Combinatorics, Probability and Computing (2016) 25, 108–129.
doi:10.1017/S0963548315000206

On the Number of B_h-Sets

DOMINGOS DELLAMONICA Jr,[1] YOSHIHARU KOHAYAKAWA[1,2,†]
SANG JUNE LEE[3,‡] VOJTĚCH RÖDL[1§] and WOJCIECH SAMOTIJ[4¶]

[1] Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA
(e-mail: domingos.junior@gmail.com, rodl@mathcs.emory.edu)

[2] Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil
(e-mail: yoshi@ime.usp.br)

[3] Department of Mathematics, Duksung Women's University, Seoul 132-714, South Korea
(e-mail: sanglee242@duksung.ac.kr, sjlee242@gmail.com)

[4] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
and
Trinity College, Cambridge CB2 1TQ, UK
(e-mail: samotij@post.tau.ac.il)

Received 8 October 2013; revised 2 July 2015; first published online 16 September 2015

A set A of positive integers is a *B_h-set* if all sums of the form $a_1 + \cdots + a_h$, with $a_1, \ldots, a_h \in A$ and $a_1 \leqslant \cdots \leqslant a_h$, are distinct. We provide asymptotic bounds for the number of B_h-sets of a given cardinality contained in the interval $[n] = \{1, \ldots, n\}$. As a consequence of our results, we address a problem of Cameron and Erdős (1990) in the context of B_h-sets. We also use these results to estimate the maximum size of a B_h-set contained in a typical (random) subset of $[n]$ with a given cardinality.

2010 *Mathematics subject classification*: Primary 11B75, 05A16
Secondary 05D40, 11B83

† Partially supported by FAPESP (2013/03447-6, 2013/07699-0), CNPq (459335/2014-6, 310974/2013-5 and 477203/2012-4), NSF (DMS∼1102086) and NUMEC/USP (Project MaCLinC/USP).
‡ The author is the corresponding author: Supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A1059913), and also by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (no. 2013042157).
§ Supported by NSF grants DMS 0800070, 1301698, and 1102086.
¶ Partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

1. Introduction

We deal with a natural extension of the concept of *Sidon sets*. For a positive integer $h \geqslant 2$, a set A of integers is called a B_h*-set* if all sums of the form $a_1 + \cdots + a_h$, where $a_i \in A$ and $a_1 \leqslant \cdots \leqslant a_h$, are distinct. We obtain Sidon sets letting $h = 2$. A central classical problem on B_h-sets is the determination of the maximum size $F_h(n)$ of a B_h-set contained in $[n] := \{1, \ldots, n\}$. Results by Chowla, Erdős, Singer and Turán [5, 9, 10, 30] from the 1940s yield that $F_2(n) = (1 + o(1))\sqrt{n}$, where $o(1)$ is a function that tends to 0 as $n \to \infty$. In 1962, Bose and Chowla [2] showed that $F_h(n) \geqslant (1 + o(1))n^{1/h}$ for $h \geqslant 3$. On the other hand, an easy argument gives that for every $h \geqslant 3$,

$$F_h(n) \leqslant (h \cdot h! \cdot n)^{1/h} \leqslant h^2 n^{1/h}. \tag{1.1}$$

Successively better bounds of the form $F_h(n) \leqslant c_h n^{1/h}$ were given in [4, 6, 8, 15, 21, 22, 24, 29]. Currently, the best known upper bound on the constant c_h is given by Green [11], who proved that

$$c_3 < 1.519, \quad c_4 < 1.627, \quad \text{and} \quad c_h \leqslant \frac{1}{2e}\left(h + \left(\frac{3}{2} + o(1)\right)\log h\right),$$

where $o(1) \to 0$ as $h \to \infty$. The interested reader is referred to the classical monograph by Halberstam and Roth [12] and to a recent survey by O'Bryant [25] and the references therein.

We study two problems related to the classical question of estimating $F_h(n)$. The first one is a natural generalization, to B_h-sets, of the problem of estimating the *number* of Sidon sets contained in $[n]$, proposed by Cameron and Erdős [3]. Second, we investigate the *maximum size* of a B_h-set contained in a *random subset of* $[n]$, in the spirit of [18, 19, 23]. This second problem belongs to the study of extremal properties of random subsets contained in $[n]$. A well-known result was given in [20] which provided a version of Roth's theorem [26] on 3-term arithmetic progressions for random subsets of $[n]$. Recently, Conlon and Gowers [7] and Schacht [28] proved the far reaching generalizations including a version of Szemerédi's theorem [31] on k-term arithmetic progressions for random subsets of $[n]$. We present and discuss our results in detail in Section 2.

Our notation is standard. We write $a \ll b$ as shorthand for the statement $a/b \to 0$ as $n \to \infty$. We omit floor $\lfloor\ \rfloor$ and ceiling $\lceil\ \rceil$ symbols when they are not essential. We are mostly interested in large n; in our statements and inequalities we often tacitly assume that n is larger than a suitably large constant.

2. The main results

Our main results are presented in two separate sections. We first discuss enumeration results and then we move on to their probabilistic consequences.

2.1. A generalization of a problem of Cameron and Erdős

Let $\mathcal{Z}_n^h$ be the family of B_h-sets contained in $[n]$. In 1990, Cameron and Erdős [3] proposed the problem of estimating $|\mathcal{Z}_n^2|$, that is, the number of Sidon sets contained in $[n]$. We investigate the problem of estimating $|\mathcal{Z}_n^h|$ for arbitrary $h \geqslant 2$. Recalling that $F_h(n)$ is the

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

maximum size of a B_h-set contained in $[n]$, one trivially has

$$2^{F_h(n)} \leqslant |\mathcal{Z}_n^h| \leqslant \sum_{i=0}^{F_h(n)} \binom{n}{i} \leqslant (1+F_h(n))\binom{n}{F_h(n)}.$$

Since $(1+o(1))n^{1/h} \leqslant F_h(n) \leqslant c_h n^{1/h}$ for some constant c_h, we have

$$2^{(1+o(1))n^{1/h}} \leqslant |\mathcal{Z}_n^h| \leqslant n^{c_h' n^{1/h}}, \tag{2.1}$$

for some constant c_h'. We improve the upper bound on $|\mathcal{Z}_n^h|$ in (2.1) as follows.

Theorem 2.1. *For every $h \geqslant 2$, we have $|\mathcal{Z}_n^h| \leqslant 2^{Cn^{1/h}}$, where $C = C(h)$ is a constant that depends only on h.*

The case $h = 2$ in Theorem 2.1 was established in [18] and later given another proof in [27]. Our proof of Theorem 2.1 is based on the solution of a refined version of the question. Let $\mathcal{Z}_n^h(t)$ be the family of B_h-sets contained in $[n]$ with t elements. Theorem 2.1 is obtained from the following result, which estimates $|\mathcal{Z}_n^h(t)|$ for all $t \geqslant n^{1/(h+1)}(\log n)^2$.

Theorem 2.2. *For every $h \geqslant 2$ and any $t \geqslant n^{1/(h+1)}(\log n)^2$,*

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{c_h n}{t^h}\right)^t, \tag{2.2}$$

where $c_h = e^6(2h)^{2h}$.

The derivation of Theorem 2.1 from Theorem 2.2 is given in Section 3, and Theorem 2.2 is proved in Section 4.2.

We now turn to lower bounds for $|\mathcal{Z}_n^h(t)|$. The bound in (2.4) in Proposition 2.3(ii) below complements (2.2) in Theorem 2.2. On the other hand, Proposition 2.3(i) shows that for small t, say, $t \ll n^{1/(2h-1)}$, the B_h-sets in $[n]$ form a much larger proportion of the total number $\binom{n}{t}$ of t-element sets (see (2.3)). Note that for large t, namely, $t \geqslant n^{1/(h+1)}(\log n)^2$, Theorem 2.2 tells us that this proportion is, very roughly speaking, of the order of

$$\left(\frac{n}{t^h}\right)^t \binom{n}{t}^{-1} \leqslant \left(\frac{n}{t^h}\right)^t \Big/ \left(\frac{n}{t}\right)^t = t^{-(h-1)t}.$$

Proposition 2.3. *The following bounds hold for every $h \geqslant 2$.*

(i) *For any $\delta > 0$, there exists an $\varepsilon > 0$ such that, for any $t \leqslant \varepsilon n^{1/(2h-1)}$,*

$$|\mathcal{Z}_n^h(t)| \geqslant (1-\delta)^t \binom{n}{t}. \tag{2.3}$$

(ii) *There are constants c_h' and $\varepsilon' = \varepsilon'(h) > 0$ such that, for all $t \leqslant \varepsilon' n^{1/h}$,*

$$|\mathcal{Z}_n^h(t)| \geqslant \left(\frac{c_h' n}{t^h}\right)^t. \tag{2.4}$$

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

The combination of the lower bounds in Proposition 2.3 and the upper bound of Theorem 2.2 naturally partitions the range of t into three intervals.

- For $t \ll n^{1/(2h-1)}$, Proposition 2.3(i) tells us that $|\mathcal{Z}_n^h(t)|$ is, up to a multiplicative factor of $(1-o(1))^t$, equal to the number $\binom{n}{t}$ of all t-element subsets of $[n]$. In this range, one might therefore say that B_h-sets are 'relatively abundant'.
- For t between $n^{1/(2h-1)}$ and $n^{1/(h+1)}(\log n)^2$, a trivial though loose upper bound follows from the monotonicity of $|\mathcal{Z}_n^h(t)|$, that is, $|\mathcal{Z}_n^h(t)| \leqslant |\mathcal{Z}_n^h(n^{1/(h+1)}(\log n)^2)|$, which is then bounded by Theorem 2.2. We note that the lower bound, given by Proposition 2.3(ii), is quite far from the upper bound. In the final section of this paper we present Conjecture 7.1, which states that the upper bound should essentially match the lower bound of Proposition 2.3(ii).
- For $t \geqslant n^{1/(h+1)}(\log n)^2$, Theorem 2.2 and Proposition 2.3(ii) determine $|\mathcal{Z}_n^h(t)|$ up to a multiplicative factor of the form c^t. In this range of t, B_h-sets are therefore much scarcer than in the first range.

2.2. Almost B_h-sets

We now consider a generalization of the notion of a B_h-set. For a set S of integers and an integer z, let

$$r_{S,h}(z) = |\{(a_1,\dots,a_h) \in S^h \ : \ a_1+\cdots+a_h = z \text{ and } a_1 \leqslant \cdots \leqslant a_h\}|. \tag{2.5}$$

Definition 1. A set S is called a $B_h[g]$-*set* if $r_{S,h}(z) \leqslant g$ for all integers z.

Observe that a $B_h[1]$-set is simply a B_h-set and hence this definition extends the notion of B_h-sets. Let $F_{h,g}(n)$ denote the maximum size of a $B_h[g]$-set contained in $[n]$. It is not hard to see that

$$(1+o(1))n^{1/h} \leqslant F_h(n) \leqslant F_{h,g}(n) \leqslant (gh\cdot h!)^{1/h}n^{1/h}. \tag{2.6}$$

Our final result in this section gives a lower bound for the cardinality of $\mathcal{Z}_n^{h,g}(t)$, the family of t-element $B_h[g]$-sets contained in $[n]$.

Theorem 2.4. *Fix an integer $h \geqslant 2$ and a function $g = g(n)$. For every fixed $\delta > 0$ and integer $1 \leqslant t \ll (n^{1-h!/g})^{1/h}$, we have*

$$(1-\delta)^t\binom{n}{t} \leqslant |\mathcal{Z}_n^{h,g}(t)| \leqslant \binom{n}{t}. \tag{2.7}$$

Notice that the bounds of Theorem 2.4 and Proposition 2.3(i) are the same, but the ranges of t for which each applies differ drastically. Indeed, for $g \gg h!$, one can take t quite close to $n^{1/h}$ in Theorem 2.4, and of course this is essentially best possible, as can be seen from (2.6). In effect, unlike in the case of B_h-sets, apart from a very narrow range of t, $B_h[g]$-sets with t-elements are either 'relatively abundant' or simply do not exist.

The proof of Theorem 2.4 is given in Section 6.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

2.3. Probabilistic results

Let $[n]_m$ be an m-element subset of $[n]$ chosen uniformly at random. We are interested in estimating the cardinality of the largest B_h-sets contained in $[n]_m$. Our bounds for the size of the families $\mathcal{Z}_n^h(t)$ presented in Section 2.1 will be useful in investigating this problem. It will be convenient to have the following definition.

Definition 2. For an integer $h \geqslant 2$ and a set R, let $F_h(R)$ denote the maximum size of a B_h-set contained in R.

The asymptotic behaviour of the random variable $F_2([n]_m)$ was investigated in [18, 19]. Our goal here is to study $F_h([n]_m)$ for arbitrary $h \geqslant 3$. A standard deletion argument implies that, with probability tending to 1 as $n \to \infty$, or *asymptotically almost surely* (a.a.s. for short), we have

$$F_h([n]_m) = (1 + o(1))m \quad \text{if } m = m(n) \ll n^{1/(2h-1)},$$

where $o(1)$ denotes some function that tends to 0 as $n \to \infty$. On the other hand, if we apply the results of Schacht [28] and Conlon and Gowers [7] to B_h-sets, we have that a.a.s.

$$F_h([n]_m) = o(m) \quad \text{if } m = m(n) \gg n^{1/(2h-1)}.$$

Thus $n^{1/(2h-1)}$ is the threshold for the property that $F_h([n]_m) = o(m)$.

The following abridged version of our results yields quite precise information about $F_h([n]_m)$ for a wide range of m and non-trivial but looser bounds for $n^{1/(2h-1)} \leqslant m \leqslant n^{h/(h+1)}$; see also Figure 1.

Theorem 2.5. *Fix $h \geqslant 3$ and let $0 \leqslant a \leqslant 1$ be a fixed constant. Suppose $m = m(n) = n^{a+o(1)}$. Then a.a.s.*

$$n^{b_1+o(1)} \leqslant F_h([n]_m) \leqslant n^{b_2+o(1)}, \tag{2.8}$$

where

$$b_1(a) = \begin{cases} a & \text{for } 0 \leqslant a \leqslant 1/(2h-1), \\ 1/(2h-1) & \text{for } 1/(2h-1) < a \leqslant h/(2h-1), \\ a/h & \text{for } h/(2h-1) < a \leqslant 1, \end{cases} \tag{2.9}$$

and

$$b_2(a) = \begin{cases} a & \text{for } 0 \leqslant a \leqslant 1/(h+1), \\ 1/(h+1) & \text{for } 1/(h+1) < a \leqslant h/(h+1), \\ a/h & \text{for } h/(h+1) < a \leqslant 1. \end{cases} \tag{2.10}$$

We prove the upper bounds in Theorem 2.5 (that is, (2.8) and (2.10)) in Section 3. The lower bounds (that is, (2.8) and (2.9)) are proved in Section 5. Theorem 2.5 determines $b = b(a)$ for which $F_h([n]_m) = n^{b+o(1)}$ when $m = n^{a+o(1)}$ whenever $a \leqslant 1/(2h-1)$ or

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

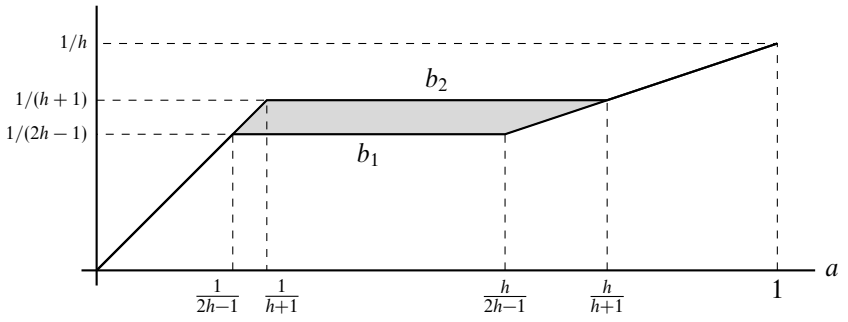

Figure 1. The graphs of $b_1 = b_1(a)$ and $b_2 = b_2(a)$ from the statement of Theorem 2.5.

$a \geqslant h/(h+1)$. An interesting open question is the existence and determination of $b = b(a)$ such that $F_h([n]_m) = n^{b+o(1)}$ for $1/(2h-1) \leqslant a \leqslant h/(h+1)$; see Conjecture 7.2 in Section 7.

As in the previous section, we now move on to consider $B_h[g]$-sets.

Definition 3. For integers $h \geqslant 2$ and $g \geqslant 1$ and a set R, let $F_{h,g}(R)$ denote the maximum size of a $B_h[g]$-set contained in R.

As a natural extension of Theorem 2.5, we investigate the random variable $F_{h,g}([n]_m)$. Trivially, we have

$$F_{h,g}([n]_m) \leqslant \min\{m, F_{h,g}(n)\}. \tag{2.11}$$

Surprisingly, as our next result shows, one can obtain a matching lower bound for this trivial upper bound, up to an $n^{o(1)}$ factor, as long as one allows g to grow with n, however slowly.

Theorem 2.6. *Let $h \geqslant 2$ be an integer and suppose $g(n) \to \infty$ as $n \to \infty$. Let $0 \leqslant a \leqslant 1$ be a fixed constant and suppose $m = m(n) = n^{a+o(1)}$. Then a.a.s.*

$$F_{h,g}([n]_m) = n^{b+o(1)}, \tag{2.12}$$

where

$$b(a) = \begin{cases} a & \text{for } 0 \leqslant a \leqslant 1/h, \\ 1/h & \text{for } 1/h \leqslant a \leqslant 1. \end{cases} \tag{2.13}$$

The upper bound on $F_{h,g}([n]_m)$ contained in Theorem 2.6 follows from (2.11). The lower bound follows from the following more precise result, which is proved in Section 6.

Theorem 2.7. *Fix an integer $h \geqslant 2$ and a function $g = g(n)$. For every fixed $\varepsilon > 0$ and $1 \leqslant m \leqslant (\varepsilon/3h)(n^{1-h!/g})^{1/h}$, we a.a.s. have $F_{h,g}([n]_m) \geqslant (1-\varepsilon)m$.*

We remark that Theorem 2.7 above is closely related to Theorem 2.4 in the previous section. Indeed, we shall derive the latter from the former at the end of Section 6.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

3. Proof of Theorem 2.1 and proof of the upper bounds in Theorem 2.5

We first derive Theorem 2.1 from Theorem 2.2.

Proof of Theorem 2.1. The total number of subsets of $[n]$ having fewer than $n^{1/(h+1)}(\log n)^2$ elements is $2^{o(n^{1/h})}$. Therefore, we may focus on B_h-sets of cardinality at least $n^{1/(h+1)}(\log n)^2$. In particular, by Theorem 2.2,

$$|\mathcal{Z}_n^h| \leqslant 2^{o(n^{1/h})} + \sum_{t \geqslant n^{1/(h+1)}(\log n)^2} \left(\frac{c_h n}{t^h}\right)^t. \tag{3.1}$$

Since the function $t \mapsto (c_h n/t^h)^t$ is maximized when $t = (c_h n)^{1/h}/e$, it follows from (3.1) that, for an appropriate choice of the constant $C = C(h)$,

$$|\mathcal{Z}_n^h| \leqslant 2^{o(n^{1/h})} + n \cdot \left(\frac{c_h n}{c_h n/e^h}\right)^{(c_h n)^{1/h}/e} \leqslant 2^{o(n^{1/h})} + n \cdot \exp\left(\frac{h(c_h n)^{1/h}}{e}\right) \leqslant 2^{Cn^{1/h}}.$$

□

We now turn to the proof of the upper bound on $F_h([n]_m)$ contained in Theorem 2.5. We start with the following easy remark.

Remark 4. At times, it will be convenient to work with the binomial random set $[n]_p$, which is a random subset of $[n]$, with each element of $[n]$ included independently with probability p. The models $[n]_m$ and $[n]_p$, with $p = m/n$, are fairly similar. If some property holds for $[n]_p$ with probability $1 - o(1/\sqrt{pn})$ then the same property holds a.a.s. for $[n]_m$ (this follows from Pittel's inequality; see [14, p. 17]).

The following theorem is a direct corollary of Theorem 2.2.

Theorem 3.1. *For every $p \geqslant n^{-1/(h+1)}(\log n)^{2h}$, we have that a.a.s.*

$$F_h([n]_p) \leqslant O(h^2(pn)^{1/h}).$$

Moreover, the probability that the inequality above fails is at most $\exp(-c(pn)^{1/h})$ for some constant $c = c(h) > 0$.

Proof. The probability that $F_h([n]_p) \geqslant t$ is the same as the probability that there exists a B_h-set of cardinality t in $[n]_p$. Hence, the union bound yields

$$\mathbb{P}[F_h([n]_p) \geqslant t] \leqslant p^t |\mathcal{Z}_n^h(t)|.$$

Let $t = (2c_h np)^{1/h}$, where $c_h = e^6(2h)^{2h}$ is the constant given in Theorem 2.2. By the assumption $p \geqslant n^{-1/(h+1)}(\log n)^{2h}$, we have that

$$\begin{aligned} t &\geqslant (2c_h n^{h/(h+1)}(\log n)^{2h})^{1/h} = (2c_h)^{1/h} n^{1/(h+1)}(\log n)^2 \\ &\geqslant n^{1/(h+1)}(\log n)^2, \end{aligned}$$

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

which satisfies the assumption of Theorem 2.2. Theorem 2.2 gives that

$$\mathbb{P}[F_h([n]_p) \geqslant t] \leqslant p^t |\mathcal{Z}_n^h(t)| \leqslant \left(\frac{pc_h n}{t^h}\right)^t = \left(\frac{1}{2}\right)^t = \exp(-(\log 2)(2c_h)^{1/h}(np)^{1/h}),$$

which completes the proof of Theorem 3.1. □

We now prove the upper bound on $F_h([n]_m)$ given in Theorem 2.5 (see (2.8) and (2.10)). Let us first recall that Remark 4 links the binomial random set $[n]_p$, appearing in Theorem 3.1, to the random set $[n]_m$ that appears in Theorem 2.5. In what follows, we establish (2.8) and (2.10) in Theorem 2.5 using Theorem 3.1. We analyse three ranges of a separately.

(i) $0 \leqslant a \leqslant 1/(h+1)$: From the trivial bound $F_h([n]_m) \leqslant m$, we see that we may take $b_2(a) = a$.

(ii) $1/(h+1) < a \leqslant h/(h+1)$: It is clear that, in probability, $F_h([n]_m)$ is non-decreasing in m. Hence, $b_2(a)$ may be taken to be non-decreasing in a as well. Since, as we show next, we may take $b_2(h/(h+1)) = 1/(h+1)$, this monotonicity lets us take $b_2(a) = 1/(h+1)$ in this range of a.

(iii) $h/(h+1) < a \leqslant 1$: In this range, $b_2(a) = a/h$ follows from Theorem 3.1. Indeed, if $p \geqslant n^{-1/(h+1)}(\log n)^{2h}$, then with probability at least

$$1 - \exp(-c(pn)^{1/h}) \geqslant 1 - o\left(\frac{1}{\sqrt{pn}}\right)$$

we have $F_h([n]_p) \leqslant C(pn)^{1/h}$ for some constant $C > 0$. Remark 4 implies that, a.a.s., $F_h([n]_m) \leqslant Cm^{1/h}$ for all $m \geqslant n^{h/(h+1)}(\log n)^{2h}$, giving that we may take $b_2(a) = a/h$ for $a > h/(h+1)$, as claimed.

4. Upper bounds for the number of B_h-sets of a given cardinality

We prove Theorem 2.2 in this section. For the case where $h = 2$, Theorem 2.2 was shown in [18] (see Theorem 2.1 of [18]). Hence, we assume that $h \geqslant 3$ in this section. We follow a strategy that may be described very roughly as follows. Suppose a B_h-set $S \subset [n]$ of cardinality s is given and one would like to extend it to a larger B_h-set of cardinality s'. We shall show that if s is not too small, then the number of such extensions is very small. To prove Theorem 2.2, we shall apply this fact iteratively, considering a sequence of cardinalities $s < s' < s'' < \cdots$.

4.1. Bounding the number of extensions of B_h-sets

We use a graph-based approach to bounding the number of extensions of a large B_h-set to a larger B_h-set. This approach is inspired by the work of Kleitman and Winston [17] and Kleitman and Wilson [16]. We start with the following simple observation. If two distinct elements $x, y \in [n] \setminus S$ satisfy

$$x + a_1 + \cdots + a_{h-1} = y + b_1 + \cdots + b_{h-1}$$
$$\text{for some } \{a_1, \ldots, a_{h-1}\}, \{b_1, \ldots, b_{h-1}\} \in \binom{S}{h-1}, \tag{4.1}$$

then $S \cup \{x, y\}$ is clearly not a B_h-set. This motivates our next definition.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

Definition 5. The *collision graph* CG_S is a graph on the vertex set $[n] \setminus S$ whose edges are all pairs of distinct elements $x, y \in [n] \setminus S$ that satisfy (4.1).

Clearly, by the construction of CG_S, we have that if $I \subseteq [n] \setminus S$ is such that $I \cup S$ is a B_h-set, then I is an independent set in CG_S.

One of our main tools is the following lemma, implicit in the work of Kleitman and Winston [17], which provides an upper bound on the number of independent sets in graphs that have many edges in each sufficiently large vertex subset (see (4.3)). Lemma 4.1 in the version presented below is stated and proved in [18, 19], where it is used to bound the number of Sidon subsets of $[n]$. For other applications of this lemma to problems in additive combinatorics, we refer the reader to [1].

Lemma 4.1. *Let δ and $\beta > 0$ and $q \in \mathbb{N}$ be numbers satisfying*

$$e^{\beta q}\delta > 1. \tag{4.2}$$

Suppose that $G = (V, E)$ is a graph satisfying

$$e_G(A) \geqslant \beta\,|A|^2 \quad \text{for all } A \subset V \text{ with } |A| \geqslant \delta\,|V|. \tag{4.3}$$

Then, for every $m \geqslant 1$, there are at most

$$\binom{|V|}{q}\binom{\delta|V|}{m} \tag{4.4}$$

independent sets in G of size $q + m$.

Remark 6. When we apply Lemma 4.1 to CG_S, we shall take $m \gg q$ to take advantage of the upper bound (4.4). In condition (4.3), there is a trade-off between β (larger is better) and δ (smaller is better) which needs to be optimized.

We wish to show that CG_S satisfies (4.3) with good parameters β and δ. To that end, we shall make use of another auxiliary graph, which we now define.

Definition 7. Let $\widetilde{\mathrm{CG}}_S$ be a multigraph version of CG_S, where the multiplicity of a pair $\{x, y\}$ of distinct $x, y \in [n] \setminus S$ is given by the number of pairs $(\{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\}) \in \binom{S}{h-1}^2$ that satisfy (4.1).

Lemma 4.2. *For every B_h-set S with $s \geqslant h$ elements and $A \subset [n] \setminus S$ with $|A| \geqslant h^{2h}n/s^{h-1}$, we have*

$$e_{\widetilde{\mathrm{CG}}_S}(A) \geqslant \frac{s^{2h-2}}{h^{2h}n}|A|^2, \tag{4.5}$$

where the edges in $\widetilde{\mathrm{CG}}_S$ are counted with multiplicity.

The proof of Lemma 4.2 will be given in Section 4.3. In view of Lemma 4.2, if the maximal multiplicity of an edge in $\widetilde{\mathrm{CG}}_S$ is at most r, then the graph CG_S satisfies the

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

conditions of Lemma 4.1 with $V = [n]$, $\beta = s^{2h-2}/(h^{2h}rn)$ and $\delta = h^{2h}/s^{h-1}$. Consequently, we are interested in bounding the multiplicity of the edges of $\widetilde{\mathrm{CG}_S}$.

Proposition 4.3. *For every B_h-set S of cardinality s, the maximal multiplicity of an edge in $\widetilde{\mathrm{CG}_S}$ does not exceed s^{h-2}.*

We postpone the proof of Proposition 4.3 to Section 4.4. The following is an immediate corollary of Lemma 4.2 and Proposition 4.3.

Corollary 4.4. *If S is a B_h-set with s elements, then for any $A \subset [n]\backslash S$ with $|A| \geqslant h^{2h}n/s^{h-1}$,*

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{s^h}{h^{2h}n}|A|^2.$$

4.2. Proof of Theorem 2.2

The case $h = 2$ of Theorem 2.2 is proved in [18], and we therefore restrict ourselves to $h \geqslant 3$ here. We shall in fact prove the following: for every $h \geqslant 3$ and

$$t \geqslant h^2 n^{1/(h+1)}(\log n)^{1+1/(h+1)}, \tag{4.6}$$

we have that

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{2^{2h}e^6h^{2h}n}{t^h}\right)^t.$$

In view of (1.1), we have $\mathcal{Z}_n^h(t) = 0$ for $t > h^2n^{1/h}$. Hence we assume

$$t \leqslant h^2n^{1/h}, \tag{4.7}$$

that is, $h^2n^{1/(h+1)}(\log n)^{1+1/(h+1)} \leqslant t \leqslant h^2n^{1/h}$. Let

$$s_0 = h^2(n\log n)^{1/(h+1)} \tag{4.8}$$

and let K be the largest integer satisfying $t2^{-K} \geqslant 2s_0$. We define three sequences $(s_k)_{0\leqslant k\leqslant K}$, $(q_k)_{0\leqslant k\leqslant K}$ and $(m_k)_{0\leqslant k\leqslant K}$ as follows. We let

$$q_0 = s_0/2 \quad \text{and} \quad m_0 = t2^{-K} - s_0 - q_0. \tag{4.9}$$

For $k = 1,\dots,K$, we let

$$s_k = t2^{-K+k-1}, \tag{4.10}$$

$$q_k = q_0 2^{-hk}, \tag{4.11}$$

$$m_k = s_{k+1} - s_k - q_k. \tag{4.12}$$

We will bound the number of sequences $S_0 \subset \cdots \subset S_K \subset S_{K+1}$ of B_h-sets with $|S_{K+1}| = t$ and $|S_k| = s_k$ for all $k = 0,\dots,K$, from which a bound on $|\mathcal{Z}_n^h(t)|$ will easily follow. Although we will only use the trivial bound $\binom{n}{s_0}$ for the number of choices for S_0, we will then employ Lemma 4.1 to obtain a non-trivial bound on the number of extensions of S_k to S_{k+1} for all k.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

Let us now estimate the number of extensions of a B_h-set S_k to a larger B_h-set S_{k+1} for some $k = 0, \dots, K$. By Corollary 4.4, the graph CG_{S_k} is such that for all $A \subset [n] \setminus S_k$ with $|A| \geqslant h^{2h} n / s_k^{h-1}$,

$$e_{\mathrm{CG}_{S_k}}(A) \geqslant \beta_k |A|^2, \quad \text{where } \beta_k = \frac{s_k^h}{h^{2h} n}.$$

Let

$$\delta_k = h^{2h} / s_k^{h-1} \geqslant 1/n \tag{4.13}$$

and observe that

$$e^{\beta_k q_k} = \exp\left(\frac{s_k^h}{h^{2h} n} \cdot \frac{q_0}{2^{hk}} \right) \overset{(4.10)}{\geqslant} \exp\left(\frac{(2^k s_0)^h \cdot s_0}{h^{2h} n \cdot 2^{hk+1}} \right) \geqslant \exp\left(\frac{s_0^{h+1}}{2h^{2h} n} \right) \overset{(4.8)}{\geqslant} n \overset{(4.13)}{\geqslant} \delta_k^{-1}.$$

Consequently, CG_{S_k}, δ_k, β_k and q_k satisfy the conditions of Lemma 4.1. Note that $S_{k+1} \setminus S_k$ must be an independent set in CG_{S_k} with cardinality $s_{k+1} - s_k = q_k + m_k$. Therefore, by Lemma 4.1, the number of extensions of S_k into a B_h-set S_{k+1} is at most

$$\binom{n}{q_k} \binom{\delta_k n}{m_k}. \tag{4.14}$$

In order to obtain an upper bound of (4.14), we first claim that

$$\binom{\delta_0 n}{m_0} \leqslant \binom{\delta_0 n}{3s_0} \tag{4.15}$$

and

$$\binom{\delta_k n}{m_k} \leqslant \binom{\delta_k n}{s_k} \tag{4.16}$$

for all $1 \leqslant k \leqslant K$. Indeed, inequality (4.15) follows from the fact that $m_0 = s_1 - s_0 - q_0 \leqslant 4s_0 - s_0 \leqslant 3s_0$ and also $3s_0 \leqslant \delta_0 n/2$. Inequality (4.16) follows from the fact that for all $1 \leqslant k \leqslant K$, $m_k \leqslant s_k \leqslant \delta_k n/2$ as

$$\frac{s_k}{\delta_k} \overset{(4.13)}{=} \frac{s_k^h}{h^{2h}} \leqslant \frac{s_K^h}{h^{2h}} \overset{(4.10)}{=} \frac{(t/2)^h}{h^{2h}} \overset{(4.7)}{\leqslant} \frac{n}{2^h}.$$

Hence,

$$\binom{n}{q_0} \binom{\delta_0 n}{m_0} \leqslant \binom{n}{q_0} \binom{\delta_0 n}{3s_0} \leqslant \binom{n}{q_0} \binom{n}{3s_0} \leqslant n^{q_0} n^{3s_0},$$

and for all $1 \leqslant k \leqslant K$

$$\binom{n}{q_k} \binom{\delta_k n}{m_k} \leqslant \binom{n}{q_k} \binom{\delta_k n}{s_k} \leqslant n^{q_k} \left(\frac{e\delta_k n}{s_k} \right)^{s_k} = n^{q_k} \left(\frac{eh^{2h} n}{s_k^h} \right)^{s_k}.$$

Applying (4.14) iteratively implies that

$$|\mathcal{Z}_n^h(t)| \leqslant \binom{n}{s_0} \prod_{k=0}^{K} \binom{n}{q_k} \binom{\delta_k n}{m_k} \leqslant n^{4s_0 + \sum_{k=0}^{K} q_k} \prod_{k=1}^{K} \left(\frac{eh^{2h} n}{s_k^h} \right)^{s_k}. \tag{4.17}$$

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

Finally, since

$$\sum_{k=0}^{K} q_k \overset{(4.11)}{=} q_0 \sum_{k=0}^{K} 2^{-hk} \leqslant 2q_0 \overset{(4.9)}{=} s_0 \overset{(4.6),(4.8)}{\leqslant} \frac{t}{\log n}$$

and

$$\begin{aligned}\prod_{k=1}^{K}\left(\frac{eh^{2h}n}{s_k^h}\right)^{s_k} &\overset{(4.10)}{=} \prod_{\ell=1}^{K}\left(\frac{eh^{2h}n}{(t2^{-\ell})^h}\right)^{t2^{-\ell}} \\ &\leqslant \left[2^{h\sum_{\ell=1}^{\infty}\ell 2^{-\ell}} \cdot \left(\frac{eh^{2h}n}{t^h}\right)^{\sum_{\ell=1}^{\infty}2^{-\ell}}\right]^t \\ &= \left(\frac{2^{2h}eh^{2h}n}{t^h}\right)^t,\end{aligned}$$

Theorem 2.2 follows from (4.17).

4.3. Proof of Lemma 4.2

Let S be a B_h-set with s elements. Let $A \subset [n] \setminus S$ be an arbitrary subset with $|A| \geqslant h^{2h}n/s^{h-1}$. Consider the auxiliary bipartite graph Γ defined as follows. The vertex classes of Γ are A and a disjoint copy of $[hn]$. The edge set of Γ is defined as

$$E(\Gamma) = \left\{(x,u) \in A \times [hn] \ : \ u = x + a_1 + \cdots + a_{h-1} \text{ for some } \{a_1,\ldots,a_{h-1}\} \in \binom{S}{h-1}\right\}.$$

Note that, because S is a B_h-set, for fixed x and u, there is at most one solution to $u = x + a_1 + \cdots + a_{h-1}$ with $\{a_1,\ldots,a_{h-1}\} \in \binom{S}{h-1}$. We will now argue that the multiplicity of a pair $\{x,y\} \in \binom{A}{2}$ in the multigraph $\widetilde{\mathrm{CG}}_S$ is the number of paths of length two connecting x to y in Γ. Indeed, there is a bijection between pairs

$$(\{a_1,\ldots,a_{h-1}\},\{b_1,\ldots,b_{h-1}\}) \in \binom{S}{h-1}^2$$

that satisfy (4.1) and paths xuy in Γ, where

$$u = x + a_1 + \cdots + a_{h-1} = y + b_1 + \cdots + b_{h-1}.$$

Consequently, $e_{\widetilde{\mathrm{CG}}_S}(A)$ is the number of paths of length two in Γ containing two vertices in the class A. By Jensen's inequality applied to the convex function $f(\alpha) = \binom{\alpha}{2} = \alpha(\alpha-1)/2$,

$$e_{\widetilde{\mathrm{CG}}_S}(A) \geqslant \sum_{u\in[hn]} \binom{\deg_\Gamma(u)}{2} \geqslant hn\binom{e(\Gamma)/hn}{2}.$$

On the other hand,

$$e(\Gamma) = \sum_{x\in A} \deg_\Gamma(x) = |A|\binom{s}{h-1} \geqslant \left(\frac{s}{h}\right)^{h-1}|A|.$$

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

It follows that $e(\Gamma) \geqslant h^h n$, and thus

$$e_{\widetilde{\mathrm{CG}}_S}(A) \geqslant hn\binom{e(\Gamma)/hn}{2} \geqslant e(\Gamma)\left(\frac{e(\Gamma)-hn}{2hn}\right) \geqslant \frac{e(\Gamma)^2}{hn}\left(\frac{h^h-h}{2h^h}\right) \geqslant \frac{e(\Gamma)^2}{3hn} \geqslant \frac{s^{2h-2}}{h^{2h}n}|A|^2.$$

This concludes the proof of Lemma 4.2.

4.4. Proof of Proposition 4.3

Let S be a B_h-set of cardinality s and let $x \neq y \in [n]$ be arbitrary. By definition, the multiplicity of $\{x, y\}$ in $\widetilde{\mathrm{CG}}_S$ is the number of pairs of sets $\{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\} \in \binom{S}{h-1}$ such that

$$y - x = a_1 + \cdots + a_{h-1} - (b_1 + \cdots + b_{h-1}).$$

Since $x \neq y$, we clearly have $\{a_1, a_2, \dots, a_{h-1}\} \neq \{b_1, b_2, \dots, b_{h-1}\}$. Hence, we may assume, without loss of generality, that $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$. Let us now bound the number of possible sets in the following way: first, pick arbitrary values of $a_1, \dots, a_{h-2} \in S$, then find values (if any exist) $a_{h-1}, b_1, \dots, b_{h-1}$, with $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$, that satisfy

$$y - x - a_1 - a_2 - \cdots - a_{h-2} = a_{h-1} - (b_1 + \cdots + b_{h-1}).$$

We claim that for each fixed sequence $a_1, \dots, a_{h-2} \in S$ there is at most one such completion (up to the order of elements b_i) that satisfies the above equality. Indeed, suppose that we also have $a'_{h-1}, b'_1, \dots, b'_{h-1} \in S$ such that

$$a'_{h-1} - (b'_1 + \cdots + b'_{h-1}) = a_{h-1} - (b_1 + \cdots + b_{h-1}).$$

Then, since S is a B_h-set, the following is a (multi)set equality:

$$\{a_{h-1}, b'_1, \dots, b'_{h-1}\} = \{a'_{h-1}, b_1, \dots, b_{h-1}\}.$$

Moreover, since $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$, we have $a'_{h-1} = a_{h-1}$, which implies that we also have $\{b'_1, \dots, b'_{h-1}\} = \{b_1, \dots, b_{h-1}\}$.

In conclusion, fixing any of the s^{h-2} choices for $a_1, \dots, a_{h-2} \in S$ completely determines both a_{h-1} and $\{b_1, \dots, b_{h-1}\}$. The proposition follows.

5. Lower bounds

In this section, we establish the lower bounds in Theorem 2.5 and prove Proposition 2.3. For conciseness, we shall be somewhat sketchy when dealing with routine arguments.

First, we show that a simple deletion argument (given in Lemma 5.1 below) yields that if $m \ll n^{1/(2h-1)}$, then $F_h([n]_m) = (1 - o(1))m$. This immediately implies that in Theorem 2.5, for $0 \leqslant a < 1/(2h-1)$, one may take $b_1(a) = a$ (see (2.8) and (2.9)). Since $F_3([n]_m)$ is non-decreasing in probability with respect to m, for $a \geqslant 1/(2h-1)$, we may take $b_1(a) = 1/(2h-1)$. Moreover, as an easy corollary of Lemma 5.1, we will also derive Proposition 2.3(i).

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

In the second part of this section, following the strategy of [18, 19], for every $t = o(n^{1/h})$, we will describe a deterministic construction of a large subfamily of $\mathcal{Z}_n^h(t)$. The existence of such a subfamily will immediately imply Proposition 2.3(ii). Moreover, we shall show that if $1 \ll m \leqslant n$, then a.a.s. the set $[n]_m$ contains a B_h-set with $\Omega(m^{1/h})$ elements from the constructed subfamily. This yields that in Theorem 2.5, we may take $b_1(a) = a/h$ for all $0 \leqslant a \leqslant 1$. Note that, in the range $1/(2h-1) \leqslant a \leqslant h/(2h-1)$, this is superseded by the bound obtained in the first part, that is, $b_1(a) = 1/(2h-1)$.

Lemma 5.1. *If $1 \leqslant m = o(n^{1/(2h-1)})$, then we a.a.s. have $m \geqslant F_h([n]_m) \geqslant (1 - o(1))m$.*

Proof. Let $1 \leqslant m \ll n^{1/(2h-1)}$ and let X be the random variable that counts the number of solutions to

$$a_1 + \cdots + a_h = b_1 + \cdots + b_h \quad \text{with } \{a_1, \ldots, a_h\} \neq \{b_1, \ldots, b_h\} \tag{5.1}$$

and $a_i, b_i \in [n]_m$ for all $i \in [h]$. Let $p = m/n$. It follows from the linearity of expectation that

$$\mathbb{E}[X] = O\left(\sum_{k=2}^{2h-1} p^{k+1} n^k\right) = O\left(p^{2h} n^{2h-1}\right) = o(m).$$

Hence, by Markov's inequality, we a.a.s. have $X = o(m)$. Since deleting from $[n]_m$ one element from the set $\{a_1, b_1, \ldots, a_h, b_h\}$ for each of the X solutions to (5.1) yields a B_h-set, the lemma follows. □

Proof of Proposition 2.3(i). Fix a constant $\delta > 0$. Choose $\beta > 0$ sufficiently small that $(1-\beta)(1-\delta/2) \geqslant 1 - \delta$ and $\binom{(1+\beta)t}{\beta t} \leqslant (1+\delta/2)^t$ for all t. Let $\varepsilon > 0$ be a small constant. Assume that $t \leqslant \varepsilon n^{1/(2h-1)}$. Lemma 5.1 with $m = (1+\beta)t$ implies that if ε is sufficiently small, then $F_h([n]_m) \geqslant t$ with probability at least $1 - \beta$. It follows that, for large enough n, we have

$$\begin{aligned} |\mathcal{Z}_n^h(t)| &\geqslant (1-\beta)\binom{n}{(1+\beta)t}\binom{n-t}{\beta t}^{-1} = (1-\beta)\binom{n}{t}\binom{(1+\beta)t}{\beta t}^{-1} \\ &\geqslant (1-\beta)(1-\delta/2)^t\binom{n}{t} \geqslant (1-\delta)^t\binom{n}{t}, \end{aligned} \tag{5.2}$$

as required. □

In order to construct a large family of B_h-sets for larger t, we will use the following theorem of Bose and Chowla [5] (with the statement adapted for our purposes).

Theorem 5.2. *Fix an integer $h \geqslant 2$. For every m, there exists a B_h-set $Y \subset \mathbb{Z}_m$ with $|Y| = \Omega(m^{1/h})$.*

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

From Theorem 5.2 we obtain the following corollary.

Corollary 5.3. *Let n and m satisfy $n \geqslant 3hm$ and suppose that $Y \subset \mathbb{Z}_m$ is a B_h-set. Then, there exists at least*

$$\left(\frac{n}{3hm}\right)^{|Y|}$$

B_h-sets of cardinality $|Y|$ in $[n]$.

More precisely, there are pairwise disjoint sets $I_0, \dots, I_{m-1} \subset [n]$, each of cardinality $\ell \geqslant n/(3hm)$, with the following property: for any B_h-set $Y \subset \mathbb{Z}_m$, all sets $S \subset [n]$ with $|S| = |Y|$ such that $|S \cap I_j| = 1$ for all $j \in Y$ are B_h-sets.

Proof. Let $k = \lfloor n/m \rfloor$ and $\ell = \lfloor n/(2hm) \rfloor \geqslant n/(3hm) \geqslant 1$. Define the integer intervals

$$I_j = [jk + 1, jk + \ell], \quad j = 0, 1, \dots, m - 1$$

and note that by construction they are all pairwise disjoint subsets of $[n]$. Let S be an arbitrary set which contains a single element of each I_j with $j \in Y$ and no additional elements.

We claim that S is a B_h-set. Indeed, suppose that

$$a_1 + \cdots + a_h = b_1 + \cdots + b_h$$

for some $a_1, b_1, \dots, a_h, b_h \in S$ with $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_h$, $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_h$. For $i = 1, \dots, h$, let $u_i \in Y$ be the unique index such that $a_i \in I_{u_i}$. Set $u = \sum_{i=1}^{h} u_i$ and notice that

$$a_1 + \cdots + a_h \in [ku + h, ku + \ell h] \subset [ku + 1, k(u + 1)].$$

Now for $i = 1, \dots, h$, let $v_i \in Y$ be the unique index such that $b_i \in I_{v_i}$, and set $v = \sum_{i=1}^{h} v_i$. Notice that the same argument as above yields

$$b_1 + \cdots + b_h \in [kv + 1, k(v + 1)],$$

which thus means that $u_1 + \cdots + u_h = u = v = v_1 + \cdots + v_h$. Since Y is a B_h-set, we have $\{u_1, \dots, u_h\} = \{v_1, \dots, v_h\}$. Since the elements a_i and b_i are in increasing order, the same holds for the u_i and v_i and thus $u_i = v_i$ for all i. Moreover, by construction, $|S \cap I_j| = 1$ for all $j \in Y$, which means that $\{a_i\} = S \cap I_{u_i} = S \cap I_{v_i} = \{b_i\}$ for all $i = 1, \dots, h$.

The above argument shows that S is a B_h-set. Since there are $\ell^{|Y|}$ choices for the construction of S, the corollary follows. □

The proof of Proposition 2.3(ii) easily follows from Corollary 5.3.

Proof of Proposition 2.3(ii). From Theorem 5.2 we obtain a constant C satisfying that for all values of t, there exists $m = m(t) \leqslant Ct^h$ such that there exists a B_h-set of cardinality t in $\mathbb{Z}_m$. We may also assume that $m(\cdot)$ is monotone.

Let $\varepsilon' > 0$ be such that $m(\varepsilon' n^{1/h}) \leqslant n/(3h)$. It follows that for any $t \leqslant \varepsilon' n^{1/h}$, there is a B_h-set $Y \subset \mathbb{Z}_{m(t)}$ with $|Y| = t$. Since $m(t) \leqslant n/(3h)$, applying Corollary 5.3 to Y shows that

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

there are at least

$$\left(\frac{n}{3hm(t)}\right)^t \geqslant \left(\frac{n}{3hCt^h}\right)^t$$

B_h-sets of cardinality t in $[n]$, which establishes the proposition with $c'_h = 1/(3hC)$. □

Next, we show that Corollary 5.3 also yields the lower bound in Theorem 2.5.

Lemma 5.4. *For any $1 \ll m \leqslant n$, we a.a.s. have $F_h([n]_m) = \Omega(m^{1/h})$.*

Proof. Lemma 5.1 implies that $F_h([n]_m) = \Omega(m^{1/h})$ for $m \ll n^{1/(2h-1)}$. We now assume that $m \gg n^{1/(2h)}$, which covers the remaining range of m (with plenty to spare). It will be convenient for us to use the model $[n]_p$ with $p = m/n$ rather than $[n]_m$ (recall Remark 4).

Without loss of generality we may assume that $n \geqslant 3hm$, since we just need to adjust the constant hidden in the Ω in the bound from the statement of the lemma. Let I_j, $j = 0, \ldots, m-1$, be sets obtained from Corollary 5.3. From Theorem 5.2, we may obtain a B_h-set $Y \subset \mathbb{Z}_m$ with $|Y| = \Omega(m^{1/h})$.

Consider, for each $j \in Y$, the intersection of the random set $[n]_p$ and I_j. The probability q that this intersection is empty satisfies

$$q = (1-p)^{|I_j|} \leqslant \exp(-p\,|I_j|) = \exp\left(-p\frac{n}{3hm}\right) = \exp\left(-\frac{1}{3h}\right).$$

Notice that q is bounded away from 1 by a constant depending only on h. Let r be the random variable denoting the number of sets I_j, $j \in Y$, that intersect $[n]_p$, namely

$$r = |\{j \in Y \ : I_j \cap [n]_p \neq \emptyset\}|.$$

Since the sets I_j are disjoint, r is a binomial random variable with parameters $|Y|$ and $1-q$. Also note that by collecting an element from each set I_j, $j \in Y$, that intersects $[n]_p$, we have a B_h-set which is a subset of the random set, thus

$$F_h([n]_p) \geqslant r.$$

By Chernoff's bound, we have that $r \geqslant (1-q)|Y|/2$ with probability

$$1 - \exp(-c\,|Y|) \geqslant 1 - \exp(-c'm^{1/h}) = 1 - o(\sqrt{pn}),$$

for some constants $c, c' > 0$. In particular, with probability $1 - o(\sqrt{pn})$ there is a B_h-set $S \subset [n]_p$ with cardinality $r = \Omega(|Y|) = \Omega(m^{1/h})$. The lemma follows from Remark 4. □

6. Proofs of Theorems 2.4 and 2.7

We need some preparation for the proofs of Theorems 2.4 and 2.7. For the remainder of this section, we fix an integer $h \geqslant 2$ and a function $g = g(n)$. Since we are only proving asymptotic results, we shall make the technical assumption that n is relatively prime to $h!$. Furthermore, it will be more convenient for us to work with modular arithmetic, that is, we consider addition modulo n. Clearly, any modular $B_h[g]$-subset of $\mathbb{Z}_n$ naturally

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

corresponds to a $B_h[g]$-subset of $[n]$ and hence the claimed lower bound results for $[n]$ follows from the corresponding results for $\mathbb{Z}_n$.

Recall the definition of $r_{S,h}$ (see (2.5) in Section 2.3). Observe that S is a $B_h[g]$-set if and only if $r_{S,h}(z) \leqslant g$ for every $z \in \mathbb{Z}_n$. In order to show that $r_{S,h}(z) \leqslant g$ for every $z \in \mathbb{Z}_n$, we define the following.

For every $1 \leqslant \ell \leqslant h$ and $\lambda > 0$ and $S \subset \mathbb{Z}_n$, let

$$E_{S,\ell}(\lambda) = \sum_{z \in \mathbb{Z}_n} \exp(\lambda r_{S,\ell}(z)). \tag{6.1}$$

Clearly, for every $z \in \mathbb{Z}_n$, we have that $r_{S,h}(z) \leqslant \lambda^{-1} \log(E_{S,h}(\lambda))$. Hence, in order to bound $r_{S,h}(z)$ for every $z \in \mathbb{Z}_n$, it suffices to bound $E_{S,h}(\lambda)$ for some appropriate choice of λ. We remark that the definition (6.1) is heavily inspired by *moment generating functions* studied in probability theory. Indeed, if z is sampled uniformly over $\mathbb{Z}_n$, then $r_{S,\ell}$ becomes a random variable whose moment generating function is $\mathbb{E}[e^{\lambda r_{S,\ell}}] = \frac{1}{n} E_{S,\ell}(\lambda)$.

The following claim bounds the average increase of $E_{S,\ell}(\lambda)$ as we add some $y \in \mathbb{Z}_n$ to S.

Claim 6.1. *Let $h \geqslant 2$ be fixed, and let n be a number relatively prime to $h!$ and $\ell \in [h]$. Then, for any $\emptyset \neq S \subset \mathbb{Z}_n$ and $\lambda > 0$, we have*

$$\mathbb{E}_{y \in \mathbb{Z}_n}[E_{S \cup \{y\},\ell}(\lambda) - E_{S,\ell}(\lambda)] \leqslant \frac{1}{n} E_{S,\ell}(\lambda)(E_{S,\ell-1}(\ell\lambda) - n). \tag{6.2}$$

Proof. Note first that

$$r_{S \cup \{y\},\ell}(z) \leqslant r_{S,\ell}(z) + \mathbf{1}[z = \ell y] + \sum_{i=1}^{\ell-1} r_{S,\ell-i}(z - iy). \tag{6.3}$$

Hence,

$$\begin{aligned} \sum_{y \in \mathbb{Z}_n} E_{S \cup \{y\},\ell}(\lambda) &= \sum_{y \in \mathbb{Z}_n} \sum_{z \in \mathbb{Z}_n} \exp(\lambda r_{S \cup \{y\},\ell}(z)) \qquad (6.4) \\ &\leqslant \sum_{y \in \mathbb{Z}_n} \sum_{z \in \mathbb{Z}_n} \exp\left\{\lambda\left(r_{S,\ell}(z) + \mathbf{1}[z = \ell y] + \sum_{i=1}^{\ell-1} r_{S,\ell-i}(z - iy)\right)\right\} \\ &= \sum_{z \in \mathbb{Z}_n} \left\{\exp(\lambda r_{S,\ell}(z)) \underbrace{\sum_{y \in \mathbb{Z}_n} \exp(\lambda \mathbf{1}[z = \ell y]) \prod_{i=1}^{\ell-1} \exp(\lambda r_{S,\ell-i}(z - iy))}_{Q(z,\lambda)}\right\}. \end{aligned}$$

We will now use a variation[1] of Hölder's inequality (see, *e.g.*, [13, p. 22]):

$$\sum_{i=1}^{m} \left| \prod_{j=1}^{\ell} a_{ij} \right| \leqslant \prod_{j=1}^{\ell} \left(\sum_{i=1}^{m} |a_{ij}|^{\ell} \right)^{1/\ell}. \tag{6.5}$$

[1] This form can be obtained from the original as follows. For $\ell = 2$, it is a special case of Hölder's inequality with reciprocals $1/2 + 1/2 = 1$. For $\ell \geqslant 3$ it follows by induction using Hölder's inequality with reciprocals $\ell^{-1} + (\ell/(\ell-1))^{-1} = 1$ applied to $\sum_{i=1}^{m} |x_i y_i|$, where $x_i = a_{i,1}$ and $y_i = \prod_{j=2}^{\ell} a_{ij}$.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

It follows from the above Hölder's inequality that, for every $z \in \mathbb{Z}_n$, the inner sum $Q(z,\lambda)$ on the right-hand side of (6.4) is bounded from above by

$$Q(z,\lambda) \leqslant \left(\sum_{y\in\mathbb{Z}_n} \exp(\lambda \mathbf{1}[z=\ell y])^\ell\right)^{1/\ell} \prod_{i=1}^{\ell-1}\left(\sum_{y\in\mathbb{Z}_n} \exp(\lambda\, r_{S,\ell-i}(z-iy))^\ell\right)^{1/\ell}.$$

Recalling that we require that $h!$ and n are co-prime and thus that each $i \in [\ell]$ is co-prime with n, it follows that, for fixed z and i, the map $y \mapsto z - iy$ is a permutation of $\mathbb{Z}_n$. In particular, in the rightmost sum above, we can substitute w for $z - iy$, simplifying the expression to

$$\sum_{w\in\mathbb{Z}_n} \exp(\lambda\, r_{S,\ell-i}(w))^\ell = E_{S,\ell-i}(\ell\lambda).$$

For the same reason, there exists only a single value $y \in \mathbb{Z}_m$ such that $z = \ell y$, and thus all but one term of the sum $\sum_{y\in\mathbb{Z}_n} \exp(\lambda \mathbf{1}[z=\ell y])^\ell$ are equal to 1, the only other term being equal to $e^{\ell\lambda}$. Consequently, we have

$$Q(z,\lambda) \leqslant \left((n + e^{\ell\lambda} - 1)\prod_{i=1}^{\ell-1} E_{S,\ell-i}(\ell\lambda)\right)^{1/\ell}.$$

Note that the bound we have obtained on $Q(z,\lambda)$ is independent of z, and hence (6.4) implies that

$$\sum_{y\in\mathbb{Z}_n} E_{S\cup\{y\},\ell}(\lambda) \leqslant E_{S,\ell}(\lambda)\left((n + e^{\ell\lambda} - 1)\prod_{i=1}^{\ell-1} E_{S,\ell-i}(\ell\lambda)\right)^{1/\ell}. \tag{6.6}$$

Observe that since $S \neq \emptyset$, then

$$E_{S,\ell-1}(\ell\lambda) = \max\{n + e^{\ell\lambda} - 1, E_{S,1}(\ell\lambda), \ldots, E_{S,\ell-1}(\ell\lambda)\}. \tag{6.7}$$

Indeed, fix some arbitrary $x \in S$ and notice that $r_{S,\ell-1}(z) \geqslant r_{S,\ell-1-i}(z - ix)$ for all $1 \leqslant i \leqslant \ell - 2$, and since $z \mapsto z - ix$ is a permutation of $\mathbb{Z}_n$, we have $E_{S,\ell-1}(\ell\lambda) \geqslant E_{S,\ell-1-i}(\ell\lambda)$. It is also clear that $E_{S,\ell-1}(\ell\lambda) \geqslant n + e^{\ell\lambda} - 1$ since, for all $x \in S$, we have $r_{S,\ell-1}((\ell-1)\cdot x) \geqslant 1$.

From (6.6) and (6.7) we conclude that for every non-empty S and all $\lambda > 0$,

$$\sum_{y\in\mathbb{Z}_n} E_{S\cup\{y\},\ell}(\lambda) \leqslant E_{S,\ell}(\lambda)\, E_{S,\ell-1}(\ell\lambda). \tag{6.8}$$

This gives inequality (6.2), and hence, the claim is proved. □

We now set

$$\lambda_\ell = \frac{h!\,\log(2n)}{\ell!\, g} \tag{6.9}$$

for each $\ell \in [h]$.

Definition 8. We shall call $y \in \mathbb{Z}_n \setminus S$ an *ε-good extension* of a set S if, for all $2 \leqslant \ell \leqslant h$,

$$E_{S\cup\{y\},\ell}(\lambda_\ell) \leqslant E_{S,\ell}(\lambda_\ell)\left(1 + \frac{2h}{\varepsilon}\,\frac{E_{S,\ell-1}(\lambda_{\ell-1}) - n}{n}\right). \tag{6.10}$$

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

Claim 6.2. *Let $h \geqslant 2$, $\varepsilon > 0$, and let n be a number relatively prime to $h!$. Moreover, suppose that $S \subset \mathbb{Z}_n$ satisfies $1 \leqslant |S| \leqslant \varepsilon n/6$. Then there are at least $(1 - 2\varepsilon/3)n$ elements $y \in \mathbb{Z}_n \setminus S$ that are ε-good extensions of S.*

Proof. Inequality (6.2) in Claim 6.1 implies that the number of elements $y \in \mathbb{Z}_n$ which satisfy

$$E_{S\cup\{y\},\ell}(\lambda_\ell) - E_{S,\ell}(\lambda_\ell) \geqslant \frac{2h}{\varepsilon} \cdot \frac{1}{n} E_{S,\ell}(\lambda_\ell)(E_{S,\ell-1}(\ell\lambda_\ell) - n)$$

is at most $\varepsilon n/(2h)$. Together with the fact that $\ell\lambda_\ell = \lambda_{\ell-1}$, we have that the number of $y \in \mathbb{Z}_n$ that violate (6.10) for a fixed ℓ is at most $\varepsilon n/(2h)$. Since ℓ can be any integer between 2 and h, there are at most $\varepsilon n/2$ violators. Recalling that $|S| \leqslant \varepsilon n/6$, we obtain that the number of $y \in \mathbb{Z}_n \setminus S$ that fail to be ε-good is at most $\varepsilon n/2 + \varepsilon n/6 = (2\varepsilon/3)n$. □

We are now in a position to prove Theorem 2.7.

Proof of Theorem 2.7. Fix $\varepsilon > 0$ and assume that $1 \leqslant m \leqslant (\varepsilon/3h)(n^{1-h!/g})^{1/h}$. We may and shall assume that $m \geqslant \log n$, since otherwise the random set $[n]_m$ is a.a.s. a B_h-set and we are done.

Let R be a random m-element subset of $\mathbb{Z}_n$. We construct a subset $S \subset R$ as follows. Let $(x_1, \dots, x_m)$ be a random ordering of the elements of R. Let $S_1 = \{x_1\}$, and for $1 < j \leqslant m$ let

$$S_j = \begin{cases} S_{j-1} \cup \{x_j\} & \text{if } x_j \text{ is an } \varepsilon\text{-good extension of } S_{j-1}, \\ S_{j-1} & \text{otherwise.} \end{cases}$$

We shall show that $S = S_m$ is a $B_h[g]$-set and that a.a.s. it has at least $(1 - \varepsilon)m$ elements. This will clearly suffice as $S_m \leqslant F_{h,g}(R)$ for all $m!$ orderings of R.

Claim 6.3. *The set $S = S_m$ is a $B_h[g]$-set.*

Proof. We shall first prove by induction that for every $1 \leqslant \ell \leqslant h$ and every $1 \leqslant j \leqslant m$, the following inequality holds:

$$\varphi(\ell, j) : E_{S_j,\ell}(\lambda_\ell) \leqslant n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} |S_j|^\ell.$$

Before proving this, let us show how $\varphi(h, m)$ implies this claim. By (6.13), for every $z \in \mathbb{Z}_n$, we have that

$$\exp(\lambda_h r_{S,h}(z)) \leqslant E_{S,h}(\lambda_h) \leqslant n + (2h/\varepsilon)^{h-1} e^{\lambda_1} m^h < 2n$$

since, recalling the definition of λ_1 in (6.9) and the assumption on m, we have

$$(2h/\varepsilon)^{h-1} e^{\lambda_1} m^h \leqslant (2h/\varepsilon)^{h-1} (2n)^{h!/g} (\varepsilon/3h)^h n^{1-h!/g} < n. \tag{6.11}$$

Consequently, from the definition of λ_h in (6.9), we have $r_{S,h}(z) \leqslant \lambda_h^{-1} \log(2n) = g$. In other words, S is a $B_h[g]$-set.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

We now resume the proof of the statements $\varphi(\ell, j)$ by induction. Observe that regardless of x_1, for every $\ell \in [h]$,

$$E_{S_1,\ell}(\lambda_\ell) = E_{\{x_1\},\ell}(\lambda_\ell) = (n-1) + e^{\lambda_\ell} \leqslant n + e^{\lambda_1},$$

and hence $\varphi(\ell, 1)$ holds for all ℓ.

Next, we consider $\varphi(1, j)$ for all j. Note that $r_{S,1}(z) = \mathbf{1}[z \in S]$ and therefore, from the definition of $E_{S,1}(\lambda)$, we have that

$$E_{S,1}(\lambda) = n - |S| + |S|e^{\lambda} = n + (e^{\lambda} - 1)|S|. \tag{6.12}$$

Hence, $\varphi(1, j)$ holds for all j.

Thus, it is enough to prove that if $\ell \geqslant 2$, then, assuming that $\varphi(\ell', j')$ holds for all pairs (ℓ', j') such that $\ell' < \ell$ or $j' < j$, the inequality $\varphi(\ell, j)$ is satisfied as well. If $S_j = S_{j-1}$, then there is nothing to show, and so we may assume that $S_j = S_{j-1} \cup \{x_j\}$, where x_j is an ε-good extension of S_{j-1}. In this case, letting $s = |S_{j-1}|$, recalling Definition 8, and invoking $\varphi(\ell, j-1)$ and $\varphi(\ell-1, j-1)$, we have

$$\begin{aligned}
E_{S_j,\ell}(\lambda_\ell) &\leqslant E_{S_{j-1},\ell}(\lambda_\ell)\left(1 + \frac{2h}{\varepsilon}\,\frac{E_{S_{j-1},\ell-1}(\lambda_{\ell-1}) - n}{n}\right) \\
&\leqslant (n + (2h/\varepsilon)^{\ell-1}e^{\lambda_1}s^{\ell})\left(1 + \frac{2h}{\varepsilon}\,\frac{(2h/\varepsilon)^{\ell-2}e^{\lambda_1}s^{\ell-1}}{n}\right) \\
&= n + (2h/\varepsilon)^{\ell-1}e^{\lambda_1}s^{\ell} + (2h/\varepsilon)^{\ell-1}e^{\lambda_1}s^{\ell-1} + \frac{(2h/\varepsilon)^{2\ell-2}e^{2\lambda_1}s^{2\ell-1}}{n} \\
&= n + (2h/\varepsilon)^{\ell-1}e^{\lambda_1}\left\{s^{\ell} + s^{\ell-1}\left(1 + \frac{(2h/\varepsilon)^{\ell-1}e^{\lambda_1}s^{\ell}}{n}\right)\right\}.
\end{aligned}$$

Since

$$(2h/\varepsilon)^{\ell-1}e^{\lambda_1}s^{\ell} \leqslant (2h/\varepsilon)^{h-1}e^{\lambda_1}m^{h} \overset{(6.11)}{\leqslant} n, \tag{6.13}$$

and $(s+1)^{\ell} \geqslant s^{\ell} + 2s^{\ell-1}$, it follows that

$$E_{S_j,\ell}(\lambda_\ell) \leqslant n + (2h/\varepsilon)^{\ell-1}e^{\lambda_1}(s+1)^{\ell},$$

thus proving the $\varphi(\ell, j)$ and concluding the induction step. □

Finally, we estimate the probability that $|S| < (1-\varepsilon)m$. If this is the case, then there are more than εm indices j for which x_j is not an ε-good extension of S_{j-1}. For each j, at least $(1 - 2\varepsilon/3)n$ elements of $\mathbb{Z}_n \setminus \{x_1, \ldots, x_{j-1}\}$ are ε-good extensions of S_{j-1}. Since x_j is a uniformly chosen random element of $\mathbb{Z}_n \setminus \{x_1, \ldots, x_{j-1}\}$, letting $\mathrm{Bin}(N, p)$ be a binomial random variable with parameters N and p, we have by Chernoff's bound

$$\mathbb{P}(|S| < (1-\varepsilon)m) \leqslant \mathbb{P}(\mathrm{Bin}(m, 1 - 2\varepsilon/3) < (1-\varepsilon)m) \leqslant \exp(-c_\varepsilon m)$$

for some constant $c_\varepsilon > 0$, and hence $|S| \geqslant (1-\varepsilon)m$ with probability $1 - o(1)$. This completes the proof of Theorem 2.7. □

We now derive Theorem 2.4 from Theorem 2.7 in the same way that we deduced Proposition 2.3(i) from Lemma 5.1.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

Proof of Theorem 2.4. Fix $\delta > 0$. Let $0 < \beta \leqslant 1/6$ be such that

$$(1-\beta)\left(1-\frac{\delta}{2}\right) \geqslant 1-\delta \quad \text{and} \quad \binom{(1+\beta)t}{\beta t} \leqslant \left(1+\frac{\delta}{2}\right)^t.$$

Now let $m = (1+\beta)t$, and note that we may suppose that $m \leqslant (\beta/6h)(n^{1-h!/g})^{1/h}$. It follows from Theorem 2.7 that $F_{h,g}([n]_m) \geqslant (1-\beta/2)m \geqslant t$ with probability at least $1-\beta$. We conclude that

$$\begin{aligned} |\mathcal{Z}_n^{h,g}(t)| &\geqslant (1-\beta)\binom{n}{(1+\beta)t}\binom{n-t}{\beta t}^{-1} = (1-\beta)\binom{n}{t}\binom{(1+\beta)t}{\beta t}^{-1} \\ &\geqslant (1-\beta)(1-\delta/2)^t\binom{n}{t} \geqslant (1-\delta)^t\binom{n}{t}, \end{aligned} \tag{6.14}$$

The lower bound in (2.7) follows. □

7. Concluding remarks

We close with two conjectures.

Conjecture 7.1. *Fix an integer $h \geqslant 3$ and $\varepsilon > 0$. For every $t \geqslant n^{1/(2h-1)+\varepsilon}$ and every large enough n, we have*

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{n}{t^{h-\varepsilon}}\right)^t. \tag{7.1}$$

Note that Proposition 2.3 implies that, if true, Conjecture 7.1 is basically optimal.

Conjecture 7.2. *Let $h \geqslant 3$ be an integer. Suppose $0 \leqslant a \leqslant 1$ is a fixed constant and $m = m(n) = (1+o(1))n^a$. Then a.a.s. $F_h([n]_m) = n^{b+o(1)}$, where $b = b_1(a)$ and $b_1(a)$ is as given in* (2.9).

It is worth mentioning that an argument following the lines of the proof of the upper bound in Theorem 2.5 shows that Conjecture 7.1 implies Conjecture 7.2. At the time of writing, we strongly believe that we are able to prove Conjecture 7.1 for $h = 3$.

Acknowledgement

The authors thank the referee for valuable comments and suggestions.

References

[1] Alon, N., Balogh, J., Morris, R. and Samotij, W. (2014) Counting sum-free sets in Abelian groups. *Israel J. Math.* **199** 309–344.
[2] Bose, R. C. and Chowla, S. (1962/1963) Theorems in the additive theory of numbers. *Comment. Math. Helv.* **37** 141–147.

https://doi.org/10.1017/S0963548315000206 Published online by Cambridge University Press

[3] Cameron, P. J. and Erdős, P. (1990) On the number of sets of integers with various properties. In *Number Theory* (Banff 1988), de Gruyter, pp. 61–79.
[4] Chen, S. (1994) On the size of finite Sidon sequences. *Proc. Amer. Math. Soc.* **121** 353–356.
[5] Chowla, S. (1944) Solution of a problem of Erdős and Turán in additive-number theory. *Proc. Nat. Acad. Sci. India. Sect. A* **14** 1–2.
[6] Cilleruelo, J. (2001) New upper bounds for finite B_h sequences. *Adv. Math.* **159** 1–17.
[7] Conlon, D. and Gowers, W. T. (2010) Combinatorial theorems in sparse random sets. Submitted.
[8] D'yachkov, A. G. and Rykov, V. V. (1984) B_s-sequences. *Mat. Zametki* **36** 593–601.
[9] Erdős, P. (1944) On a problem of Sidon in additive number theory and on some related problems: Addendum. *J. London Math. Soc.* **19** 208.
[10] Erdős, P. and Turán, P. (1941) On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.* **16** 212–215.
[11] Green, B. (2001) The number of squares and $B_h[g]$ sets. *Acta Arith.* **100** 365–390.
[12] Halberstam, H. and Roth, K. F. (1983) *Sequences*, second edition, Springer.
[13] Hardy, G. H., Littlewood, J. E. and Polya, G. (1934) *Inequalities*, Cambridge University Press.
[14] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience.
[15] Jia, X. D. (1993) On finite Sidon sequences. *J. Number Theory* **44** 84–92.
[16] Kleitman, D. J. and Wilson, D. B. (1996) On the number of graphs which lack small cycles. Unpublished Manuscript.
[17] Kleitman, D. J. and Winston, K. J. (1982) On the number of graphs without 4-cycles. *Discrete Math.* **41** 167–172.
[18] Kohayakawa, Y., Lee, S. J., Rödl, V. and Samotij, W. (2015) The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers. *Random Struct. Alg.* **46** 1–25.
[19] Kohayakawa, Y., Lee, S. and Rödl, V. (2011) The maximum size of a Sidon set contained in a sparse random set of integers. In *Proc. Twenty-Second Annual ACM–SIAM Symposium on Discrete Algorithms* (Philadelphia, PA), SIAM, pp. 159–171.
[20] Kohayakawa, Y., Łuczak, T. and Rödl, V. (1996) Arithmetic progressions of length three in subsets of a random set. *Acta Arith.* **75** 133–163.
[21] Kolountzakis, M. N. (1996) The density of $B_h[g]$ sequences and the minimum of dense cosine sums. *J. Number Theory* **56** 4–11.
[22] Krückeberg, F. (1961) B_2-Folgen und verwandte Zahlenfolgen. *J. Reine Angew. Math.* **206** 53–60.
[23] Lee, S. J. On Sidon sets in a random set of vectors. *J. Korean Math. Soc.*, to appear.
[24] Lindström, B. (1969) A remark on B_4-sequences. *J. Combin. Theory* **7** 276–277.
[25] O'Bryant, K. (2004) A complete annotated bibliography of work related to Sidon sequences. *Electron. J. Combin.* Dynamic Surveys **11** (electronic).
[26] Roth, K. F. (1953) On certain sets of integers. *J. London Math. Soc.* **28** 104–109.
[27] Saxton, D. and Thomason, A. (2015) Hypergraph containers, *Invent. Math.*, **201** 925–992
[28] Schacht, M. Extremal results for random discrete structures. Submitted.
[29] Shparlinskiĭ, I. E. (1986) On B_s-sequences. In *Combinatorial Analysis*, no. 7 (Russian), Moscow State University, pp. 42–45.
[30] Singer, J. (1938) A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.* **43** 377–385.
[31] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** 199–245.