A NOTE ON SOLUTION-FREE SETS OF 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 $\mathcal{L}$. We determine the precise size of the largest $\mathcal{L}$-free subset of $[n]$ for several general classes of linear equations $\mathcal{L}$ of the form $px + qy = rz$ for fixed $p, q, r \in \mathbb{N}$ where $p \geq q \geq r$. Further, for *all* such linear equations $\mathcal{L}$, we give an upper bound on the number of maximal $\mathcal{L}$-free subsets of $[n]$. In the case when $p = q \geq 2$ and $r = 1$ this bound is exact up to an error term in the exponent. We make use of container and removal lemmas of Green [12] to prove this result. Our results also extend to various linear equations with more than three variables.

MSC2010: 11B75, 05C69.

1. Introduction

In this note we study solution-free sets of integers, that is, sets that contain no solution to a given linear equation $\mathcal{L}$. In particular, we investigate the size of the largest such subset of $[n] := \{1, \dots, n\}$ and the number of maximal solution-free subsets of $[n]$.

More precisely, consider a fixed linear equation $\mathcal{L}$ of the form

$$a_1x_1 + \cdots + a_kx_k = b$$

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*. 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*.

Many of the most famous results in combinatorial number theory concern solution-free sets. For example, Schur's theorem [25] states that if n is sufficiently large then $[n]$ cannot be partitioned into r *sum-free* sets (i.e. $\mathcal{L}$ is $x + y = z$). Roth's theorem [20] states that the largest *progression-free* set (i.e. $\mathcal{L}$ is $x_1 + x_2 = 2x_3$) has size $o(n)$ whilst a classical result of Erdős and Turán [10] determines, up to an error term, the size of the largest *Sidon set* (i.e. $\mathcal{L}$ is $x_1 + x_2 = x_3 + x_4$).

1.1. The size of the largest solution-free set.

As indicated above, a key 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 $[n]$. In general $\mu_{\mathcal{L}}(n) = o(n)$ if $\mathcal{L}$ is translation-invariant and $\mu_{\mathcal{L}}(n) = \Omega(n)$ otherwise (see [21]). When $\mathcal{L}$ is $x + y = z$ it is easy to see that $\mu_{\mathcal{L}}(n) = \lceil n/2 \rceil$. Recently, the authors [14] established that if $\mathcal{L}$ is the equation $px + qy = z$ with $p, q \in \mathbb{N}$, $p \geq 2$, then $\mu_{\mathcal{L}}(n) = n - \lfloor n/(p+q) \rfloor$ for sufficiently large n. This bound is attained by the interval $[\lfloor n/(p+q) \rfloor + 1, n]$. For other (exact) bounds on $\mu_{\mathcal{L}}(n)$ for various linear equations $\mathcal{L}$ see, for example, [21, 22, 6, 9, 15].

Date: May 6, 2016.

The second author is supported by EPSRC grant EP/M016641/1.

Our first result determines $\mu_{\mathcal{L}}(n)$ for a wide class of equations of the form $px + qy = rz$ where $p \geq q \geq r$ and p, q, r are fixed positive integers.

Theorem 1. *Let $\mathcal{L}$ denote the equation $px + qy = rz$ where $p \geq q \geq r$ and p, q, r are fixed positive integers satisfying $\gcd(p, q, r) = 1$. Let $t := \gcd(p, q)$. Write $r_1 := p/t$ and $r_2 := q/t$.*

(i) If q divides p and $p + q \leq rq$ then $\mu_{\mathcal{L}}(n) = \lceil (q-1)n/q \rceil$;

(ii) If q divides p and $p + q \geq rq$ then $\mu_{\mathcal{L}}(n) = \lceil (p+q-r)(n-a)/(p+q) \rceil + a$ where a is the unique non-negative integer $0 \leq a < q$ such that $n - a$ is divisible by q;

(iii) If q does not divide p, $t > 1$ and

$$r > (r_1 r_2 - r_1 - r_2 + 4) r_2 \left(r_1 + 1 + \frac{r_2 - 1}{r_1^2 + (r_1 - 1)(r_2 - 1)} \right)$$

then $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil$.

Theorem 1(ii) was already proven (for large enough n) in [14] in the special case when $r = 1$. (Note though that our work in [14] determines $\mu_{\mathcal{L}}(n)$ for many equations $\mathcal{L}$ not covered by Theorem 1.) Previously, Hegarty [15] proved Theorem 1(i) in the case when $p = q$. In Section 3 we also give a generalisation of Theorem 1 concerning some linear equations with more variables (see Corollary 14).

Notice that in the case when q divides p, Theorem 1 gives a dichotomy for the value of $\mu_{\mathcal{L}}(n)$: when $p + q \leq rq$ the set $A := \{x \in [n] : x \not\equiv 0 \mod q\}$ is a largest $\mathcal{L}$-free subset of $[n]$, whilst when $p + q \geq rq$ the interval $I := [\lfloor r(n-a)/(p+q) \rfloor + 1, n]$ is a largest $\mathcal{L}$-free subset of $[n]$. To see that I is $\mathcal{L}$-free observe that since $\gcd(p, q, r) = 1$ and $\gcd(p, q) = q$, any solution (x, y, z) to $\mathcal{L}$ with $x, y, z \in I$ must have z divisible by q. Since $px + qy > r(n-a)$, z must lie in $[n - a + 1, n]$ however then z is not divisible by q and so I is $\mathcal{L}$-free. Note that when $r = 1$, $I := [\lfloor r(n-a)/(p+q) \rfloor + 1, n] = [\lfloor rn/(p+q) \rfloor + 1, n]$, though this does not hold in general when $r > 1$.

Theorem 1 does not provide us with as much information for the case when q does not divide p; it would be interesting to establish whether a similar dichotomy occurs in this case.

1.2. **The number of maximal solution-free sets.** Given a linear equation $\mathcal{L}$, write $f(n, \mathcal{L})$ for the number of $\mathcal{L}$-free subsets of $[n]$. Observe that all possible subsets of an $\mathcal{L}$-free set are also $\mathcal{L}$-free, and so $f(n, \mathcal{L}) \geq 2^{\mu_{\mathcal{L}}(n)}$. In fact, in general this trivial lower bound is not too far from the precise value of $f(n, \mathcal{L})$. Indeed, Green [12] showed that given a homogenous linear equation $\mathcal{L}$, then $f(n, \mathcal{L}) = 2^{\mu_{\mathcal{L}}(n) + o(n)}$ (where here the $o(n)$ may depend on $\mathcal{L}$). (Recall though, in the case when $\mathcal{L}$ is translation-invariant $\mu_{\mathcal{L}}(n) = o(n)$, so Green's theorem only tells us that $f(n, \mathcal{L}) = 2^{o(n)}$ for such $\mathcal{L}$.) In the case of sum-free sets, Green [11] and independently Sapozhenko [23], showed that there are constants C_1 and C_2 such that $f(n, \mathcal{L}) = (C_i + o(1))2^{n/2}$ for all $n \equiv i \mod 2$. This resolved a conjecture of Cameron and Erdős [7].

Far less is known about the number $f_{\max}(n, \mathcal{L})$ of maximal $\mathcal{L}$-free subsets of $[n]$. (We say that $A \subseteq [n]$ 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 $[n]$.) In the case when $\mathcal{L}$ is $x + y = z$, Cameron and Erdős [8] asked whether $f_{\max}(n, \mathcal{L}) \leq f(n, \mathcal{L})/2^{\varepsilon n}$ for some constant $\varepsilon > 0$; a few years later Łuczak and Schoen [19] confirmed this to be true. After further progress on the problem [26, 2], Balogh, Liu, Sharifzadeh and Treglown [3] proved the following sharp result for maximal sum-free sets: For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $f_{\max}(n, \mathcal{L}) = (C_i + o(1))2^{n/4}$.

For other linear equations $\mathcal{L}$, it is also natural to ask whether there are significantly fewer maximal $\mathcal{L}$-free subsets of $[n]$ than there are $\mathcal{L}$-free subsets. In [14] we showed that this is the case for all non-translation-invariant three-variable homogeneous equations $\mathcal{L}$. In particular in this case $f_{\max}(n, \mathcal{L}) \leq 3^{(\mu_{\mathcal{L}}(n) - \mu^*_{\mathcal{L}}(n))/3 + o(n)}$ where $\mu^*_{\mathcal{L}}(n)$ denotes the number of elements in $[n]$ which do not lie in *any* non-trivial solution to $\mathcal{L}$ that only consists of elements from $[n]$. We also gave other

upper bounds on $f_{\max}(n, \mathcal{L})$ for equations of the form $px + qy = z$ for fixed $p, q \in \mathbb{N}$ which in most cases yielded better bounds (see [14]). In this paper we prove the following result.

Theorem 2. *Let $\mathcal{L}$ denote the equation $px + qy = rz$ where $p \geq q \geq r$ and p, q, r are fixed positive integers satisfying $\gcd(p, q, r) = 1$. Let $t := \gcd(p, q)$. Then*

$$f_{\max}(n, \mathcal{L}) \leq 2^{\frac{Crn}{q} + o(n)} \text{ where } C := 1 - \frac{t}{p+q}\Big(\frac{p^2 + (p-t)(q-t)}{p^2}\Big).$$

For a wide class of equations $\mathcal{L}$ this is the current best known upper bound on $f_{\max}(n, \mathcal{L})$; see the appendix for more details. In the case when $p = q \geq 2$ and $r = 1$, the upper bound given by Theorem 2 is actually exact up to the error term in the exponent.

Theorem 3. *Let $\mathcal{L}$ denote the equation $qx + qy = z$ where $q \geq 2$ is an integer. Then*

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

In Section 4 we will also generalise Theorem 2 to consider some linear equations with more variables (see Theorem 15).

For the proof of both Theorems 1 and 2, a simple but crucial tool is a result (Lemma 8) which ensures a certain auxiliary graph contains a large collection of disjoint edges. To prove Theorem 2 we also make use of container and removal lemmas of Green [12] (see Section 2.1).

In the next section we collect together a number of useful tools and lemmas. We prove our results on the size of the largest solution-free subset of $[n]$, and on the number of maximal solution-free subsets of $[n]$, in Section 3 and 4 respectively.

2. Containers, link hypergraphs and the main lemmas

2.1. **Container and removal lemmas.** Observe that we can formulate the study of $\mathcal{L}$-free sets in terms of independent sets in hypergraphs. Let H denote the hypergraph with vertex set $[n]$ and edges corresponding to non-trivial solutions to $\mathcal{L}$. Then an independent set in H is precisely an $\mathcal{L}$-free set. The method of containers roughly states that for certain (hyper)graphs G, the independent sets of G lie only in a small number of subsets of $V(G)$ called *containers*, where each container is an 'almost independent set'. In general, the method of containers has had a wide number of applications to combinatorics and other areas; the method for *graphs* was first developed by Kleitman and Winston [16, 17]. More recently, the *hypergraph* container method was developed by Balogh, Morris and Samotij [4] and independently by Saxton and Thomason [24]. In this section we introduce a container lemma of Green [12] for $\mathcal{L}$-free sets.

Lemma 4(i)–(iii) is stated explicitly in Proposition 9.1 of [12]. Lemma 4(iv) follows as an immediate consequence of Lemma 4(i) and Lemma 5 below.

Lemma 4. [12] *Fix a k-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 $o(n^{k-1})$ solutions to $\mathcal{L}$.
(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 call the elements of $\mathcal{F}$ *containers*. Observe that Lemma 4(iv) gives a bound on the size of the containers in terms of $\mu_{\mathcal{L}}(n)$, even in the case when $\mu_{\mathcal{L}}(n)$ is not known.

The following removal lemma is a special case of a result of Green (Theorem 1.5 in [12]). This result was also generalised to systems of linear equations by Král', Serra and Vena (Theorem 2 in [18]).

Lemma 5. [12] *Fix a k-variable homogeneous linear equation $\mathcal{L}$. Suppose that $A \subseteq [n]$ is a set containing $o(n^{k-1})$ solutions to $\mathcal{L}$. Then there exist B and C such that $A = B \cup C$ where B is $\mathcal{L}$-free and $|C| = o(n)$.*

We will use both the above results to obtain bounds on the number of maximal $\mathcal{L}$-free sets.

2.2. **Link hypergraphs.** One can turn the problem of counting the number of maximal $\mathcal{L}$-free subsets of $[n]$ into one of counting maximal independent sets in an auxiliary graph. Similar techniques were used in [26, 2, 3, 14], and in the graph setting in [5, 1]. To be more precise let B and S be disjoint subsets of $[n]$ and fix a three-variable linear equation $\mathcal{L}$. The *link graph $L_S[B]$ of S on B* has vertex set B, and an edge set consisting of the following two types of edges:

(i) Two vertices x and y are adjacent if there exists an element $z \in S$ such that $\{x, y, z\}$ is an $\mathcal{L}$-triple;

(ii) There is a loop at a vertex x if there exists an element $z \in S$ or elements $z, z' \in S$ such that $\{x, x, z\}$ or $\{x, z, z'\}$ is an $\mathcal{L}$-triple.

Here by *$\mathcal{L}$-triple* we simply mean a multiset $\{x, y, z\}$ which forms a solution to $\mathcal{L}$.

Consider the following generalisation of a link graph $L_S[B]$ to that of a *link hypergraph*: Let B and S be disjoint subsets of $[n]$ and let $\mathcal{L}$ denote the equation $p_1x_1 + \cdots + p_kx_k = 0$ where $p_1, \ldots, p_k$ are fixed integers. The *link hypergraph $L_S[B]$* of S on B has vertex set B; It has an edge set consisting of hyperedges between $s \leq k$ distinct vertices $v_1, \ldots, v_s$ of B, whenever $(x_1, \ldots, x_k)$ is a solution to $\mathcal{L}$ in which $\{x_1, \ldots, x_k\} \subseteq S \cup \{v_1, \ldots, v_s\}$ and $\{v_1, \ldots, v_s\} \subseteq \{x_1, \ldots, x_k\}$. In this definition one could have edges corresponding to trivial solutions. However in our applications, since we only consider non-translation-invariant equations, there are no trivial solutions.

The link graph lemmas used by the authors in [14] (Lemmas 12 and 15) can easily be extended to the hypergraph case.

Lemma 6. *Let $\mathcal{L}$ denote a non-translation-invariant linear equation. Suppose that B, S are disjoint $\mathcal{L}$-free subsets of $[n]$. If $I \subseteq B$ is such that $S \cup I$ is a maximal $\mathcal{L}$-free subset of $[n]$, then I is a maximal independent set in the link hypergraph $L_S[B]$.* □

Let $\mathrm{MIS}(G)$ denote the number of maximal independent sets in G. The above result can be used in conjunction with the container lemma as follows. Let $F = A \cup B$ be a container as in Lemma 4 where $|A| = o(n)$ and B is $\mathcal{L}$-free. Observe that any maximal $\mathcal{L}$-free subset of $[n]$ in F can be found by first selecting an $\mathcal{L}$-free subset $S \subseteq A$, and then extending S in B. Then the number of extensions of S in B is bounded by $\mathrm{MIS}(L_S[B])$ by Lemma 6.

We can also use link graphs to obtain lower bounds.

Lemma 7. *Let $\mathcal{L}$ denote a non-translation-invariant linear equation. Suppose that B, S are disjoint $\mathcal{L}$-free subsets of $[n]$. Let H be an induced subgraph of the link graph $L_S[B]$. Then $f_{\max}(n, \mathcal{L}) \geq \mathrm{MIS}(H)$.* □

2.3. **The main lemmas.** Here we use a specific link graph as a means to bound the number of elements in a solution-free subset of $[n]$.

Let $\mathcal{L}$ denote the equation $px + qy = rz$ where $p \geq q \geq r$ and p, q, r are fixed positive integers satisfying $\gcd(p, q, r) = 1$. Let $t := \gcd(p, q)$ and write $r_1 := p/t$, $r_2 := q/t$. Fix $M \in [n]$ such that M is divisible by t. We define the graph G_M to have vertex set $[\lceil rM/q \rceil - 1]$ and an edge between x and y whenever $px + qy = rM$.

Lemma 8. *The graph G_M contains a collection E of vertex-disjoint edges where*

$$|E| = \left\lfloor \frac{rM}{r_2(p+q)} \right\rfloor + (r_1r_2 - r_1 - r_2 + 1) \left\lfloor \left\lfloor \frac{rM}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor \frac{1}{r_1r_2} \right\rfloor$$

and at most one edge in E is a loop.

Proof. All edges in G_M are pairs of the form $\{s, (rM - sp)/q\}$ for some $s \in \mathbb{N}$ since $ps + q(rM - sp)/q = rM$. Since $p = r_1 t$ and $q = r_2 t$ where r_1 and r_2 are coprime, for a fixed integer s precisely one element in $\{(rM - (s-j)p)/q : 0 \le j < r_2\}$ is an integer. (Note here we are using that M is divisible by t.) In other words there exists a unique $x \in \mathbb{N}$, $1 \le x \le r_2$ such that $(rM - xp)/q$ is an integer, and all edges in G_M are of the form $\{x + ar_2, (rM - xp)/q - ar_1\}$ for some non-negative integer a. In particular there is an edge provided a satisfies $(rM - xp)/q - ar_1 \in \mathbb{N}$.

Write $y := (rM - xp)/q$. Note that if $x + ar_2 \le rM/(p+q)$ then

$$y - ar_1 = \frac{rM - (x + ar_2)p}{q} \ge \frac{rM - prM/(p+q)}{q} = \frac{rM}{p+q}.$$

Hence there are $\lfloor rM/(r_2(p+q)) \rfloor$ distinct edges in G_M of the form $\{x + ar_2, y - ar_1\}$ with $x + ar_2 \le rM/(p+q) \le y - ar_1$. Note that one of these edges may be a loop. (This will be at $rM/(p+q)$ in the case when $rM/(p+q) \in \mathbb{N}$.) Call this collection of edges E_1. Our next aim is to find an additional collection E_2 of edges in G_M that is vertex-disjoint from E_1.

Note that $x + ar_2 \equiv x \bmod r_2$ and $y - ar_1 \equiv y \bmod r_1$. Also $p(rM/p) + q(0) = rM$ and $\lceil rM/p \rceil \le \lceil rM/q \rceil$, hence there are at least

$$\left\lfloor \left(\left\lceil \frac{rM}{p} \right\rceil - 1 - \left\lfloor \frac{rM}{p+q} \right\rfloor \right) / r_2 \right\rfloor \ge \left\lfloor \left(\frac{rM}{p} - \frac{rM}{p+q} - 1 \right) / r_2 \right\rfloor = \left\lfloor \frac{rM}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor$$

edges in G_M of the form $\{x + ar_2, y - ar_1\}$ with $x + ar_2 > rM/(p+q)$. Consider a set of $r_1 r_2$ edges $\{\{x + ar_2 + br_2, y - ar_1 - br_1\} : 0 \le b < r_1 r_2\}$ for a fixed a. Since r_1 and r_2 are coprime, precisely r_2 of these edges (1 in r_1 of them) have $x + ar_2 + br_2 \equiv y \bmod r_1$, and precisely r_1 of these edges have $y - ar_1 - br_1 \equiv x \bmod r_2$. (Also, precisely 1 edge satisfies both.) In all other cases since $x + ar_2 + br_2 \not\equiv y \bmod r_1$ and $y - ar_1 - br_1 \not\equiv x \bmod r_2$, the edge $\{x + ar_2 + br_2, y - ar_1 - br_1\}$ is vertex-disjoint from E_1. Hence we obtain a set E_2 of at least $(r_1 r_2 - r_1 - r_2 + 1)\lfloor \lfloor rM/(r_1(p+q)) - 1/r_2 \rfloor / (r_1 r_2) \rfloor$ additional distinct edges. Thus $E := E_1 \cup E_2$ is our desired set. □

Observe that the graph G_M is a link graph $L_S[B]$, where $S := \{M\}$ and $B := [\lceil rM/q \rceil - 1]$. If we wish to extend a solution-free set S into a solution-free subset of $S \cup B$, then we must pick an independent set in $L_S[B]$. Similarly here if we wish to obtain a solution-free subset of $[n]$ which contains M divisible by t, then we must pick an independent set in G_M. This is the idea behind the following key lemma, which allows us to bound the number of elements in such an $\mathcal{L}$-free set.

Lemma 9. *Let $\mathcal{L}$ denote the equation $px + qy = rz$ where $p \ge q \ge r$ and p, q, r are fixed positive integers satisfying $\gcd(p, q, r) = 1$. Let $t := \gcd(p, q)$ and write $r_1 := p/t$ and $r_2 := q/t$. Let S be an $\mathcal{L}$-free subset of $[n]$. If $M \in S$ is divisible by t, then S contains at most*

$$\left\lceil \frac{rM}{q} \right\rceil - 1 - \left\lfloor \frac{rM}{r_2(p+q)} \right\rfloor - (r_1 r_2 - r_1 - r_2 + 1) \left\lfloor \left\lfloor \frac{rM}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor \frac{1}{r_1 r_2} \right\rfloor$$

elements from $[\lceil rM/q \rceil - 1]$.

Proof. Consider the graph G_M and observe that its edges correspond to $\mathcal{L}$-triples: since $p \ge q \ge r$ there is an edge between x and y precisely when $\{x, y, M\}$ is an $\mathcal{L}$-triple. Hence if $I \subseteq V(G_M)$ is such that $I \cup \{M\}$ is an $\mathcal{L}$-free subset of $[n]$ then I is an independent set in G_M. As a consequence if we find a set of vertex-disjoint edges in G_M of size J, then S contains at most $\lceil rM/q \rceil - 1 - J$ elements from $[\lceil rM/q \rceil - 1]$. The result then follows by applying Lemma 8. □

First note that if $\mathcal{L}$ denotes the equation $x + y = z$, then in Lemma 9 we are simply saying that if a sum-free set S contains M, then it cannot contain both 1 and $M - 1$, it cannot contain both 2 and $M - 2$, and so on. So in a sense this lemma is a generalisation of the proof that sum-free subsets of $[n]$ cannot contain more than $\lceil n/2 \rceil$ elements.

Let $\mathcal{L}$ denote the equation $px + qy = rz$ where $p \geq q \geq r$ and p, q, r are fixed positive integers satisfying $\gcd(p, q, r) = 1$ and let $t := \gcd(p, q)$. Then observe that the set $T := \{x \in [n] : x \not\equiv 0 \mod t\}$ is $\mathcal{L}$-free: note that in any solution $\{x, y, z\}$ to $\mathcal{L}$, z must be divisible by t since $\gcd(r, t) = 1$. But T contains no elements divisible by t. Lemma 9 roughly implies that every $\mathcal{L}$-free subset of $[n]$ must have 'not too many small elements' or must 'look like' T. Clearly this lemma gives rise to an upper bound on the size of the largest $\mathcal{L}$-free subset of $[n]$. In Section 4 we also show that this lemma can be used to obtain an upper bound on the number of maximal $\mathcal{L}$-free subsets of $[n]$.

The following simple proposition allows us to extend our results for linear equations with three variables to linear equations with more than three variables.

Proposition 10. *Let $\mathcal{L}_1$ denote the equation $p_1x_1 + \cdots + p_kx_k = b$ where $p_1, \ldots, p_k, b \in \mathbb{Z}$ and let $\mathcal{L}_2$ denote the equation $(p_1 + p_2)x_1 + p_3x_2 + \cdots + p_kx_{k-1} = b$. Then $\mu_{\mathcal{L}_1}(n) \leq \mu_{\mathcal{L}_2}(n)$.*

Proof. If $(p_1 + p_2)x_1 + p_3x_2 + \cdots + p_kx_{k-1} = b$ for some $x_i \in [n]$, $1 \leq i \leq k - 1$, then $p_1x_1 + p_2x_1 + p_3x_2 + \cdots + p_kx_{k-1} = b$. Hence any solution to $\mathcal{L}_2$ in $[n]$ gives rise to a solution to $\mathcal{L}_1$ in $[n]$. So if $A \subseteq [n]$ is $\mathcal{L}_1$-free, then A is also $\mathcal{L}_2$-free. Hence the size of the largest $\mathcal{L}_2$-free set is at least the size of the largest $\mathcal{L}_1$-free set. □

We will also make use of the following trivial fact.

Fact 11. *Suppose $\mathcal{L}_1$ is a linear equation and $\mathcal{L}_2$ is a positive integer multiple of $\mathcal{L}_1$. Then the set of $\mathcal{L}_1$-free subsets of $[n]$ is precisely the set of $\mathcal{L}_2$-free subsets of $[n]$. In particular $\mu_{\mathcal{L}_1}(n) = \mu_{\mathcal{L}_2}(n)$, $f(n, \mathcal{L}_1) = f(n, \mathcal{L}_2)$ and $f_{\max}(n, \mathcal{L}_1) = f_{\max}(n, \mathcal{L}_2)$.*

The two results above allow us to extend the use of Lemma 9 to equations with more than three variables.

Lemma 12. *Let $\mathcal{L}$ denote the equation $p_1x_1 + \cdots + p_kx_k = 0$ where $p_i \in \mathbb{Z}$. Suppose there is a partition of the p_i into three non-empty parts P_1, P_2 and P_3 where $p' := \sum_{p_j \in P_1} p_j$, $q' := \sum_{p_j \in P_2} p_j$ and $r' := -\sum_{p_j \in P_3} p_j$ satisfy $p' \geq q' \geq r' \geq 1$. Let $t' := \gcd(p', q', r')$ and write $p := p'/t'$, $q := q'/t'$ and $r := r'/t'$. Let $t := \gcd(p, q)$ and write $r_1 := p/t$ and $r_2 := q/t$. Let S be an $\mathcal{L}$-free subset of $[n]$. If $M \in S$ is divisible by t, then S contains at most*

$$\left\lceil \frac{rM}{q} \right\rceil - 1 - \left\lfloor \frac{rM}{r_2(p+q)} \right\rfloor - (r_1r_2 - r_1 - r_2 + 1) \left\lfloor \left\lfloor \frac{rM}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor \frac{1}{r_1r_2} \right\rfloor$$

elements from $[\lceil rM/q \rceil - 1]$.

Proof. Let $\mathcal{L}'$ denote the equation $px + qy = rz$. Now observe by repeatedly applying Proposition 10 and Fact 11 that any $\mathcal{L}$-free set is also an $\mathcal{L}'$-free set. Hence S must be $\mathcal{L}'$-free and so we simply apply Lemma 9. □

This bounds the number of 'small elements' in solution-free sets for equations with more than three variables, and in Theorem 15 we will use this lemma to obtain a result for the number of maximal solution-free sets.

3. The size of the largest solution-free set

The aim of this section is to use our results from the previous section to obtain bounds on $\mu_{\mathcal{L}}(n)$ for linear equations $\mathcal{L}$ of the form $px + qy = rz$ with $p \geq q \geq r$ positive integers and also linear

equations with more than three variables. As previously mentioned we can use Lemma 9 to obtain a bound on the size of a solution-free set.

Corollary 13. *Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p \geq q \geq r$ and p,q,r are fixed positive integers satisfying $\gcd(p,q,r)=1$. Let S be an $\mathcal{L}$-free subset of $[n]$ and suppose M is the largest element of S divisible by $t := \gcd(p,q)$. Write $r_1 := p/t$ and $r_2 := q/t$. Then*

$$|S| \leq M - \left\lfloor \frac{rM}{r_2(p+q)} \right\rfloor - (r_1r_2 - r_1 - r_2 + 1) \left\lfloor \left\lfloor \frac{rM}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor \frac{1}{r_1r_2} \right\rfloor + \left\lceil \frac{(n-M)(t-1)}{t} \right\rceil.$$

Proof. By Lemma 9, S contains at most $\lceil rM/q \rceil - 1 - \lfloor rM/(r_2(p+q)) \rfloor - (r_1r_2 - r_1 - r_2 + 1)\lfloor \lfloor rM/(r_1(p+q)) - 1/r_2 \rfloor / (r_1r_2) \rfloor$ elements from $[\lceil rM/q \rceil - 1]$. It also cannot contain any element larger than M and divisible by t. □

Note in the statement of Corollary 13 we are implicitly assuming that M exists. If it does not then $|S| \leq \lceil n(t-1)/t \rceil$.

Recall that if $\mathcal{L}$ denotes the equation $px + qy = rz$ where p,q,r are fixed positive integers satisfying $\gcd(p,q,r)=1$ and $t := \gcd(p,q)$, then the set $T := \{x \in [n] : x \not\equiv 0 \bmod t\}$ is $\mathcal{L}$-free. Let a be the unique non-negative integer $0 \leq a < t$ such that $n-a$ is divisible by t. The interval $I := [\lfloor r(n-a)/(p+q) \rfloor + 1, n]$ is also $\mathcal{L}$-free. So $\mu_{\mathcal{L}}(n) \geq \max\{|T|, |I|\}$.

In the following theorem we find equations where the set T is a largest solution-free set; We also find equations where I is a largest solution-free set.

Theorem 1. *Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p \geq q \geq r$ and p,q,r are fixed positive integers satisfying $\gcd(p,q,r)=1$. Let $t := \gcd(p,q)$. Write $r_1 := p/t$ and $r_2 := q/t$.*

(i) If q divides p and $p+q \leq rq$ then $\mu_{\mathcal{L}}(n) = \lceil (q-1)n/q \rceil$;

(ii) If q divides p and $p+q \geq rq$ then $\mu_{\mathcal{L}}(n) = \lceil (p+q-r)(n-a)/(p+q) \rceil + a$ where a is the unique non-negative integer $0 \leq a < q$ such that $n-a$ is divisible by q;

(iii) If q does not divide p, $t > 1$ and

$$r > (r_1r_2 - r_1 - r_2 + 4)r_2\left(r_1 + 1 + \frac{r_2 - 1}{r_1^2 + (r_1-1)(r_2-1)}\right)$$

then $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil$.

Proof. Let S be an $\mathcal{L}$-free subset of $[n]$ and suppose M is the largest element of S divisible by t. If S does not contain an element divisible by t, set $M := 0$. If q divides p then $t = q$ and $r_2 = 1$ and hence by Corollary 13 we have

$$|S| \leq \left\lceil \frac{(p+q-r)M}{p+q} \right\rceil + \left\lceil \frac{(n-M)(q-1)}{q} \right\rceil. \tag{1}$$

(This is true even in the case $M = 0$.)

If $p+q \leq rq$ then $|S| \leq \lceil (q-1)M/q \rceil + \lceil (n-M)(q-1)/q \rceil = \lceil n(q-1)/q \rceil$ since M is divisible by q. Observe that the set $T := \{x \in [n] : x \not\equiv 0 \bmod t\}$ is an $\mathcal{L}$-free set obtaining this size, and so this proves (i).

For (ii) we will show that (1) is an increasing function of M (when restricted to running through M divisible by t) and hence it will be maximised by taking $M = n-a$. Then $|S| \leq \lceil (p+q-r)(n-a)/(p+q) \rceil + a$. Observe that the interval $I := [\lfloor r(n-a)/(p+q) \rfloor + 1, n]$ is an $\mathcal{L}$-free set obtaining this size and so this proves (ii), provided (1) is an increasing function of M.

Since M must be divisible by $t = q$, write $M' := M/q$ and so (1) can be written as

$$\left\lceil \frac{((r_1+1)q - r)M'}{r_1+1} \right\rceil + \left\lceil \frac{n(q-1)}{q} \right\rceil - M'(q-1) = M' + \left\lceil \frac{-rM'}{r_1+1} \right\rceil + \left\lceil \frac{n(q-1)}{q} \right\rceil.$$

Now observe that the difference between successive terms M' and $M'+1$ is given by

$$M'+1+\left\lceil\frac{-r(M'+1)}{r_1+1}\right\rceil-M'-\left\lceil\frac{-rM'}{r_1+1}\right\rceil=1+\left\lceil\frac{-rM'}{r_1+1}-\frac{r}{r_1+1}\right\rceil-\left\lceil\frac{-rM'}{r_1+1}\right\rceil\geq 0$$

where the inequality follows since $r_1+1\geq r$. Hence (1) is an increasing function of M as required.

For (iii) if $M=0$ then $|S|\leq\lceil n(t-1)/t\rceil$ as required. So assume $M\geq t$. Then by Corollary 13 we have

$$\begin{aligned}
|S|&\leq M-\left\lfloor\frac{rM}{r_2(p+q)}\right\rfloor-(r_1r_2-r_1-r_2+1)\left\lfloor\left\lfloor\frac{rM}{r_1(p+q)}-\frac{1}{r_2}\right\rfloor\frac{1}{r_1r_2}\right\rfloor+\left\lceil\frac{(n-M)(t-1)}{t}\right\rceil\\
&\leq M-\frac{rM}{r_2(p+q)}+1-\frac{r_1r_2-r_1-r_2+1}{r_1r_2}\left(\frac{rM}{r_1(p+q)}-\frac{1}{r_2}-1\right)+r_1r_2-r_1-r_2+1\\
&\quad-\frac{M(t-1)}{t}+\left\lceil\frac{n(t-1)}{t}\right\rceil\\
&\leq\left\lceil\frac{n(t-1)}{t}\right\rceil+r_1r_2-r_1-r_2+3-M\left(\frac{r(r_1^2+(r_1-1)(r_2-1))}{tr_1^2r_2(r_1+r_2)}-\frac{1}{t}\right)\\
&=\left\lceil\frac{n(t-1)}{t}\right\rceil+r_1r_2-r_1-r_2+3-M\left(\frac{r}{tr_2}\left(r_1+1+\frac{r_2-1}{r_1^2+(r_1-1)(r_2-1)}\right)^{-1}-\frac{1}{t}\right)\\
&\leq\left\lceil\frac{n(t-1)}{t}\right\rceil+r_1r_2-r_1-r_2+3-M\left(\frac{r_1r_2-r_1-r_2+4}{t}-\frac{1}{t}\right)\\
&\leq\left\lceil\frac{n(t-1)}{t}\right\rceil,
\end{aligned}$$

where the penultimate inequality follows by our lower bound on r and the last inequality follows by using $M\geq t$. □

For Theorem 1(iii) it is easy to check that actually given the conditions on r we must always have $t>1$ (we just state $t>1$ in the theorem for clarity). As an example, $p:=3t$, $q:=2t$, $r\geq 41$, and $t\geq r/2$ gives a set of equations which satisfy the conditions of Theorem 1(iii).

Theorem 1 together with Proposition 10 yield results for $\mu_{\mathcal{L}}(n)$ where $\mathcal{L}$ is an equation with more than three variables. Full details can be found in [13].

Corollary 14. *Let $\mathcal{L}$ denote the equation $a_1x_1+\cdots+a_kx_k+b_1y_1+\cdots+b_\ell y_\ell=c_1z_1+\cdots+c_mz_m$ where the $a_i,b_i,c_i\in\mathbb{N}$ and $p':=\sum_i a_i$, $q':=\sum_i b_i$ and $r':=\sum_i c_i$ satisfy $p'\geq q'\geq r'$. Let $t':=\gcd(p',q',r')$ and write $p:=p'/t'$, $q:=q'/t'$ and $r:=r'/t'$. Let $t:=\gcd(p,q)$.*

(i) If $m=1$, $\ell=1$, $q'=b_1$ divides a_i for all $1\leq i\leq k$ and $p+q\leq rq$ then $\mu_{\mathcal{L}}(n)=\lceil(q-1)n/q\rceil$;

(ii) If q divides p and $p+q\geq rq$ then $\lceil(p+q-r)n/(p+q)\rceil\leq\mu_{\mathcal{L}}(n)\leq\lceil(p+q-r)(n-a)/(p+q)\rceil+a$ where a is the unique non-negative integer $0\leq a<q$ such that $n-a$ is divisible by q;

(iii) Write $r_1:=p/t$ and $r_2:=q/t$. If q does not divide p, $m=1$, tt' divides a_i and b_j for $1\leq i\leq k$, $1\leq j\leq\ell$ and

$$r>(r_1r_2-r_1-r_2+4)r_2\left(r_1+1+\frac{r_2-1}{r_1^2+(r_1-1)(r_2-1)}\right)$$

then $\mu_{\mathcal{L}}(n)=\lceil(t-1)n/t\rceil$.

□

4. The number of maximal solution-free sets

We start this section with the proof of Theorem 2.

Theorem 2. *Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p \geq q \geq r$ and p,q,r are fixed positive integers satisfying* $\gcd(p,q,r)=1$. *Let $t := \gcd(p,q)$. Then*

$$f_{\max}(n,\mathcal{L}) \leq 2^{\frac{Crn}{q}+o(n)} \text{ where } C := 1 - \frac{t}{p+q}\Big(\frac{p^2+(p-t)(q-t)}{p^2}\Big).$$

Proof. First note that C lies between $1/2$ and $1-t/(p+q)$. To see this, note that if q divides p, then $C = 1-q/(p+q) \geq 1/2$ since $p \geq q$. Otherwise, $p > q > t$, and so $(p-t)(q-t) < p^2$. Hence $t(p^2+(p-t)(q-t))/(p^2(p+q)) < 2t/(p+q) \leq 2(q/2)/(p+q) < 1/2$ and so $C > 1/2$. We observe that $C \leq 1-t/(p+q)$ since $p \geq q \geq t$.

Let $\mathcal{F}$ denote the set of containers obtained by applying Lemma 4. Since every maximal $\mathcal{L}$-free subset of $[n]$ lies in at least one of the $2^{o(n)}$ containers, it suffices to show that every $F \in \mathcal{F}$ houses at most $2^{Crn/q+o(n)}$ maximal $\mathcal{L}$-free sets.

Let $F \in \mathcal{F}$. By Lemmas 4(i) and 5, $F = A \cup B$ where $|A| = o(n)$, $|B| \leq \mu_{\mathcal{L}}(n)$ and B is $\mathcal{L}$-free. Define $M := \max B$, $M' := \max\{x \in B : x \equiv 0 \bmod t\}$ and $u := \max\{\lfloor rM'/q \rfloor, \lfloor rn/2q \rfloor\}$. Every maximal $\mathcal{L}$-free set which lies in such a container can be constructed by:

(i) Picking $S_1 \subseteq A$ to be $\mathcal{L}$-free;
(ii) Adding a set $S_2 \subseteq [u] \cap B$ so that $S_1 \cup S_2$ is $\mathcal{L}$-free;
(iii) Extending to a maximal $\mathcal{L}$-free set in $[u+1, M] \cap B$.

There are $2^{o(n)}$ ways to pick S_1. If $M' \leq n/2$ then $u = \lfloor rn/2q \rfloor$ and so there are at most $2^{rn/2q} \leq 2^{Crn/q}$ ways to pick S_2 so that $S_1 \cup S_2$ is $\mathcal{L}$-free. Write $r_1 := p/t$ and $r_2 := q/t$. If $M' \geq n/2$ then since M' is divisible by t, we apply Lemma 9 to show that

$$\begin{aligned}
|[u] \cap B| &= |[\lfloor rM'/q \rfloor] \cap B| \\
&\leq \left\lfloor \frac{rM'}{q} \right\rfloor - \left\lfloor \frac{rM'}{r_2(p+q)} \right\rfloor - (r_1r_2 - r_1 - r_2 + 1)\left\lfloor \left\lfloor \frac{rM'}{r_1(p+q)} - \frac{1}{r_2} \right\rfloor \frac{1}{r_1r_2} \right\rfloor \\
&= \frac{CrM'}{q} + o(n).
\end{aligned}$$

Hence there are at most $2^{CrM'/q+o(n)} \leq 2^{Crn/q+o(n)}$ ways to pick S_2 so that $S_1 \cup S_2$ is $\mathcal{L}$-free.

Let $B' := [u+1, M] \cap B$. For step (iii) we calculate the number of extensions of $S_1 \cup S_2$ into B'. Observe by Lemma 6, this is bounded above by $\mathrm{MIS}(L_{S_1 \cup S_2}[B'])$. We will show that this link graph has only one maximal independent set. Then combining steps (i)-(iii) we have that F contains at most $2^{o(n)} \times 2^{Crn/q+o(n)} = 2^{Crn/q+o(n)}$ maximal $\mathcal{L}$-free sets as desired.

If the link graph only contains loops and isolated vertices, then it has only one maximal independent set. For it to have an edge between distinct vertices, we either must have $x, z \in B'$, $y \in S_1 \cup S_2$ such that $px+qy=rz$ or $py+qx=rz$, or we must have $x, y \in B'$, $z \in S_1 \cup S_2$ such that $px+qy=rz$.

The first of these events does not occur since otherwise $rz \geq q(x+y) > qx \geq q(\lfloor rM'/q \rfloor + 1) > rM'$ and so $z > M'$. Note that since z is part of the solution $px+qy=rz$ and $\gcd(p,q,r)=1$, it must be divisible by t. However this contradicts $z > M'$ as we have $z \in B$, but M' was defined to be the largest element in B divisible by t.

If $M' > n/2$ then the second event does not occur since $rz = px+qy \geq q(x+y) \geq 2q(\lfloor rM'/q \rfloor + 1) > 2rM' > rn$ and so $z > n$. If $M' \leq n/2$ then the second event does not occur since $rz = px+qy \geq q(x+y) \geq 2q(\lfloor rn/2q \rfloor + 1) > rn$ and so again $z > n$. □

Note that when $r = 1$, Theorem 2 gives us new results for equations of the form $px+qy = z$. The authors previously obtained results for such equations in [14]. In the appendix we give a summary describing which result gives the best upper bound for various values of p and q.

When $\mathcal{L}$ denotes the equation $qx + qy = z$ for some positive integer $q \geq 2$, Proposition 26(iii) from [14] gives a lower bound of $f_{\max}(n, \mathcal{L}) \geq 2^{(n-6q)/2q}$. Combining this with Theorem 2 gives us the following asymptotically exact result.

Theorem 3. *Let $\mathcal{L}$ denote the equation $qx + qy = z$ where $q \geq 2$ is an integer. Then*

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

By adapting the proof of Theorem 2 we obtain the following result for $f_{\max}(n, \mathcal{L})$ for some equations with more than three variables.

Theorem 15. *Let $\mathcal{L}$ denote the equation $p_1x_1 + \cdots + p_kx_k = rz$ where $p_1, \ldots, p_k, r \in \mathbb{N}$ satisfy $\gcd(p_1, \ldots, p_k, r) = 1$ and $p_1 \geq \cdots \geq p_k \geq r$. Suppose that $p := \sum_{i=1}^{k-1} p_i$ and $q := p_k$ satisfy $t := \gcd(p, q) = \gcd(p_1, \ldots, p_k)$. Then*

$$f_{\max}(n, \mathcal{L}) \leq 2^{\frac{Crn}{q}+o(n)} \text{ where } C := 1 - \frac{t}{p+q}\Big(\frac{p^2 + (p-t)(q-t)}{p^2}\Big).$$

Proof. We follow the proof used in Theorem 2 precisely (except for using Lemma 12 instead of Lemma 9 in step (ii)) up until counting the number of ways of extending $S_1 \cup S_2$ to a maximal $\mathcal{L}$-free set in $B' := [u+1, M] \cap B$. Observe by Lemma 6, this is bounded above by $\mathrm{MIS}(L_{S_1 \cup S_2}[B'])$ since B' and $S_1 \cup S_2$ are $\mathcal{L}$-free. To see that B' is $\mathcal{L}$-free, suppose $(x_1, \ldots, x_k, z)$ is a solution within B' and note that $rz = p_1x_1 + \cdots + p_kx_k > p_kx_k \geq q(rM'/q) = rM'$ and so $z > M'$. (Here we needed that each p_i is positive.) Since $\gcd(p_1, \ldots, p_k, r) = 1$ and $\gcd(p, q) = \gcd(p_1, \ldots, p_k)$ we have $\gcd(t, r) = 1$ and so in any solution to $\mathcal{L}$, z must be divisible by t. However this contradicts $z > M'$ as we have $z \in B$, but M' was defined to be the largest element in B divisible by t.

We will show that this link hypergraph $L_{S_1 \cup S_2}[B']$ has only one maximal independent set (and hence the number of maximal $\mathcal{L}$-free sets contained in F is at most $2^{Crn/q+o(n)}$ as required). If the link hypergraph only contains loops and isolated vertices, then it has only one maximal independent set.

For it to have a hyperedge between at least two vertices, there must exist a solution $(x_1, \ldots, x_k, z)$ where either there is a hyperedge with distinct vertices $x_i, z \in B'$ for some $1 \leq i \leq k$ and $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k\} \in B' \cup S_1 \cup S_2$, or there is a hyperedge with distinct vertices $x_i, x_j \in B'$ for some $1 \leq i < j \leq k$ and $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k, z\} \in B' \cup S_1 \cup S_2$.

Suppose the first event occurs with $(x_1, \ldots, x_k, z)$. Then $rz = p_1x_1 + \cdots + p_kx_k > p_ix_i \geq p_kx_i = qx_i \geq q(rM'/q) = rM'$ and so $z > M'$. But since z is part of a solution, it must be divisible by t. This contradicts $z \in B$, since M' was defined to be the largest element in B divisible by t.

If $M' > n/2$ then the second event does not occur since $rz = p_1x_1 + \cdots + p_kx_k > p_k(x_i + x_j) \geq 2q(\lfloor rM'/q \rfloor + 1) > 2rM' > rn$ and so $z > n$. If $M' \leq n/2$ then the second event does not occur since $rz = p_1x_1 + \cdots + p_kx_k > p_k(x_i + x_j) \geq 2q(\lfloor rn/(2q) \rfloor + 1) > rn$ and so again $z > n$. □

We end the section with a lower bound.

Proposition 16. *Let $\mathcal{L}$ denote the equation $qx + qy = rz$ where $q > r$ and q, r are fixed positive integers satisfying $\gcd(q, r) = 1$. Then*

$$f_{\max}(n, \mathcal{L}) \geq 2^{\lceil \lfloor rn/2q - rq/2 \rfloor (q-1)/q \rceil - 1}.$$

Proof. Let B be the $\mathcal{L}$-free set $\{z \in [n] : z \not\equiv 0 \bmod q\}$. Let $M := \max\{z \in [n] : rz/q^2 \in [n]\}$; so $M > n - q^2$. Let $S := \{M\}$ and consider the link graph $L_S[B]$. Note that if $i \in B$ where $i < rM/q$ then $rM/q - i \in B$. This follows since $rM/q^2 \in \mathbb{N}$ and so $rM/q - i \not\equiv 0 \bmod q$. Hence there is an edge in $L_S[B]$ between every such i and $rM/q - i$ since $q(i + rM/q - i) = rM$. By running through

all $i \in B$ we obtain a total of $\lceil \lfloor rM/2q \rfloor (q-1)/q \rceil$ disjoint edges in $L_S[B]$ of which at most one is a loop (at $rM/2q$ if it is an integer not congruent to 0 modulo q). Hence we obtain an *induced* matching E in $L_S[B]$ of size $\lceil \lfloor rM/2q \rfloor (q-1)/q \rceil - 1 \geq \lceil \lfloor rn/2q - rq/2 \rfloor (q-1)/q \rceil - 1$. It is easy to see that this gives rise to $2^{|E|}$ maximal independent sets in $L_S[B]$. By applying Lemma 7 we obtain the result. □

Question 17. *Let $\mathcal{L}$ denote the equation $qx + qy = rz$ where $q > r \geq 2$ and q, r are fixed positive integers satisfying* $\gcd(q, r) = 1$. *Does $f_{\max}(n, \mathcal{L}) = 2^{rn(q-1)/(2q^2)+o(n)} (= 2^{(\mu_{\mathcal{L}}(n) - \mu^*_{\mathcal{L}}(n))/2 + o(n)})$?*

5. Concluding Remarks

In this paper we have used Lemma 8 as a tool to prove both results on the size of the largest solution-free subset of $[n]$ as well as on the number of maximal solution-free subsets of $[n]$. Recall that Green [12] showed that given a homogenous linear equation $\mathcal{L}$, then $f(n, \mathcal{L}) = 2^{\mu_{\mathcal{L}}(n)+o(n)}$. One can actually very easily apply Lemma 8 to obtain that $f(n, \mathcal{L}) = \Theta(2^{\mu_{\mathcal{L}}(n))})$ for *some* linear equations $\mathcal{L}$ of the form $px + qy = rz$ where $p \geq q \geq r$ are positive integers. However, the results we obtain seem quite niche. Thus, we defer their statement and proof to the thesis of the first author [13].

The crucial trick used in the proof of Theorems 2 and 15 was to choose our sets S carefully so that the link hypergraphs $L_S[B]$ each contain precisely one maximal independent set. In other applications of this method [2, 3, 14] the approach had been instead to obtain other structural properties of the link graphs (such as being triangle-free) to ensure there are not too many maximal independent sets in $L_S[B]$. It would be interesting to see if the approach of our paper can be applied to obtain other (exact) results in the area.

Although we have found an initial bound on $f_{\max}(n, \mathcal{L})$ for some equations with more than three variables, we still do not know in general if there are significantly fewer maximal $\mathcal{L}$-free subsets of $[n]$ than there are $\mathcal{L}$-free subsets of $[n]$. Progress on giving general upper bounds on the number of maximal independent sets in (non-uniform) hypergraphs should (through the method of link hypergraphs) yield results in this direction.

References

[1] J. Balogh, H. Liu, S. Petříčková and M. Sharifzadeh, The typical structure of maximal triangle-free graphs, *Forum Math. Sigma*, 3, (2015), e20 (19 pages).

[2] 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.

[3] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, Sharp bound on the number of maximal sum-free subsets of integers, submitted.

[4] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, *J. Amer. Math. Soc.*, 28, (2015), 669–709.

[5] J. Balogh and S. Petříčková, The number of the maximal triangle-free graphs, *Bull. London Math. Soc.*, 46, (2014), 1003–1006.

[6] 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.

[7] 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.

[8] P. Cameron and P. Erdős, Notes on sum-free and related sets, *Combin. Probab. Comput.*, 8, (1999), 95–107.

[9] K. Dilcher and L. Lucht, On finite pattern-free sets of integers, *Acta Arith.*, 121, (2006), 313–325.

[10] 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.

[11] B. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.*, 36, (2004), 769–778.

[12] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.*, 15, (2005), 340–376.

[13] R. Hancock, *PhD thesis, University of Birmingham*, in preparation.

[14] R. Hancock and A. Treglown, On solution-free sets of integers, submitted.
[15] P. Hegarty, Extremal subsets of $\{1,\dots,n\}$ avoiding solutions to linear equations in three variables, *Electron. J. Combin.*, 14, (2007), R74.
[16] D. Kleitman and K.J. Winston, The asymptotic number of lattices, *Ann. Discrete Math.*, 6, (1980), 243–249, Combinatorial mathematics, optimal designs and their applications (Proc. Sympos. Combin. Math. and Optimal Design, Colorado State Univ., Fort Collins, Colo., 1978)
[17] D. Kleitman and K.J. Winston, On the number of graphs without 4-cycles, *Discrete Math.*, 6, (1982), 167–172.
[18] D. Král', O. Serra and L. Vena, A removal lemma for systems of linear equations over finite fields, *Israel J. Math.*, 187, (2012), 193–207
[19] T. Łuczak and T. Schoen, On the number of maximal sum-free sets, *Proc. Amer. Math. Soc.*, 129, (2001), 2205–2207.
[20] K.F. Roth, On certain sets of integers, *J. London Math. Soc.*, 28, (1953), 104–109.
[21] I.Z. Ruzsa, Solving a linear equation in a set of integers I, *Acta Arith.*, 65, (1993), 259–282.
[22] I.Z. Ruzsa, Solving a linear equation in a set of integers II, *Acta Arith.*, 72, (1995), 385–397.
[23] A.A. Sapozhenko, The Cameron-Erdős conjecture, (Russian) *Dokl. Akad. Nauk.*, 393, (2003), 749–752.
[24] D. Saxton and A. Thomason, Hypergraph containers, *Invent. Math.*, 201, (2015), 925–992.
[25] I. Schur, Uber die Kongruenz $x^m + y^m \equiv z^m \pmod p$, *ber. Deutsch. Mat. Verein.*, 25, (1916), 114–117.
[26] G. Wolfovitz, Bounds on the number of maximal sum-free sets, *European J. Combin.*, 30, (2009), 1718–1723.

Robert Hancock, Andrew Treglown
School of Mathematics
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK

E-mail addresses: {rah410,a.c.treglown}@bham.ac.uk

Appendix A

In this appendix we give a summary of the best known upper bound on $f_{\max}(n,\mathcal{L})$ for equations of the form $px+qy=rz$ where $p\geq q\geq r$. First we recall some of the results from [14].

Theorem 18. [14] *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 19. [14] *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*

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

Theorem 20. [14] *Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p\geq q\geq r$ and $p,q,r\in\mathbb{N}$. Then*

$$f_{\max}(n,\mathcal{L}) \leq 2^{\mu_{\mathcal{L}}(rn/q)+o(n)}.$$

Proposition 21. [14] *Let $\mathcal{L}$ denote the equation $px+qy=z$ where $p\geq q$, $p\geq 2$ and $p,q\in\mathbb{N}$. The best upper bound on $f_{\max}(n,\mathcal{L})$ given by Theorems 18, 19 and 20 is:*

(i) $f_{\max}(n,\mathcal{L}) \leq 3^{(\mu_{\mathcal{L}}(n)-\mu^*_{\mathcal{L}}(n))/3+o(n)}$ *if* $\gcd(p,q)=q$, $p\geq q^2$, *and either* $q\leq 9$ *or* $10\leq q\leq 17$ *and* $p<(a-1)(q^2-q)/(q(2-a)-1)$ *where* $a:=\log_3(8)$;
(ii) $f_{\max}(n,\mathcal{L}) \leq 2^{(\mu_{\mathcal{L}}(n)-\mu^*_{\mathcal{L}}(n))/2+o(n)}$ *if* $\gcd(p,q)=q$ *and* $p\leq q^2-q$;
(iii) $f_{\max}(n,\mathcal{L}) \leq 2^{\mu_{\mathcal{L}}(n/q)+o(n)}$ *otherwise.*

Recall $\mu^*_{\mathcal{L}}(n)$ denotes the number of elements in $[n]$ which do not lie in *any* non-trivial solution to $\mathcal{L}$ that only consists of elements from $[n]$. Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p\geq q\geq r$ and p,q,r are fixed positive integers satisfying $\gcd(p,q,r)=1$. Let $t:=\gcd(p,q)$. Then notice that $S:=\{s\in[n]: s>\lfloor(rn-p)/q\rfloor, t\nmid s\}$ is a set of elements which do not lie in any solution to $\mathcal{L}$ in $[n]$. This follows since if $s>\lfloor(rn-p)/q\rfloor$ then $ps+q\geq qs+p>rn$ and so s cannot play the role of x or

y in an $\mathcal{L}$-triple in $[n]$. If $t \nmid s$ then as $t|(px+qy)$ for any $x, y \in [n]$ but $\gcd(r,t)=1$ we have that s cannot play the role of z in an $\mathcal{L}$-triple in $[n]$. Actually, for large enough n, every element that does not lie in any solution to $\mathcal{L}$ in $[n]$ is in S, and so we have $\mu^*_{\mathcal{L}}(n) = |S| = \lceil (n - \lfloor (rn-p)/q \rfloor)(t-1)/t \rceil$ for all such $\mathcal{L}$. We omit the proof of this here.

We can now state the new summary.

Proposition 22. *Let $\mathcal{L}$ denote the equation $px+qy=rz$ where $p \geq q \geq r$, $p \geq 2$ and p,q,r are fixed positive integers satisfying $\gcd(p,q,r)=1$. Let $t := \gcd(p,q)$ and let $a := \log_2 3$. The best upper bound on $f_{\max}(n,\mathcal{L})$ given by Theorems 18, 19, 20 and 2 is:*

(i) $f_{\max}(n,\mathcal{L}) \leq 3^{(\mu_{\mathcal{L}}(n)-\mu^*_{\mathcal{L}}(n))/3+o(n)}$ if
- *(a) $r=1$, $\gcd(p,q)=q$, $p \geq \max\{q^2, (q^2-q)a/(q(3-2a)+a)\}$, and $q \leq 9$;*
- *(b) $r \geq 2$, $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil$, and additionally (1) $p \neq q$ or (2) $2 \leq q \leq 18$;*
- *(c) $r \geq 2$, q divides p, $p+q \geq rq$ and additionally (1) $p \neq q$ or (2) $2 \leq q \leq 18$;*

(ii) $f_{\max}(n,\mathcal{L}) \leq 2^{Crn/q+o(n)}$ where $C := 1 - t(p^2+(p-t)(q-t))/(p^2(p+q))$ if
- *(a) $r=1$, $\gcd(p,q) \neq q$ or $q > 9$ or $p < q^2$ or $p < (q^2-q)a/(q(3-2a)+a)$;*
- *(b) $r \geq 2$, $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil$, and $p = q \geq 19$.*

Proof. First suppose that $r=1$ (and so $\mu_{\mathcal{L}}(n) = \lceil (p+q-1)n/(p+q) \rceil$). Note that $C \leq 1-t/(p+q) = (p+q-t)/(p+q) \leq (p+q-1)/(p+q)$ and so the exponent given by Theorem 2 is at most the exponent given by Theorem 20. For Theorem 19 we require $\gcd(p,q)=q$ and $p \leq q^2-q$. In this case $C = p/(p+q)$ and $(p+q-1)/(2(p+q)) - (q-1)^2/(2q^2) = (2pq+q^2-p-q)/(2q^2(p+q)) \geq p/(q(p+q)) = C/q$ and so the exponent given by Theorem 2 is at most the exponent given by Theorem 19.

It remains to check when the bound given by Theorem 18 is still better than the bound given by Theorem 2. By Proposition 21 this can only possibly be the case if $\gcd(p,q)=q$ and $p \geq q^2$. To prove (i)(a) it suffices to show that

$$3^{\frac{(p+q-1)n}{3(p+q)} - \frac{(q-1)^2 n}{3q^2}} \leq 2^{\frac{pn}{q(p+q)}},$$

or rearranging

$$p(q(3-2a)+a) \geq (q^2-q)a.$$

If $q \geq 10$ then $(q(3-2a)+a)$ is negative, but then we would require p negative, a contradiction. Hence we must have $q \leq 9$ and then the inequality holds if $p > (q^2-q)a/(q(3-2a)+a)$.

Now suppose that $r \geq 2$ and $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil$. Then $\mu_{\mathcal{L}}(n) - \mu^*_{\mathcal{L}}(n) = r(t-1)n/(qt) + o(n)$ and $3^{x/3} < 2^x$ and so Theorem 18 gives a better bound than Theorem 20. We wish to know when

$$3^{\frac{r}{q}\frac{t-1}{3t}} < 2^{\frac{r}{q}(1-\frac{t}{p+q}(\frac{p^2+(p-t)(q-t)}{p^2}))}.$$

Write $r_1 := p/t$ and $r_2 := q/t$. The above rearranges to give

$$t((a-3)r_1^2(r_1+r_2) + 3r_1^2 + 3(r_1-1)(r_2-1)) < ar_1^2(r_1+r_2).$$

The right hand side is positive and the left hand side is negative unless $r_1 = r_2 = 1$. In this case $p=q=t$ and so we now require $3^{(q-1)/(3q)} < 2^{1/2}$, which holds when $q \leq 18$.

Finally suppose that $r \geq 2$, q divides p and $p+q \geq rq$ (so $\mu_{\mathcal{L}}(n) = \lceil (p+q-r)n/(p+q) \rceil$). Since q divides p, we have $t=q$ and $p=r_1 q$, and so Theorem 2 gives a bound of $2^{rpn/(q(p+q))+o(n)}$. This is better than Theorem 20 which gives a bound of $2^{r(p+q-r)n/(q(p+q))+o(n)}$ since $q \geq r$. Therefore we wish to know when

$$3^{\frac{p+q-r}{3(p+q)} - \frac{(q-r)(q-1)}{3q^2}} < 2^{\frac{rp}{q(p+q)}}.$$

Rearranging, we require $r_1(a(q+qr-r)/3 - qr) \leq a(r-q)/3$. Now note $a(q+qr-r)/3 - qr$ is negative when $r \geq 2$, so this rearranges to give $r_1 \geq (q-r)/(r-q-rq+3rq/a)$. If $p > q$ (so $r_1 \geq 2$), it suffices to have $2 \geq (q-r)/(r-q-rq+3rq/a)$ or rearranging, $q(r(2-6/a)+3) \leq 3r$.

This holds since $r(2-6/a)+3$ is negative for $r \geq 2$. Otherwise $p = q$, and so since $p + q \geq rq$, we have that $r = 2$. So we require $1 \geq (q-2)/((6/a-3)q+2)$ which holds when $q \leq 18$. (In this final case, $\mu_{\mathcal{L}}(n) = \lceil (t-1)n/t \rceil = \lceil (p+q-r)n/(p+q) \rceil = \lceil (q-1)n/q \rceil$.)

□