Lower bounds for the Turán densities of daisies

David Ellis[a]   Dylan King[b]

Abstract

For integers $r \geqslant 2$ and $t \geqslant 2$, an r-uniform t-daisy $\mathcal{D}_r^t$ is a family of $\binom{2t}{t}$ r-element sets of the form

$$\{S \cup T \ : T \subset U, \ |T| = t\}$$

for some sets S, U with $|S| = r - t$, $|U| = 2t$ and $S \cap U = \emptyset$. It was conjectured by Bollobás, Leader and Malvenuto (and independently by Bukh) that the Turán densities of t-daisies satisfy $\lim_{r \to \infty} \pi(\mathcal{D}_r^t) = 0$ for all $t \geqslant 2$ (an equivalent conjecture was made independently by Johnson and Talbot). This has become a well-known problem, and it is still open for all values of t. In this paper, we give lower bounds for the Turán densities of r-uniform t-daisies. To do so, we introduce (and make some progress on) the following natural problem in additive combinatorics: for integers $m \geqslant 2t \geqslant 4$, what is the maximum cardinality $g(m, t)$ of a subset R of $\mathbb{Z}/m\mathbb{Z}$ such that for any $x \in \mathbb{Z}/m\mathbb{Z}$ and any $2t$-element subset X of $\mathbb{Z}/m\mathbb{Z}$, there are t distinct elements of X whose sum is not in the translate $x + R$? This is a slice-analogue of an extremal Hilbert cube problem considered by Gunderson and Rödl as well as Cilleruelo and Tesoro.

Mathematics Subject Classifications: 05C65, 05D40

1   Introduction

For integers $r \geqslant 2$ and $t \geqslant 2$, an r-uniform t-daisy $\mathcal{D}_r^t$ is a collection of $\binom{2t}{t}$ r-element sets of the form

$$\{S \cup T \ : T \subset U, \ |T| = t\}$$

for some sets S, U with $|S| = r - t$, $|U| = 2t$ and $S \cap U = \emptyset$. As usual, for an integer $n \geqslant 3$ we write $\mathrm{ex}(n, \mathcal{D}_r^t)$ for the nth *Turán number* of $\mathcal{D}_r^t$, i.e. the maximum possible cardinality of a family of r-element subsets of $\{1, 2, \ldots, n\}$ which is $\mathcal{D}_r^t$-free, and we write

$$\pi(\mathcal{D}_r^t) = \lim_{n \to \infty} \frac{\mathrm{ex}(n, \mathcal{D}_r^t)}{\binom{n}{r}}$$

[a]School of Mathematics, University of Bristol, UK (david.ellis@bristol.ac.uk).

[b]Department of Mathematics, California Institute of Technology, Pasadena, California, USA (dking@caltech.edu).

for the *Turán density* of $\mathcal{D}_r^t$.

In the case $r = 2$, we have $\pi(\mathcal{D}_2^2) = 2/3$ by Turán's theorem for K_4's. The first unknown case occurs when $r = 3$; in this case, Bollobás, Leader and Malvenuto show in [2] that $\pi(\mathcal{D}_3^2) \geqslant \frac{1}{2}$, by taking the complement of the Fano plane, blowing up and iterating, and conjecture that in fact equality holds. In Proposition 4.3 of [5], Falgas-Ravry and Vaughan use the semidefinite programming approach developed by Razborov to show that $\pi(\mathcal{D}_3^2) \leqslant 0.504081$. For larger $t \geqslant 2$ and $r \geqslant 3$, even less is known concerning $\pi(\mathcal{D}_r^t)$. The following conjecture was made by Bollobás, Leader and Malvenuto in [2] (and independently by Bukh, see [2]).

Conjecture 1 (Bollobás-Leader-Malvenuto / Bukh)**.** For all $t \geqslant 2$, $\lim\limits_{r \to \infty} \pi(\mathcal{D}_r^t) = 0$.

This is still open even for $t = 2$. Johnson and Talbot independently made an equivalent conjecture in [8], which we now describe. To state it, we need the (standard) definition of a subcube of the Boolean cube.

Definition 2. For $n, d \in \mathbb{N}$ with $1 \leqslant d \leqslant n$, a *$d$-dimensional subcube* of the n-dimensional Boolean cube $\{0, 1\}^n$ is a subset of $\{0, 1\}^n$ of the form

$$\{x \in \{0, 1\}^n \ : \ x_i = a_i \ \forall \ i \in I\}$$

for some set $I \in \binom{[n]}{n-d}$ and values a_i such that $a_i \in \{0, 1\}$ for each $i \in I$. (The elements of the set I are called the *fixed coordinates* of the subcube; the elements of $[n] \setminus I$ are called the *moving coordinates*.)

(Here, and henceforth, we write $[n] := \{1, 2, \ldots, n\}$ for the standard n-element set.)

Conjecture 3 (Johnson-Talbot)**.** Let $d \geqslant 2$ and $\delta \in (0, 1]$. Then for n sufficiently large depending on d and δ, and any set $A \subset \{0, 1\}^n$ with $|A| \geqslant \delta 2^n$, there exists a d-dimensional subcube $\mathcal{C}$ with $|A \cap \mathcal{C}| \geqslant \binom{d}{\lfloor d/2 \rfloor}$.

It is easy to verify Conjecture 3 for $d = 2$ and $d = 3$, but it remains open for all $d \geqslant 4$. It is easy to see that Conjectures 1 and 3 are equivalent for $d = 2t$; the reader is referred to [2] for details.

The value $\pi(\mathcal{D}_r^t)$ is clearly nondecreasing in t for fixed r, since an r-uniform family that is free of t-daisies is also free of t'-daisies for all $t' > t$. It is also easy to see (by averaging over links of vertices) that the value $\pi(\mathcal{D}_r^t)$ is nonincreasing in r for each fixed t.

In [2], a lower bound of $\pi(\mathcal{D}_r^2) \geqslant r!/r^r$ is observed; this comes from considering the r-partite r-uniform hypergraph on $[n]$ with parts of sizes as equal as possible. However, this lower bound is exponentially small in r. In this paper, we obtain the following improved lower bound, which is polynomial in r, and linear when $t = 2$, using an additive-combinatorial construction. We also raise a question in additive combinatorics which may be of interest in its own right.

Theorem 4. *We have $\pi(D_r^2) = \Omega(1/r)$. Furthermore, for each $t \geqslant 3$, we have*

$$\pi(D_r^t) \geqslant r^{-\frac{4t-2}{\binom{2t}{t}-1} - O(1/\sqrt{\log r})}.$$

Our proof of Theorem 4 relies upon the existence of a subset of $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z}$ that avoids a certain additive structure, which we define now.

Definition 5. For positive integers $m, t \geqslant 2$ with $m \geqslant 2t$, let $g(m,t)$ denote the maximum possible size of a subset $R \subset \mathbb{Z}_m$ such that for any $x_0 \in \mathbb{Z}_m$ and any $(2t)$-element subset X of $\mathbb{Z}_m$, there are t distinct elements of X whose sum is not contained in $R - x_0$, i.e.

$$\left\{ x_0 + \sum_{x \in T} x \ : \ T \subset X, \ |T| = t \right\} \not\subset R.$$

For brevity, given a set $X \in \binom{\mathbb{Z}_m}{2t}$ we write

$$C(X) := \left\{ \sum_{x \in T} x \ : \ T \subset X, \ |T| = t \right\}$$

for the set of sums of t distinct elements of X; $g(m,t)$ is the maximum size of a subset of $\mathbb{Z}_m$ containing no translate of $C(X)$ for any $|X| = 2t$.

The function $g(m,t)$ is related to a question raised by Gunderson and Rödl in [7], concerning Hilbert cubes.

Definition 6. If B is a ring, the *d-dimensional Hilbert cube* generated by $x_1, \ldots, x_d \in B$ is the set

$$\left\{ \sum_{i \in I} x_i \ : \ I \subset \{1, 2, \ldots, d\} \right\} \subset B.$$

Gunderson and Rödl considered large sets of integers which do not contain any translate of a Hilbert cube (working over the integers, i.e., the $B = \mathbb{Z}$ case of Definition 6). In particular, they prove the following (Theorem 2.3 and Theorem 2.5 of [7]).

Theorem 7. *For each integer $d \geqslant 3$, there exists $c_d > 0$ such that any set of integers $A \subset [m]$ with $|A| \geqslant c_d(\sqrt{m}+1)^{2 - \frac{1}{2^{d-2}}}$ contains a translate of a d-dimensional Hilbert cube. Furthermore, for all m there exists a set of integers $A \subset [m]$ with $|A| \geqslant m^{1 - \frac{d}{2^d-1} - O(1/\sqrt{\log m})}$ that does not contain any translate of a Hilbert cube.*

We consider the case $R = \mathbb{Z}_m$, but for our purposes, the structural differences between $[m]$ and $\mathbb{Z}_m$ will not be particularly important. Estimating $g(m,t)$ is a natural variant of the Gunderson-Rödl problem, where we avoid only the middle slice of a Hilbert cube of dimension $2t$. We make a small but important change in that, in Definition 5, we require that X be composed of $2t$ distinct elements, while a Hilbert cube may even have $x_1 = x_2 \cdots = x_d$. In this case, the Hilbert cube is a $(d+1)$-element arithmetic progression, but $C(X)$ is a singleton.

We obtain the following bounds on $g(m,t)$.

Theorem 8. *For all $t \geqslant 3$ and $m \geqslant 4$ we have*

$$g(m,t) \geqslant m^{1 - \frac{2t}{\binom{2t}{t} - 1} - O(1/\sqrt{\log m})},$$

and if furthermore m is prime, then

$$g(m,t) \geqslant m^{1 - \frac{2t-1}{\binom{2t}{t} - 1} - O(1/\sqrt{\log m})}.$$

For $t = 2$ and $m \geqslant 64$ we have $g(m,2) \geqslant \sqrt{m}/8$.

Theorem 9. *For each $t \geqslant 2$ and all m sufficiently large depending on t, we have*

$$g(m,t) \leqslant 4^{1 - 1/2^{2t}} (\sqrt{m} + \sqrt{2t})^{2 - 1/2^{2t-1}}.$$

(Here, we use the standard asymptotic notation: if X is a set and $f, h : X \to \mathbb{R}^+$, we write $f = O(h)$ if there exists an absolute constant $C > 0$ such that $f(x) \leqslant Ch(x)$ for all $x \in X$.)

It would be of interest to narrow the gap between our upper and lower bounds on $g(m,t)$.

The proofs of Theorem 9 and of the first part of Theorem 8 (i.e. the $t > 2$ case) are very similar to those used in [7] to prove Theorem 7 above. The proof of our lower bound for $t > 2$ consists of a probabilistic construction very similar to that of Gunderson and Rödl in [7]; we have to choose a random set of slightly lower density as we must avoid the middle slice of a Hilbert cube, as opposed to an entire Hilbert cube. At first sight, it might seem that an upper bound on $g(m,t)$ follows from the upper bound given in Theorem 7, since a set which contains a Hilbert cube contains its middle layer, but one must make small changes to the proof in [7] so as to ensure that the generators of the Hilbert cube we find (viz., the x_i in the above definition), are distinct. In fact, we find an entire Hilbert cube generated by these x_i, not just the middle layer of such, so there may be some room for improvement.

We give an explicit construction for the $t = 2$ case in Theorem 8: this construction outperforms the probabilistic one in the case $t = 2$. The probabilistic approach does not sample directly from the ground set but instead from a large 3-AP-free set. This choice is critical in (optimizing) the argument, since it destroys additive structure in the (still large) set we obtain.

Having introduced the background and general structure of our approach, we briefly consider the asymmetric version of the daisy problem. For $t \geqslant 1$ and $s \geqslant t + 1$, an *r-uniform (s,t)-daisy* $\mathcal{D}_r^{s,t}$ is defined to be a set of the form

$$\{S \cup T \ : T \subset U, \ |T| = t\}$$

for some sets S, U with $|S| = r - t$, $|U| = s$ and $S \cap U = \emptyset$. (So $\mathcal{D}_r^{2t,t} := \mathcal{D}_r^t$.) In this language, we have so far only considered $(2t,t)$-daisies. Conjecture 1 immediately implies the analog for asymmetric daisies ($s \neq 2t$), since an (s,t)-daisy is contained in a

$(2\max\{t, s-t\}, \max\{t, s-t\})$-daisy. In [2] Bollobás, Leader, and Malvenuto focus on the symmetric case ($s = 2t$), and in this article we do the same. However, the proofs in the sequel may be modified in the obvious way to obtain the following asymmetric analogue of Theorem 4.

Theorem 10. *For each $t \geqslant 2$ and $s \geqslant 4$ with $(s,t) \neq (4,2)$, we have*

$$\pi(D_r^{s,t}) \geqslant r^{-\frac{2s-2}{\binom{s}{t}-1} - O(1/\sqrt{\log r})}.$$

The remainder of this paper is structured as follows. In Section 2, we prove Theorem 4 using Theorem 8. In Section 3, we prove Theorems 8 and 9. Note that we do not require the latter in our study of the Turán density of daisies, but it may be of independent interest.

2 The Proof of Theorem 4

Proof. In proving Theorem 4, by an appropriate choice of c (and of the absolute constant implicit in the big-O notation), we may clearly assume that $r \geqslant 8$. For $r, t \in \mathbb{N}$ with $r \geqslant 8$ and n sufficiently large depending on r and t, we proceed to construct a $\mathcal{D}_r^t$-free family of r-element subsets of $[n]$. We may assume that $n \geqslant 2r^2$. Let L be a prime number such that $r^2 < L < 2r^2$ (such exists, by Bertrand's postulate). We use a 'partite' construction, partitioning $[n]$ into L blocks, and then taking only r-sets containing at most one element from each block. Formally, the block of vertex i will be denoted by a variable x_i, where for $i \in [n]$ we set $x_i = \lfloor L\frac{i-1}{n} \rfloor$; note that $0 \leqslant x_i < L$ for each $i \in [n]$. By Definition 5, there exists a set $R \subset \mathbb{Z}_L$ of size $|R| = g(L,t)$ with the property that for any $X \subset \mathbb{Z}_L$ with $|X| = 2t$ and for any $x_0 \in \mathbb{Z}_L$, we have $x_0 + C(X) \not\subset R$. Define a family $\mathcal{F}_R \subset \binom{[n]}{r}$ by

$$\mathcal{F}_R = \left\{ S \in \binom{[n]}{r} \;\middle|\; \sum_{i \in S} x_i \in R \quad \text{and} \quad (\forall i, j \in S)(x_i = x_j \Rightarrow i = j) \right\}.$$

First, we check that $\mathcal{F}_R$ is $\mathcal{D}_r^t$-free. Indeed, suppose for the sake of a contradiction that $\mathcal{F}_R$ contains a daisy $\mathcal{D} = \{S_0 \cup T : T \subset U, |T| = t\}$, where $S_0, U \subset [n]$ with $|S_0| = r - t$, $|U| = 2t$ and $S_0 \cap U = \emptyset$. Let $x_0 := \sum_{i \in S_0} x_i$. By the above property of R, the $(2t)$-element set

$$X = \{x_i : i \in U\} \subset \mathbb{Z}_L$$

must satisfy $x_0 + C(X) \not\subset R$, and therefore there is a t-sum, indexed by $T = \{i_1, i_2, \ldots, i_t\} \subset U$, say, such that

$$x_0 + x_{i_1} + x_{i_2} + \cdots + x_{i_t} = x_0 + \sum_{i \in T} x_i \notin R.$$

It follows that $S := S_0 \cup T \notin \mathcal{F}_R$ and therefore $\mathcal{D} \not\subset \mathcal{F}_R$, a contradiction, as required.

Now to finish the proof of Theorem 4 we bound $|\mathcal{F}_R|$ from below. First note that there are at least $\frac{1}{2}\binom{n}{r}$ sets $S \in \binom{[n]}{r}$ with $x_i \neq x_j$ for all $i \neq j$, $i, j \in S$. Indeed, choose

a set S uniformly at random from $\binom{[n]}{r}$. Since the probability that a uniformly random two-element subset $\{i, j\}$ of $[n]$ has $x_i = x_j$ is at most $1/L$, we have

$$\mathbb{P}(x_i = x_j \text{ for some } i \neq j, \ i, j \in S) \leqslant (1/L)\binom{n}{2}\binom{n-2}{r-2}/\binom{n}{r}$$
$$= r(r-1)/(2L)$$
$$\leqslant 1/2.$$

The family $\mathcal{F} = \mathcal{F}_R$, defined above, is $\mathcal{D}_r^t$-free even if the set R is replaced by a translate $R_a := R + a$ for some $a \in \mathbb{Z}_L$. Averaging over all such translates yields some translate R_a of R such that $|\mathcal{F}_{R_a}| \geqslant \frac{1}{2}\binom{n}{r}\frac{|R|}{L}$, and therefore

$$\pi(D_r^t) \geqslant \frac{g(L,t)}{2L}.$$

Now we may apply Theorem 8, recalling that $r^2 < L < 2r^2$ is prime. When $t = 2$ and $L = m \geqslant 64$ (which follows from $r \geqslant 8$), we have

$$\pi(D_r^2) \geqslant \frac{g(L,2)}{2L} \geqslant \frac{\sqrt{L}}{16L} \geqslant \frac{\sqrt{r^2}}{32r^2} = \frac{1}{32r}$$

and when $t > 2$ we have

$$\pi(D_r^t) \geqslant \frac{L^{1 - \frac{2t-1}{\binom{2t}{t}-1} - O(1/\sqrt{\log L})}}{L} \geqslant \left(r^2\right)^{-\frac{2t-1}{\binom{2t}{t}-1} - O(1/\sqrt{\log r})} = r^{-\frac{4t-2}{\binom{2t}{t}-1} - O(1/\sqrt{\log r})},$$

as required. $\qquad\square$

3   Bounds on $g(m, t)$

The focus of this section is the analysis of $g(m, t)$.

3.1   Proof of Theorem 8

Proof. First assume $t \geqslant 3$. In this case, we use the idea of Gunderson and Rödl (in [7]) of passing to a fairly dense subset of $\mathbb{Z}_m$ which is free of 3-term arithmetic progressions; this 'destroys' a lot of the additive structure we want to avoid.

By the well-known construction of Behrend in [1], there exists a set $R_0 \subset [\lfloor m/2 \rfloor]$ with $|R_0| = m^{1-\gamma(m)}$ for $\gamma(m) := \frac{4}{\sqrt{\log(m/2)}}$ (here, and elsewhere, log denotes the natural logarithm), such that R_0 contains no 3-term arithmetic progression. Let $R_1 \subset \mathbb{Z}_m$ be the natural embedding of R_0 into $\mathbb{Z}_m$. Then R_1 also contains no 3-term arithmetic progression. Set

$$p = \begin{cases} \frac{1}{8}m^{-\frac{2t-1+\gamma(m)}{\binom{2t}{t}-1}} & \text{if } m \text{ is prime,} \\ \frac{1}{8}m^{-\frac{2t+\gamma(m)}{\binom{2t}{t}-1}} & \text{otherwise,} \end{cases}$$

and choose a set $R_2 \subset R_1$ by including each element of R_1 independently at random with probability p. A standard Chernoff bound (for example Theorem 4.5 in [9]) yields

$$\mathbb{P}(|R_2| \leqslant |R_1|p/2) \leqslant e^{-|R_1|p/8}. \tag{1}$$

Define the random variable

$$Y = |\{x_0 + C(X) : \ x_0 \in \mathbb{Z}_m, \ X \subset \mathbb{Z}_m, \ |X| = 2t, \ x_0 + C(X) \subset R_2\}|.$$

Since R_2 does not contain any 3-term arithmetic progressions, for any set of the form $x_0 + C(X)$ lying within R_2, we must have $|x_0 + C(X)| = |C(X)| = \binom{2t}{t}$. Indeed, suppose for a contradiction that $X = \{x_1, \ldots, x_{2t}\}$ is a $(2t)$-element subset of $\mathbb{Z}_m$ with $x_0 + C(X) \subset R_2$, where $x_0 \in \mathbb{Z}_m$ and $|C(X)| < \binom{2t}{t}$. Then there exist two distinct t-element subsets of X, $\{x_{i_1}, \ldots, x_{i_t}\} = S_1$ and $\{x_{i'_1}, \ldots, x_{i'_t}\} = S_2$ say, such that

$$x_{i_1} + \cdots + x_{i_t} = x_{i'_1} + \cdots + x_{i'_t};$$

we may assume without loss of generality that $x_{i_1} \in S_1 \setminus S_2$ and $x_{i'_1} \in S_2 \setminus S_1$, so that $x_{i_1} \neq x_{i'_j}$ for all j and $x_{i'_1} \neq x_{i_j}$ for all j. Then

$$\{x_{i'_1} + x_{i_2} + \cdots + x_{i_t}, x_{i_1} + x_{i_2} + \cdots + x_{i_t}, x_{i_1} + x_{i'_2} + \cdots + x_{i'_t}\} \subset R_2$$

is a (nontrivial) 3-term arithmetic progression in R_2, a contradiction.

We now proceed to bound $\mathbb{E}Y$ from above. In the case that m is not prime we may crudely bound the number of possible sets of the form $x_0 + C(X)$ from above by m^{2t+1}, which is the number of choices for $x_0, x_1, \ldots, x_{2t} \in \mathbb{Z}_m$. If m is prime then we may assume each such set has $x_0 = 0$, by translating each of $x_1, \ldots, x_{2t}$ by $-t^{-1}x_0$, leaving only m^{2t} choices. The probability that each fixed set of the form $x_0 + C(X)$ lies in R_2 is of course $p^{\binom{2t}{t}}$. It follows that

$$\mathbb{E}Y \leqslant \begin{cases} m^{2t}p^{\binom{2t}{t}} & \text{if } m \text{ is prime} \\ m^{2t+1}p^{\binom{2t}{t}} & \text{otherwise} \end{cases} \Bigg\} \leqslant m^{1-\gamma(m)}\frac{p}{8}.$$

It follows from Markov's inequality that

$$\mathbb{P}(Y \geqslant m^{1-\gamma(m)}p/4) \leqslant 1/2. \tag{2}$$

Combining (1) and (2), we obtain

$$\mathbb{P}\left(|R_2| > m^{1-\gamma(m)}p/2 \text{ and } Y < m^{1-\gamma(m)}p/4\right) \geqslant 1 - e^{-m^{1-\gamma(m)}p/8} - \tfrac{1}{2}. \tag{3}$$

Clearly, for any $t \geqslant 2$ and m sufficiently large depending on t, we have $1 - \gamma(m) - \frac{\gamma(m)+2t}{\binom{2t}{t}-1} > 0$, so for large enough m, the probability in (3) is positive, and therefore there exists a set $R_2 \subset \mathbb{Z}_m$ with $|R_2| > m^{1-\gamma(m)}p/2$ and $Y < m^{1-\gamma(m)}p/4$. Now for each set of the form $x_0 + C(X) \subset R_2$ for $(x_0, X) = (x_0, \{x_1, \ldots, x_{2t}\})$ we remove a single element from R_2,

chosen arbitrarily from $x_0 + C(X)$. The total number of elements deleted from R_2 is at most $Y < m^{1-\gamma(m)}p/4$ and we are still left with

$$|R_2| - Y \geqslant m^{1-\gamma(m)}\frac{p}{4} = \begin{cases} \frac{1}{32}m^{1-\gamma(m)-\frac{2t-1+\gamma(m)}{\binom{2t}{t}-1}} & \text{if } m \text{ is prime} \\ \frac{1}{32}m^{1-\gamma(m)-\frac{2t+\gamma(m)}{\binom{2t}{t}-1}} & \text{otherwise} \end{cases}$$

elements, finishing the proof of the first statement of Theorem 8.

Finally, in the case $t = 2$, we give an algebraic construction that improves upon the random one. First, we recall the definition of a Sidon set.

Definition 11. A *Sidon set* in an Abelian group G is a subset $S \subset G$ such that the only solutions to the equation $a + b = c + d$ with $a, b, c, d \in S$, are the trivial ones (meaning, those with $\{a, b\} = \{c, d\}$).

It follows from the classical construction of Singer [10] that for any prime p there is a Sidon set of size $p + 1$ inside $\mathbb{Z}_{p^2+p+1}$. Assume that $m \geqslant 64$ and let p be a prime with $\sqrt{m}/8 \leqslant p \leqslant \sqrt{m}/4$ (such is furnished by Bertrand's postulate). Let R_0 be a Sidon set of size at least $\sqrt{m}/8$ inside $\mathbb{Z}_{p^2+p+1}$. The image R of R_0 under the natural inclusion map from $\mathbb{Z}_{p^2+p+1}$ to $\mathbb{Z}_m$ is a Sidon set in $\mathbb{Z}_m$ (here, we use $p^2+p+1 \leqslant m/16+\sqrt{m}/4+1 < m/2$). Now we will show that for any x_0 and $X = \{x_1, x_2, x_3, x_4\} \in \binom{\mathbb{Z}_m}{4}$ we have $x_0 + C(X) = x_0 + \{x_1 + x_2, x_1 + x_3, x_1 + x_4, x_2 + x_3, x_2 + x_4, x_3 + x_4\} \not\subset R$. Suppose for a contradiction that $x_0 + C(X) \subset R$; then

$$(x_0 + x_1 + x_2) + (x_0 + x_3 + x_4) = (x_0 + x_1 + x_3) + (x_0 + x_2 + x_4)$$

and each term in brackets is an element of R. Since R is a Sidon set, this implies $x_2 = x_3$ or $x_1 = x_4$, contradicting the fact that the x_i are distinct. We have $|R| \geqslant \sqrt{m}/8$ and therefore we are done in the case $t = 2$. $\qquad\square$

Remark 12. We were not able to generalize the Sidon set approach to $t > 2$, hence our reliance on the probabilistic construction avoiding middle layers of Hilbert cubes. There are some improvements upon Behrend's construction, for example by Elkin in [4] and Green and Wolf in [6]. Using these yields, for $t \geqslant 3$, a slightly better error-term in the exponent of r, but this does not affect the main term in the exponent.

3.2 Proof of Theorem 9

Proof. We begin with a quick calculation.

Lemma 13. *If $m, d,$ and b are positive real numbers with $m \geqslant d+1$ and $b \geqslant \max\{\frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}, 4d+ 1\}$, then $\frac{\binom{b}{2}-db}{m-d} \geqslant \frac{b^2}{4(\sqrt{m}+\sqrt{d})^2}$.*

Proof. Since

$$\frac{1}{b} \leqslant \frac{2\sqrt{d}}{\sqrt{m} + \sqrt{d}},$$

we have

$$\frac{b-1}{b} \geqslant \frac{\sqrt{m} - \sqrt{d}}{\sqrt{m} + \sqrt{d}}.$$

Since $b \geqslant 4d+1$ we have $\binom{b}{2} - db \geqslant \frac{b(b-1)}{4}$, and therefore

$$\binom{b}{2} - db \geqslant \frac{b^2}{4} \frac{\sqrt{m} - \sqrt{d}}{\sqrt{m} + \sqrt{d}}.$$

Dividing by $m - d$ yields the result. □

We may now obtain our upper bound on $g(m,t)$. Let $A \subset \mathbb{Z}_m$ such that

$$|A| \geqslant 4^{1-1/2^{2t}}(\sqrt{m} + \sqrt{2t})^{2-1/2^{2t-1}}. \tag{4}$$

We will show that $x_0 + C(X) \subset A$ for some $x_0 \in \mathbb{Z}_m$ and $X \in \binom{\mathbb{Z}_m}{2t}$. For $x_1, \ldots, x_d \in \mathbb{Z}_m$ and $A \subset \mathbb{Z}_m$ we define $A_{x_1} := A \cap (A - x_1)$, $A_{x_1,x_2} := A_{x_1} \cap (A_{x_1} - x_2)$, and more generally,

$$A_{x_1,\ldots,x_{d-1},x_d} := A_{x_1,\ldots,x_{d-1}} \cap (A_{x_1,\ldots,x_{d-1}} - x_d).$$

Then $A_{x_1,\ldots,x_d} = \{x \in \mathbb{Z}_m : x + \sum_{i \in I} x_i \in A \; \forall \; I \subset [d]\}$ and so A will contain a translate of $C(\{x_1, \ldots, x_{2t}\})$ if $|A_{x_1,\ldots,x_{2t}}| \geqslant 1$. We will find these x_i inductively, using the following claim.

Claim 14. *Provided m is sufficiently large depending on d, for each $0 \leqslant d \leqslant 2t$ there exist d distinct elements $x_1, \ldots, x_d \in \mathbb{Z}_m$ such that*

$$|A_{x_1,\ldots,x_d}| \geqslant \frac{|A|^{2^d}}{4^{2^d-1}(\sqrt{m} + \sqrt{2t})^{2^{d+1}-2}}.$$

(Note that when $d = 0$, the left-hand side of the above is defined to be $|A|$.)

Proof of Claim. The proof is by induction on d (with base case $d = 0$, for which the claim holds trivially). Suppose the claim holds for d for elements $x_1, \ldots, x_d$. Since every pair of elements in $A_{x_1,\ldots,x_d}$ uniquely determine a difference y, we have

$$\sum_{y \in \mathbb{Z}_m} |A_{x_1,\ldots,x_d,y}| \geqslant \binom{|A_{x_1,\ldots,x_d}|}{2}.$$

Therefore, when forbidding y to assume the values $\{x_1, \ldots, x_d\}$, we crudely obtain

$$\sum_{y \in \mathbb{Z}_m \setminus \{x_1,\ldots,x_d\}} |A_{x_1,\ldots,x_d,y}| \geqslant \binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|.$$

By averaging over $y \in \mathbb{Z}_m \setminus \{x_1, \ldots, x_d\}$, there exists $y' \in \mathbb{Z}_m \setminus \{x_1, \ldots, x_d\}$ such that

$$|A_{x_1,\ldots,x_d,y'}| \geqslant \frac{\binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|}{m - d}.$$

We now wish to apply Lemma 13 with $b = |A_{x_1,\ldots,x_d}|$. The hypotheses that $m \geqslant d+1$ and $b \geqslant 4d+1$ are satisfied for m large enough (depending on t), so to apply Lemma 13 it remains only to check that $b \geqslant \frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}$, which follows from our inductive hypothesis and our lower bound (4) on $|A|$, as well as the fact that $d < 2t$ in the inductive step:

$$
\begin{aligned}
b = |A_{x_1,\ldots,x_d}| & \\
&\geqslant \frac{|A|^{2^d}}{4^{2^d-1}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2}} \\
&\geqslant \frac{4^{2^d-2^{d-2t}}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2^{d-2t+1}}}{4^{2^d-1}(\sqrt{m}+\sqrt{2t})^{2^{d+1}-2}} \\
&\geqslant 4^{1-2^{d-2t}}(\sqrt{m}+\sqrt{2t})^{2-2^{d-2t+1}} \\
&\geqslant \frac{\sqrt{m}+\sqrt{d}}{2\sqrt{d}}.
\end{aligned}
$$

Hence, applying Lemma 13 we have

$$
\begin{aligned}
|A_{x_1,\ldots,x_d,y'}| &\geqslant \frac{\binom{|A_{x_1,\ldots,x_d}|}{2} - d|A_{x_1,\ldots,x_d}|}{m-d} \\
&\geqslant \frac{|A_{x_1,\ldots,x_d}|^2}{4(\sqrt{m}+\sqrt{d})^2} \\
&\geqslant \frac{|A|^{2^{d+1}}}{4^{2^{d+1}-2+1}(\sqrt{m}+\sqrt{d})^{2^{d+2}-4+2}},
\end{aligned}
$$

as required. We may therefore choose $x_{d+1} = y'$. $\qquad\square$

Applying Claim 14 with $d = 2t$, and using our lower bound (4) on $|A|$, we obtain distinct $x_1,\ldots,x_{2t} \in \mathbb{Z}_m$ such that $|A_{x_1,\ldots,x_{2t}}| \geqslant 1$, completing the proof of Theorem 9. $\qquad\square$

Acknowledgements

References

[1] F.A. Behrend, On sets of integers which contain no three term arithmetical progression. *Proc. Natl. Acad. Sci. USA* 32 (1946), 331-332.

[2] B. Bollobás, I. Leader and C. Malvenuto, Daisies and other Turán problems. *Combin. Probab. Comput.* 20 (2011), 743-747.

[3] J. Cilleruelo and R. Tesoro, On sets free of sumsets with summands of prescribed size. *Combinatorica* 38 (2017), 511-546.

[4] M. Elkin, An improved construcion of progrssion-free sets. *Israel J. Math* 184 (2011), 93-128.

[5] V. Falgas-Ravry and E. Vaughan, Applications of the Semi-Definite Method to the Turán Density Problem for 3-Graphs. *Combin. Probab. Comput.* 22 (2013), 21-54.

[6] B. Green and J. Wolf, A note on Elkin's improvement of Behrend's construction. *Additive Number Theory* (2010), 141-144.

[7] D.S Gunderson and V. Rödl, Extremal problems for affine cubes of integers. *Combin. Probab. Comput.* 7 (1998), 65-79.

[8] J. R. Johnson and J. Talbot, Vertex Turán problems in the hypercube. *J. Combin. Theory, Series A* 117 (2010), 454-465.

[9] M. Mitzenmacher, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Cambridge University Press, 2005.

[10] J. Singer, A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.* 43 (1938), 377-385.