Discrete Applied Mathematics 243 (2018) 219–238

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On the complexity of finding and counting solution-free sets of integers

Kitty Meeks [a], Andrew Treglown [b,*]

[a] School of Computing Science, Sir Alwyn Williams Building, University of Glasgow, Glasgow G12 8QQ, UK
[b] School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

ARTICLE INFO

Article history:
Received 16 October 2017
Accepted 1 February 2018
Available online 26 March 2018

Keywords:
(Parameterised) complexity
Solution-free sets

ABSTRACT

Given a linear equation $\mathcal{L}$, a set A of integers is $\mathcal{L}$-free if A does not contain any 'non-trivial' solutions to $\mathcal{L}$. This notion incorporates many central topics in combinatorial number theory such as sum-free and progression-free sets. In this paper we initiate the study of (parameterised) complexity questions involving $\mathcal{L}$-free sets of integers. The main questions we consider involve deciding whether a finite set of integers A has an $\mathcal{L}$-free subset of a given size, and counting all such $\mathcal{L}$-free subsets. We also raise a number of open problems.

1. Introduction

Sets of integers which do not contain any solutions to some linear equation have received a lot of attention in the field of combinatorial number theory. Two particularly well-studied examples are *sum-free sets* (sets avoiding solutions to the equation $x + y = z$) and *progression-free sets* (sets that do not contain any 3-term arithmetic progression x, y, z or equivalently avoid solutions to the equation $x + z = 2y$). A lot of effort has gone into determining the size of the largest solution-free subset of $\{1, \ldots, n\}$ and other sets of integers, and into computing (asymptotically) the number of (maximal) solution-free subsets of $\{1, \ldots, n\}$.

In this paper we initiate the study of the computational complexity of problems involving solution-free subsets. We are primarily concerned with determining the size of the largest subset of an arbitrary set of integers A which avoids solutions to a specified linear equation $\mathcal{L}$; in particular, we focus on sum-free and progression-free sets, but many of our results also generalise to larger families of linear equations. For suitable equations $\mathcal{L}$, we demonstrate that the problem of deciding whether A contains a solution-free subset of size at least k is NP-complete (see Section 2); we further show that it is hard to approximate the size of the largest solution-free subset within a factor $(1 + \epsilon)$ (see Section 3), or to determine for a constant $c < 1$ whether A contains a solution-free subset of size at least $c|A|$ (see Section 6). On the other hand, in Section 5 we see that the decision problem is fixed-parameter tractable when parameterised by either the cardinality of the desired solution-free set, or by the number of elements of A we can exclude from such a set. We also consider the complexity, with respect to various parameterisations, of counting the number of solution-free sets of a specified size (see Section 7): while there is clearly no polynomial-time algorithm in general, the problem is fixed-parameter tractable when parameterised by the number of elements we can exclude from A; we show that there is unlikely to be a fixed-parameter algorithm to solve the counting problem exactly when the size of the solution-free sets is taken as the parameter, but we give an efficient approximation algorithm for this setting. Finally, in Section 8 we consider all of these questions in a variant of the problem, where we specify that a given solution-free subset $B \subset A$ must be included in any solution.

* Corresponding author.
E-mail addresses: kitty.meeks@glasgow.ac.uk (K. Meeks), a.c.treglown@bham.ac.uk (A. Treglown).

https://doi.org/10.1016/j.dam.2018.02.008

Many of our results are based on the fact that we can set up polynomial-time reductions in both directions between our problem and (different versions of) the well-known hitting set problem for hypergraphs. In particular, in Section 2.1 we provide a construction that has several applications throughout the paper. In Section 4 we also derive some new lower-bounds on the size of the largest solution-free subset of an arbitrary set of integers for certain equations $\mathcal{L}$, which may be of independent interest. Our approach here utilises a trick of Alon and Kleitman [3] which transfers the problem into the setting of solution-free sets in cyclic groups.

Our aim is to provide a thorough introduction to the study of (parameterised) complexity questions involving $\mathcal{L}$-free sets of integers. As such, some of the results presented have straightforward proofs, such as the parameterised complexity results discussed in Section 5, whilst other proofs are more involved. However, even the simplest of our results lead to natural open questions. In Section 9 we collect together a number of open problems which we hope will stimulate further interest in the topic.

In the remainder of this section, we give some background on solution-free sets in Section 1.1 and review the relevant notions from the study of computational complexity in Section 1.2. In Section 1.3 we outline the main results of the paper.

1.1. Background on solution-free sets

Consider a fixed linear equation $\mathcal{L}$ of the form

$$a_1x_1 + \cdots + a_\ell x_\ell = b \tag{1}$$

where $a_1, \ldots, a_\ell, b \in \mathbb{Z}$. We say that $\mathcal{L}$ is *homogeneous* if $b = 0$. If

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

then we say that $\mathcal{L}$ is *translation-invariant*. (Here $[k]$ denotes the set $\{1, \ldots, k\}$.) Let $\mathcal{L}$ be translation-invariant. Then notice that $(x, \ldots, x)$ is a 'trivial' solution of (1) for any x. More generally, a solution $(x_1, \ldots, x_k)$ to $\mathcal{L}$ is said to be *trivial* if there exists a partition $P_1, \ldots, P_\ell$ 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 of integers 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*.

1.1.1. Sum-free sets

A set S (of integers or elements of a group) is *sum-free* if there does not exist x, y, z in S such that $x + y = z$. The topic of sum-free sets has a rich history spanning a number of branches of mathematics. In 1916 Schur [42] proved that, given $r \in \mathbb{N}$, if n is sufficiently large, then any r-colouring of $[n] := \{1, \ldots, n\}$ yields a monochromatic triple x, y, z such that $x + y = z$. (Equivalently, $[n]$ cannot be partitioned into r sum-free sets.) This theorem was followed by other seminal related results such as van der Waerden's theorem [46], and ultimately led to the birth of arithmetic Ramsey theory.

Paul Erdős had a particular affinity towards sum-free sets. In 1965 he [20] proved one of the cornerstone results in the subject: every set of n non-zero integers A contains a sum-free subset of size at least $n/3$. Employing the probabilistic method, Alon and Kleitman [3] improved this bound to $(n+1)/3$ and further, Kolountzakis [33] gave a polynomial time algorithm for constructing such a sum-free subset. Then, using a Fourier-analytical approach, Bourgain [11] further improved the bound to $(n+2)/3$ in the case when A consists of positive integers. Erdős [20] also raised the question of determining upper bounds for this problem: recently Eberhard, Green and Manners [18] asymptotically resolved this important classical problem by proving that there is a set of positive integers A of size n such that A does not contain any sum-free subset of size greater than $n/3 + o(n)$. This result raises the question of whether one can decide *efficiently* whether a set A of non-negative integers contains a sum-free subset of size at least $c|A|$ for some $c > 1/3$. As we shall see in Section 6, the answer is likely to be no.

In Section 7 we consider the complexity, with respect to various parameterisations, of counting the number of sum-free sets of a specified size. A number of important questions concerning (counting) sum-free sets were raised in two papers of Cameron and Erdős [13,14]. In [13], Cameron and Erdős conjectured that there are $\Theta(2^{n/2})$ sum-free subsets of $[n]$. Here, the lower bound follows by observing that the largest sum-free subset of $[n]$ has size $\lceil n/2 \rceil$; this is attained by the set of odds in $[n]$ and by $\{\lfloor n/2 \rfloor + 1, \ldots, n\}$. Then, for example, by taking all subsets of $[n]$ containing only odd numbers one obtains at least $2^{n/2}$ sum-free subsets of $[n]$. After receiving much attention, the Cameron–Erdős conjecture was proven independently by Green [28] and Sapozhenko [40]. Given a set A of integers we say $S \subseteq A$ is a *maximal sum-free subset of* A if S is sum-free and it is not properly contained in another sum-free subset of A. Cameron and Erdős [14] raised the question of how many maximal sum-free subsets there are in $[n]$. Very recently, this question has been resolved via a combinatorial approach by Balogh, Liu, Sharifzadeh and Treglown [7,8].

Sum-free sets have also received significant attention with respect to groups. One highlight in this direction is work of Diananda and Yap [15] and Green and Ruzsa [29] that determines the size of the largest sum-free subset for every finite abelian group. In each case the largest sum-free set has size linear in the size of the abelian group. Another striking result in the area follows from Gowers' work on quasirandom groups. Indeed, Gowers [27] proved that there are non-abelian groups for which the largest sum-free subset has sublinear size, thereby answering a question of Babai and Sós [5]. See the survey of Tao and Vu [44] for a discussion on further problems concerning sum-free sets in groups.