sets. In the case of Theorem 1.4, we do not use any structural information about our auxiliary graphs and therefore just use a general bound of Moon and Moser [23] which states that *any* graph on N vertices contains at most $3^{N/3}$ maximal independent sets. Similar ideas were used to prove results in [3, 1, 2].

Acknowledgments

The second author is supported by EPSRC grant EP/M016641/1. The research in this conference abstract was presented at the 1st IMA Conference on Theoretical and Computational Discrete Mathematics. The first author would like to thank the Institute of Mathematics and its Applications for a travel grant which allowed him to attend this conference.

References

[1] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, The number of maximal sum-free subsets of integers, *Proc. Amer. Math. Soc.*, 143, (2015), 4713–4721.
[2] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, Sharp bound on the number of maximal sum-free subsets of integers, submitted.
[3] J. Balogh and S. Petříčková, The number of the maximal triangle-free graphs, *Bull. London Math. Soc.*, 46, (2014), 1003–1006.
[4] A. Baltz, P. Hegarty, J. Knape, U. Larsson and T. Schoen, The structure of maximum subsets of $\{1, \ldots, n\}$ with no solutions to $a + b = kc$, *Electron. J. Combin.*, 12, (2005), R19.
[5] F. Behrend. On sets of integers which contain no three terms in arithmetic progression, *Proc. Nat. Acad. Sci.*, 32, (1946), 331–332.
[6] Y. Bilu, Sum-free sets and related sets, *Combinatorica*, 18, (1998), 449–459.
[7] N.J. Calkin and J.M. Thomason, Counting generalized sum-free sets, *J. Number Theory*, 68, (1996), 151–159.
[8] P. Cameron and P. Erdős, On the number of sets of integers with various properties, in Number Theory (R.A. Mollin, ed.), 61–79, Walter de Gruyter, Berlin, 1990.
[9] P. Cameron and P. Erdős, Notes on sum-free and related sets, *Combin. Probab. Comput.*, 8, (1999), 95–107.

[10] J. Deshouillers, G. Freiman, V. Sós and M. Temkin, On the structure of sum-free sets II, *Astérisque*, 258, (1999), 149–161.
[11] K. Dilcher and L. Lucht, On finite pattern-free sets of integers, *Acta Arith.*, 121, (2006), 313–325.
[12] M. Elkin, An improved construction of progression-free sets, *Israel J. Math.*, 184, (2011), 93–128.
[13] P. Erdős and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, *J. London Math. Soc.*, 1, (1941), 212–215.
[14] B. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.*, 36, (2004), 769–778.
[15] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.*, 15, (2005), 340–376.
[16] B. Green and J. Wolf, A note on Elkins improvement of Behrend's construction, *Additive number theory: Festschrift in honor of the sixtieth birthday of Melvyn B. Nathanson*, pages 141–144. Springer-Verlag, 1st edition, 2010.
[17] R. Hancock and A. Treglown, On solution-free sets of integers, *European J. Combin.*, to appear.
[18] R. Hancock and A. Treglown, On solution-free sets of integers II, submitted.
[19] P. Hegarty, Extremal subsets of $\{1,\dots,n\}$ avoiding solutions to linear equations in three variables, *Electron. J. Combin.*, 14, (2007), R74.
[20] M. Hujter and Z. Tuza, The number of maximal independent sets in triangle-free graphs, *SIAM J. Discrete Math.*, 6, (1993), 284–288.
[21] Y. Kohayakawa, S. Lee, V. Rödl, and W. Samotij, The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers, *Random Structures & Algorithms*, 46, (2015), 1–25.
[22] T. Łuczak and T. Schoen, On the number of maximal sum-free sets, *Proc. Amer. Math. Soc.*, 129, (2001), 2205–2207.
[23] J.W. Moon and L. Moser, On cliques in graphs, *Israel J. Math.*, 3, (1965), 23–28.
[24] K.F. Roth, On certain sets of integers, *J. London Math. Soc.*, 28, (1953), 104–109.
[25] I.Z. Ruzsa, Solving a linear equation in a set of integers I, *Acta Arith.*, 65, (1993), 259–282.
[26] I.Z. Ruzsa, Solving a linear equation in a set of integers II, *Acta Arith.*, 72, (1995), 385–397.
[27] T. Sanders, On Roth's theorem on progressions, *Ann. of Math.*, 174, (2011), 619–636.
[28] A.A. Sapozhenko, The Cameron-Erdős conjecture, (Russian) *Dokl. Akad. Nauk.*, 393, (2003), 749–752.
[29] D. Saxton and A. Thomason, Hypergraph containers, *Invent. Math.*, 201, (2015), 925–992.
[30] T. Schoen, The number of $(2,3)$-sum-free subsets of $\{1,\dots,n\}$, *Acta Arith.*, 98, (2001), 155–163.
[31] I. Schur, Uber die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *ber. Deutsch. Mat. Verein.*, 25, (1916), 114–117.
[32] G. Wolfovitz, Bounds on the number of maximal sum-free sets, *European J. Combin.*, 30, (2009), 1718–1723.

School of Mathematics, University of Birmingham, Birmingham, UK
E-mail address: rah410@bham.ac.uk

School of Mathematics, University of Birmingham, Birmingham, UK
E-mail address: a.c.treglown@bham.ac.uk