ON THE NUMBER OF B_h-SETS

DOMINGOS DELLAMONICA JR., YOSHIHARU KOHAYAKAWA, SANG JUNE LEE, VOJTĚCH RÖDL, AND WOJCIECH SAMOTIJ

ABSTRACT. A set A of positive integers is a B_h*-set* if all the sums of the form $a_1+\cdots+a_h$, with $a_1,\ldots,a_h \in A$ and $a_1 \le \cdots \le a_h$, 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, better upper bounds for a problem of Cameron and Erdős (1990) in the context of B_h-sets are obtained. We 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.

1. INTRODUCTION

We deal with a natural extension of the concept of *Sidon sets*: For a positive integer $h \ge 2$, a set A of integers is called a B_h*-set* if all sums of the form $a_1+\cdots+a_h$ are distinct, where $a_i \in A$ and $a_1 \le \cdots \le a_h$. 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 of Chowla, Erdős, Singer, and Turán [5, 9, 10, 26] 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 $F_h(n) \ge (1+o(1))n^{1/h}$ for $h \ge 3$. On the other hand, an easy argument gives that for every $h \ge 3$,

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

Successively better bounds of the form $F_h(n) \le c_h n^{1/h}$ were given in [4, 6, 8, 14, 19, 20, 21, 25]. Currently, the best known upper bound on the constant c_h is given by Green [11], who proved that

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

where $o(1)\to 0$ as $h\to\infty$. The interested reader is referred to the classical monograph of Halberstam and Roth [12] and to a recent survey by O'Bryant [22] and the references therein.

We study two problems related to the classical problem of estimating $F_h(n)$. The first problem 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 [17, 18]. We present and discuss our results in detail in Section 2.

Our notation is standard. Let us remark that we use the notation $a \ll b$ 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

Date: 2013/10/07, 9:49am.
The second author was partially supported by FAPESP (2013/03447-6, 2013/07699-0), CNPq (308509/2007-2 and 477203/2012-4), NSF (DMS 1102086) and NUMEC/USP (Project MaCLinC/USP). The third author was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIP) (No. 2013042157). The fourth author was supported by the NSF grants DMS 0800070, 1301698, and 1102086. The fifth author was partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

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 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 $h \geq 2$. Recalling that $F_h(n)$ is the maximum size of a B_h-set contained in $[n]$, one trivially has

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

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

$$2^{(1+o(1))n^{1/h}} \leq |\mathcal{Z}_n^h| \leq n^{c'_h n^{1/h}}, \tag{2}$$

for some constant c'_h. We improve the upper bound on $|\mathcal{Z}_n^h|$ in (2) as follows.

Theorem 2.1. *For every $h \geq 2$, we have $|\mathcal{Z}_n^h| \leq 2^{C_h n^{1/h}}$, where C_h is a constant that depends only on h.*

The case $h = 2$ in Theorem 2.1 was established in [17] and later given another proof in [23]. The proof of Theorem 2.1 is based on 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 $t \geq n^{1/(h+1)}(\log n)^2$.

Theorem 2.2. *For every $h \geq 2$, there is a constant $c_h > 0$ such that, for any $t \geq n^{1/(h+1)}(\log n)^2$, we have*

$$|\mathcal{Z}_n^h(t)| \leq \left(\frac{c_h n}{t^h}\right)^t. \tag{3}$$

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 (4) in Proposition 2.3(i) below complements (3) in Theorem 2.2. On the other hand, Proposition 2.3(ii) shows that for small t, say, $t \ll n^{1/(2h-1)}$, the B_h-sets form a much larger proportion of the total number $\binom{n}{t}$ of t-element sets (see (5)). Note that, for large t, namely, $t \geq n^{1/(h+1)}(\log n)^2$, Theorem 2.2 tells us that this proportion is, very roughly speaking, of the order of $(n/t^h)\binom{n}{t}^{-1} \leq (n/t^h)^t/(n/t)^t = t^{-(h-1)t}$.

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

(i) *There is a constant $c'_h > 0$ such that*

$$|\mathcal{Z}_n^h(t)| \geq \left(\frac{c'_h n}{t^h}\right)^t. \tag{4}$$

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

$$|\mathcal{Z}_n^h(t)| \ge (1-\delta)^t \binom{n}{t}. \tag{5}$$

Let us compare the bounds we have for $|\mathcal{Z}_n^h(t)|$ as t varies. For $t \ll n^{1/(2h-1)}$, Proposition 2.3(*ii*) tells us that $|\mathcal{Z}_n^h(t)|$ is, up to a multiplicative factor of $(1-o(1))^t$, equal to the total number $\binom{n}{t}$ of t-element subsets of $[n]$. In this range, one might therefore say that B_h-sets are 'relatively abundant'. On the other hand, for $n^{1/(h+1)}(\log n)^2 \le t \ll n^{1/h}$, Theorem 2.2 and Proposition 2.3(*i*) determine $|\mathcal{Z}_n^h(t)|$ up to a multiplicative factor of the form c^t, and we see that the probability that a random t-element subset of $[n]$ is a B_h-set is roughly of the form $t^{-(h-1)t}$. In this second range, B_h-sets are therefore scarcer. Finally, note that, by (1), if $t > h^2 n^{1/h}$, we have $\mathcal{Z}_n^h(t) = \emptyset$, that is, there are no B_h-sets in this third range.

Note that, in the discussion above, we did not cover the whole range of t. In particular, we left open the interval $n^{1/(2h-1)} \le t \le n^{1/(h+1)}$. We believe that the hypothesis on t in Theorem 2.2 may be weakened to a bound comparable to the one in Proposition 2.3(*ii*). We make this precise in Conjecture 7.1, given in Section 7. If true, this conjecture implies that, roughly speaking, there is a sudden change of behaviour around $t_0 = n^{1/(2h-1)}$. Indeed, this conjecture implies that, for t considerably larger than this 'critical' value t_0, we have that $|\mathcal{Z}_n^h(t)|$ is of the form $\big(O(n/t^h)\big)^t$; this is in contrast to the fact that, as we have already seen, for t of smaller order than t_0, we have that $|\mathcal{Z}_n^h(t)|$ is of the form $(1-o(1))^t\binom{n}{t} = \big(\Theta(n/t)\big)^t$.

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

$$r_{S,h}(z) = \left|\left\{(a_1,\dots,a_h) \in S^h \colon a_1 + \dots + a_h = z \text{ and } a_1 \le \dots \le a_h\right\}\right|. \tag{6}$$

A set S is called a $B_h[g]$-*set* if $r_{S,h}(z) \le g$ 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

$$(1+o(1))n^{1/h} \le F_h(n) \le F_{h,g}(n) \le (gh \cdot h!)^{1/h} n^{1/h}. \tag{7}$$

Our final result in this section gives a lower bound for the number $Z_n^{h,g}(t)$ of $B_h[g]$-sets of cardinality t contained in $[n]$. We shall see that a bound of the form (5) in Proposition 2.3(*ii*) holds for $Z_n^{h,g}(t)$ even for t quite close to $n^{1/h}$, at least if $g = g(n) \to \infty$. This is somewhat surprising, as $Z_n^{h,g}(t) = 0$ if $t > g^{1/h} h^2 n^{1/h}$ (see (7)). Furthermore, note that, therefore, there are basically only two 'regimes' for $B_h[g]$-sets if $g \to \infty$, in contrast to the case of B_h-sets, for which we have identified three distinct regimes (B_h-sets are relatively abundant for small t (see (5)), rather scarce for intermediate t (see (3)) and non-existent for large t (see (1))).

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

$$(1-\delta)^t \binom{n}{t} \le Z_n^{h,g}(t) \le \binom{n}{t}. \tag{8}$$

The proof of Theorem 2.4 is given in Section 6.

2.2. **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 $[n]_m$.

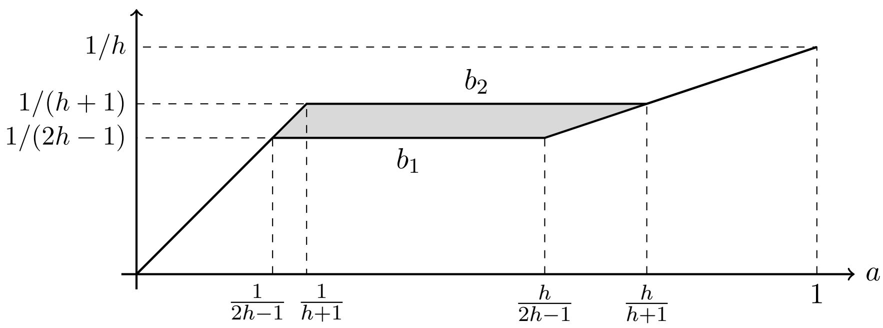

FIGURE 1. The graphs of $b_1 = b_1(a)$ and $b_2 = b_2(a)$ from the statement of Theorem 2.6

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

The asymptotic behavior of the random variable $F_2([n]_m)$ was investigated in [17, 18]. Our goal here is to study $F_h([n]_m)$ for arbitrary $h \geq 3$. 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 [24] 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 gives us quite precise information on $F_h([n]_m)$ for a wide range of m and non-trivial but looser bounds for $n^{1/(2h-1)} \leq m \leq n^{h/(h+1)}$; see also Figure 1.

Theorem 2.6. *Fix $h \geq 3$ and let $0 \leq a \leq 1$ be a fixed constant. Suppose $m = m(n) = (1 + o(1))n^a$. Then* ***a.a.s.***

$$n^{b_1+o(1)} \leq F_h([n]_m) \leq n^{b_2+o(1)}, \tag{9}$$

where

$$b_1(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/(2h-1); \\ 1/(2h-1), & \text{for } 1/(2h-1) \leq a \leq h/(2h-1); \\ a/h, & \text{for } h/(2h-1) \leq a \leq 1; \end{cases} \tag{10}$$

and

$$b_2(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/(h+1); \\ 1/(h+1), & \text{for } 1/(h+1) \leq a \leq h/(h+1); \\ a/h, & \text{for } h/(h+1) \leq a \leq 1. \end{cases} \tag{11}$$

We prove the upper bounds in Theorem 2.6 (that is, (9) and (11)) in Sections 3. The lower bounds (that is, (9) and (10)) are proved in Section 5. Theorem 2.6 determines $b = b(a)$ for which $F_h([n]_m) = n^{b+o(1)}$ when $m = (1+o(1))n^a$ whenever $a \leq 1/(2h-1)$ or $a \geq h/(h+1)$. An interesting open question is the existence and determination of $b = b(a)$ such that $F_h([n]_m) = n^{b+o(1)}$ for $1/(2h-1) \leq a \leq h/(h+1)$ (see Conjecture 7.2 in Section 7).

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

Definition 2.7. For integers $h \geq 2$ and $g \geq 1$ and a set R, denote by $F_{h,g}(R)$ the maximum size of a $B_h[g]$-set contained in R.

As a natural extension of Theorem 2.6, we investigate the random variable $F_{h,g}([n]_m)$. Trivially, one has

$$F_{h,g}([n]_m) \leq \min\{m, F_{h,g}(n)\}. \tag{12}$$

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

Theorem 2.8. *Let $h \geq 2$ be an integer and suppose $g(n) \to \infty$ as $n \to \infty$. Let $0 \leq a \leq 1$ be a fixed constant and suppose $m = m(n) = (1+o(1))n^a$. Then* ***a.a.s.***

$$F_{h,g}([n]_m) = n^{b+o(1)}, \tag{13}$$

where

$$b(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/h; \\ 1/h, & \text{for } 1/h \leq a \leq 1. \end{cases} \tag{14}$$

The upper bound on $F_{h,g}([n]_m)$ contained in Theorem 2.8 follows from (12). The lower bound follows from the following more precise result, which is proved in Section 6.

Theorem 2.9. *Fix an integer $h \geq 2$ and a function $g = g(n)$. For every fixed $\varepsilon > 0$ and $1 \leq m \leq (\varepsilon/3h)\left(n^{1-h!/g}\right)^{1/h}$, we* ***a.a.s.*** *have $F_{h,g}([n]_m) \geq (1-\varepsilon)m$.*

We remark that Theorem 2.9 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.

3. Proof of Theorem 2.1 and proof of the upper bounds in Theorem 2.6

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,

$$|\mathcal{Z}_n^h| \leq 2^{o(n^{1/h})} + \sum_{t \geq n^{1/(h+1)}(\log n)^2} \left(\frac{c_h n}{t^h}\right)^t. \tag{15}$$

Since the function $t \mapsto (c_h n/t^h)^t$ is maximized when $t = (c_h n)^{1/h}/e$, it follows from (15) that, for an appropriate choice of the constant C_h,

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

We now turn to the proof of the upper bound on $F_h([n]_m)$ contained in Theorem 2.6. We start with the following easy remark.

Remark 3.1. 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 [13, p. 17]).

The following theorem is a direct corollary of Theorem 2.2.

Theorem 3.2. *There is an absolute constant C such that for every $p \geq n^{-1/(h+1)}(\log n)^{2h}$,* ***a.a.s.****,*

$$F_h([n]_p) \leq C(pn)^{1/h}.$$

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

To derive Theorem 3.2 from Theorem 2.2, it suffices to use the following proposition.

Proposition 3.3. *The expected number of B_h-sets of cardinality t in $[n]_p$ is $p^t|\mathcal{Z}_n^h(t)|$. In particular,*

$$\mathbf{P}\bigl[F_h([n]_p) \geq t\bigr] \leq p^t|\mathcal{Z}_n^h(t)|.$$ □

We now prove the upper bound on $F_h([n]_m)$ given in Theorem 2.6 (see (9) and (11)). Let us first recall that Remark 3.1 links the binomial random set $[n]_p$, appearing in Theorem 3.2, to the random set $[n]_m$ that appears in Theorem 2.6. In what follows, we establish (9) and (11) in Theorem 2.6 using Theorem 3.2. We analyse three ranges of a separately.

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

(*ii*) $1/(h+1) < a \leq h/(h+1)$: 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 $b_2\bigl(h/(h+1)\bigr) = 1/(h+1)$, this monotonicity lets us take $b_2(a) = 1/(h+1)$ in this range of a.

(*iii*) $h/(h+1) < a \leq 1$: In this range, $b_2(a) = a/h$ follows from Theorem 3.2. Indeed, if $p \geq n^{-1/(h+1)}(\log n)^{2h}$, then with probability at least $1 - \exp\bigl(-c(pn)^{1/h}\bigr) \geq 1 - o\bigl(1/\sqrt{pn}\bigr)$ we have $F_h([n]_p) \leq C(pn)^{1/h}$ for some absolute constant $C > 0$. Remark 3.1 implies that, **a.a.s.**, $F_h([n]_m) \leq Cm^{1/3}$ for all $m \geq n^{h/(h+1)}(\log n)^{2h}$, giving that we may take $b_2(a) = a/3$ 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. 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 $s < s' < s'' < \dots$.

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 [16] and Kleitman and Wilson [15]. We start with the following simple observation. If two distinct elements $x, y \in [n] \setminus S$ satisfy

$$x + a_1 + \cdots + a_{h-1} = y + b_1 + \cdots + b_{h-1}$$
$$\text{for some } \{a_1, \ldots, a_{h-1}\}, \{b_1, \ldots, b_{h-1}\} \in \binom{S}{h-1}, \tag{16}$$

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

Definition 4.1. 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 (16).

Clearly, by the construction of CG_S, any set I of elements of $[n] \setminus S$ that extends S to a larger B_h-set $S \cup I$ must be an independent set in CG_S.

One of our main tools is the following lemma, implicit in the work of Kleitman and Winston [16], which provides an upper bound on the number of independent sets in graphs that have many edges in each sufficiently large vertex subset (see (18)). Lemma 4.2 in the version presented below is stated and proved in [17, 18], 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 4.2. *Let δ and $\beta > 0$ and $q \in \mathbb{N}$ be numbers satisfying*

$$e^{\beta q} \delta > 1. \tag{17}$$

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

$$e_G(A) \geq \beta\, |A|^2 \text{ for all } A \subset V \text{ with } |A| \geq \delta\, |V|. \tag{18}$$

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

$$\binom{|V|}{q}\binom{\delta |V|}{m} \tag{19}$$

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

Remark 4.3. When we apply Lemma 4.2 to CG_S, we shall take $m \gg q$ to take advantage of the upper bound (19). In condition (18), 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 (18) with good parameters β and δ. To that end, we shall make use of two auxiliary graphs, which we now define.

Definition 4.4. 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 $\big(\{a_1, \ldots, a_{h-1}\}, \{b_1, \ldots, b_{h-1}\}\big) \in \binom{S}{h-1}^2$ that satisfy (16).

It will be convenient for us to work with a certain subgraph of $\widetilde{\mathrm{CG}}_S$ that we define as follows. For a set S with s elements, let

$$S_1, \ldots, S_{h-1} \tag{20}$$

be a fixed partition of S into sets with cardinalities that differ by at most one. Let $\widetilde{\mathrm{CG}}'_S$ be the subgraph of $\widetilde{\mathrm{CG}}_S$ in which the multiplicity of a pair $x, y \in [n] \setminus S$ is the number of pairs

$\big(\{a_1,\dots,a_{h-1}\},\{b_1,\dots,b_{h-1}\}\big)\in\binom{S}{h-1}^2$ that satisfy (16) and, moreover, are such that $a_i,b_i\in S_i$ for each $i\in[h-1]$.

Lemma 4.5. *For every B_h-set S with s elements and every $A\subset[n]\setminus S$ with $|A|\ge h^{2h}n/s^{h-1}$, we have*

$$e_{\widetilde{\mathrm{CG}}_S}(A)\ge e_{\widetilde{\mathrm{CG}}'_S}(A)\ge\frac{s^{2h-2}}{h^{2h}n}|A|^2, \tag{21}$$

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

The proof of Lemma 4.5 will be given in Section 4.3. In view of Lemma 4.5, if the maximal multiplicity of an edge in $\widetilde{\mathrm{CG}}'_S$ is at most r, then the graph CG_S satisfies the conditions of Lemma 4.2 with $\beta=s^{2h-2}/h^{2h}rn$ and $\delta=h^{2h}/s^{h-1}$. Consequently, we are interested in bounding the multiplicity of the edges of $\widetilde{\mathrm{CG}}'_S$.

Proposition 4.6. *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 4.6 to Section 4.4. The following is an immediate corollary of Lemma 4.5 and Proposition 4.6.

Corollary 4.7. *If S is a B_h-set with s elements, then for every $A\subset[n]\setminus S$ with $|A|\ge h^{2h}n/s^{h-1}$,*

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

4.2. **Proof of Theorem 2.2.** The case $h=2$ of Theorem 2.2 is proved in [17] and we therefore restrict ourselves to $h\ge3$ here. We shall in fact prove the following: for every $h\ge3$ and $t\ge h^2n^{1/(h+1)}(\log n)^{1+1/(h+1)}$,

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

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

$$t\le h^2n^{1/h}, \tag{22}$$

that is, $h^2n^{1/(h+1)}(\log n)^{1+1/(h+1)}\le t\le h^2n^{1/h}$. Let $s_0=h^2(n\log n)^{1/(h+1)}$ and let K be the largest integer satisfying $t2^{-K}\ge 2s_0$. We define three sequences $(s_k)_{0\le k\le K}$, $(q_k)_{0\le k\le K}$ and $(m_k)_{0\le k\le K}$ as follows. We let $q_0=s_0/2$ and $m_0=t2^{-K}-s_0-q_0$. Moreover, we let $s_1=t2^{-K}\ge 2s_0$, $q_1=q_0/2^h$ and $m_1=t2^{-K+1}-s_1-q_1$. For $k=2,\dots,K$, we let $s_k=2s_{k-1}=t2^{-K+k-1}$, $q_k=q_{k-1}/2^h=q_02^{-hk}$ and $m_k=t2^{-K+k}-s_k-q_k$.

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 4.2 to obtain a non-trivial bound on the number of extensions of S_k to S_{k+1} for all k.

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 4.7, the graph CG_{S_k} is such that for all $A\subset[n]\setminus S_k$ with $|A|\ge h^{2h}n/s_k^{h-1}$,

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

Let $\delta_k = h^{2h}/s_k^{h-1} \geq 1/n$ and observe that

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

Consequently, CG_{S_k}, δ_k, β_k and q_k satisfy the conditions of Lemma 4.2. 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 4.2, the number of extensions of S_k into a B_h-set S_{k+1} is at most $\binom{n}{q_k}\binom{\delta_k n}{m_k}$. Note that

$$\binom{\delta_0 n}{m_0} \leq \binom{\delta_0 n}{3s_0} \quad \text{and} \quad \binom{\delta_k n}{m_k} \leq \binom{\delta_k n}{s_k}$$

for all $1 \leq k \leq K$. Indeed, we have that $m_0 = s_1 - s_0 - q_0 \leq 4s_0 - s_0 \leq 3s_0$ and also $3s_0 \leq \frac{\delta_0 n}{2}$ and that for all $1 \leq k \leq K$, $m_k \leq s_k \leq \frac{\delta_k n}{2}$ as

$$\frac{s_k}{\delta_k} = \frac{s_k^h}{h^{2h}} \leq \frac{s_K^h}{h^{2h}} = \frac{(t/2)^h}{h^{2h}} \leq \frac{n}{2^h},$$

where the last inequality follows from our assumption on t. Hence,

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

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

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

It follows that

$$|\mathcal{Z}_n^h(t)| \leq \binom{n}{s_0}\prod_{k=0}^{K}\binom{n}{q_k}\binom{\delta_k n}{m_k} \leq n^{4s_0 + \sum_{k=0}^K q_k}\prod_{k=1}^{K}\left(\frac{eh^{2h}n}{s_k^h}\right)^{s_k}. \tag{23}$$

Finally, since

$$\sum_{k=0}^{K} q_k = q_0 \sum_{k=0}^{K} 2^{-hk} \leq 2q_0 = s_0 \leq \frac{t}{\log n}$$

and

$$\prod_{k=1}^{K}\left(\frac{eh^{2h}n}{s_k^h}\right)^{s_k} \leq \prod_{k=1}^{K+1}\left(\frac{eh^{2h}n}{(t2^{-k})^h}\right)^{t2^{-k}} \leq \left[\left(\frac{eh^{2h}n}{t^h}\right)^{\sum_{k\geq 1}2^{-k}} \cdot 2^{h\sum_{k\geq 1}k2^{-k}}\right]^t \leq \left(\frac{2^{2h}eh^{2h}n}{t^h}\right)^t,$$

Theorem 2.2 follows from (23). □

4.3. **Proof of Lemma 4.5.** Let S be a B_h-set with s elements and let $S_1, \ldots, S_{h-1}$ be the partition (20) of S from the definition of $\widetilde{\mathrm{CG}}'_S$. Let $A \subset [n] \setminus S$ be an arbitrary subset with $|A| \geq h^{2h}n/s^{h-1}$. 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

$$E(\Gamma) = \big\{(x,u) \in A \times [hn] \colon u = x + a_1 + \cdots + a_{h-1} \text{ for some } a_1 \in S_1, \ldots, a_{h-1} \in S_{h-1}\big\}.$$

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 $a_1 \in S_1, \ldots a_{h-1} \in S_{h-1}$. 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 $\big(\{a_1,\ldots,a_{h-1}\},\{b_1,\ldots,b_{h-1}\}\big) \in \binom{S}{h-1}^2$

with $a_i, b_i \in S_i$ for all $i \in [h-1]$ that satisfy (16) 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 $f(\alpha) = \binom{\alpha}{2} = \alpha(\alpha-1)/2$,

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

On the other hand, since $|A| \geq h^{2h} n / s^{h-1}$, we may assume that $s \geq h^2$ and hence,

$$e(\Gamma) = \sum_{x \in A} \deg_\Gamma(x) = |A||S_1| \ldots |S_{h-1}| \geq \left(\left\lfloor \frac{s}{h-1} \right\rfloor \right)^{h-1} |A| \geq \left(\frac{s}{h} \right)^{h-1} |A|.$$

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

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

This concludes the proof of Lemma 4.5. □

4.4. **Proof of Proposition 4.6.** Let S be a B_h-set of cardinality s and let $S_1, \ldots, S_{h-1}$ be the partition (20) of S from the definition of $\widetilde{\mathrm{CG}}'_S$. For each pair $i, j \in [h]$ with $i \leq j$ and each $x \in \mathbb{Z}$, let

$$N_i^j(x) = \big\{ x + a_i + \cdots + a_{j-1} \colon a_i \in S_i, \ldots, a_{j-1} \in S_{j-1} \big\},$$

where $N_i^i(x) = \{x\}$, and note that (since S is a B_h-set) the multiplicity of an edge $\{x, y\}$ in the multigraph $\widetilde{\mathrm{CG}}'_S$ is $|N_1^h(x) \cap N_1^h(y)|$. The following claim implies the postulated bound on the multiplicity of $\{x, y\}$, as trivially $x \in N_1^1(x) \setminus N_1^1(y)$.

Claim 4.8. *Fix x and $y \in \mathbb{Z}$ with $x \neq y$. For every $i \in [h]$, and every $z \in N_1^i(x) \setminus N_1^i(y)$,*

$$\big| N_i^h(z) \cap N_1^h(y) \big| \leq s^{h-i-1}. \tag{24}$$

Proof. We prove the claim by induction on $h - i$. If $i = h$, then there is nothing to prove as $N_h^h(z) = \{z\}$ is disjoint from $N_1^h(y)$. Assume then that $i < h$ and let z be an arbitrary element of $N_1^i(x) \setminus N_1^i(y)$. If $N_i^{i+1}(z) \cap N_1^{i+1}(y) = \emptyset$, then, as $N_i^{i+1}(z) \subset N_1^{i+1}(x)$, the induction assumption implies that

$$\begin{aligned} \big| N_i^h(z) \cap N_1^h(y) \big| &\leq \sum_{u \in N_i^{i+1}(z)} \big| N_{i+1}^h(u) \cap N_1^h(y) \big| \\ &\leq \big| N_i^{i+1}(z) \big| \cdot s^{h-i-2} = |S_i| \cdot s^{h-i-2} \leq s^{h-i-1}. \end{aligned}$$

Otherwise, there is a $u \in N_i^{i+1}(z) \cap N_i^{i+1}(y)$. If $N_{i+1}^h(u') \cap N_1^h(y) = \emptyset$ for all $u' \in N_i^{i+1}(z) \setminus \{u\}$, then

$$\big| N_i^h(z) \cap N_1^h(y) \big| = \big| N_{i+1}^h(u) \cap N_1^h(y) \big| \leq \big| N_{i+1}^h(u) \big| \leq |S_{i+1}| \cdots |S_{h-1}| \leq s^{h-i-1}.$$

Hence, we may assume that there is a $u' \in N_i^{i+1}(z) \setminus \{u\}$ such that $N_{i+1}^h(u') \cap N_1^h(y) \neq \emptyset$. In this case, let $j \in \{i, \ldots, h-1\}$ be the smallest index such that $N_{i+1}^{j+1}(u') \cap N_1^{j+1}(y) \neq \emptyset$ and let $w \in N_{i+1}^{j+1}(u') \cap N_1^{j+1}(y)$ be arbitrary. Moreover, let $k \in \{1, \ldots, i\}$ be the largest index such that there is a $w' \in N_1^k(y)$ satisfying $u \in N_k^i(w')$ and $w \in N_k^{j+1}(w')$. Observe that

$$\begin{aligned} u &= w' + a_k + \cdots + a_i && \text{for some } a_k \in S_k, \ldots, a_i \in S_i, \\ w &= z + b_i + \cdots + b_j && \text{for some } b_i \in S_i, \ldots, b_j \in S_j, \\ w &= w' + c_k + \cdots + c_j && \text{for some } c_k \in S_k, \ldots, c_j \in S_j, \\ u &= z + d && \text{for some } d \in S_i. \end{aligned}$$

Moreover, the minimality of j implies that $b_j \neq c_j$ and the maximality of k implies that $a_k \neq c_k$. Also, since $b_i = u' - z$ and $u' \neq u$, then $b_i \neq d$. It follows that

$$a_k + \cdots + a_i + b_i + \cdots + b_j = c_k + \cdots + c_j + d.$$

Since S is a B_h-set and $j - k + 2 \leq h$, we must have

$$\{a_k, \ldots, a_i, b_i, \ldots, b_j\} = \{c_k, \ldots, c_j, d\}. \tag{25}$$

Recall that the sets $S_1, \ldots, S_{h-1}$ are pairwise disjoint. If $j > i$, then $b_j \neq c_j$ are the only elements of S_j in (25) and hence (25) cannot hold. If $k = j = i$, then (25) cannot hold as $b_i \notin \{c_i, d\}$. Therefore, it must be that $k < i$. But in this case, as $a_k \neq c_k$ are the only elements of S_k, equality (25) again cannot hold. This contradiction completes the proof of the claim. □

5. Lower bounds

In this section, we establish the lower bounds in Theorem 2.6 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 2.6, for $0 \leq a \leq 1/(2h-1)$, one may take $b_1(a) = a$ (see (9) and (10)). Since $F_3([n]_m)$ is non-decreasing in probability with respect to m, for $a > 1/(2h-1)$, we may take $b_1(a) = b_1\big(1/(2h-1)\big) = 1/(2h-1)$. Moreover, as an easy corollary of Lemma 5.1, we will also derive Proposition 2.3(ii).

In the second part of this section, following the strategy of [17, 18], 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(i). Moreover, we shall show that if $1 \ll m \leq n$, 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 2.6, we may take $b_1(a) = a/h$ for all $0 \leq a \leq 1$. Note that, in the range $1/(2h-1) \leq a \leq h/(2h-1)$, this is superseded by the bound obtained in the first part, that is, $b_1(a) = 1/(2h-1)$.

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

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

$$a_1 + \cdots + a_h = b_1 + \cdots + b_h \quad \text{with} \quad \{a_1, \ldots, a_h\} \neq \{b_1, \ldots, b_h\} \tag{26}$$

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

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

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 (26) yields a B_h-set, the lemma follows. $\square$

Proof of Proposition 2.3(ii). Fix a constant $\delta > 0$. Choose $\beta > 0$ small enough so that $(1 - 2\beta)(1-\delta/3) \geq 1-\delta$ and $\binom{(1+\beta)t}{\beta t} \leq (1+\delta/3)^t$ for all t. Let $\varepsilon > 0$ be a small constant. Assume that $t \leq \varepsilon n^{1/(2h-1)}$. Lemma 5.1 with $m = (1+\beta)t$ implies that if ε is sufficiently small, then $F_h([n]_m) \geq t$ with probability at least $1-\beta$. It follows that, for large enough n, we have

$$|\mathcal{Z}_n^h(t)| \geq (1-\beta)\binom{n}{(1+\beta)t}\binom{n}{\beta t}^{-1} \geq (1-2\beta)\binom{n}{(1+\beta)t}\binom{n-t}{\beta t}^{-1}$$
$$= (1-2\beta)\binom{n}{t}\binom{(1+\beta)t}{\beta t}^{-1} \geq (1-2\beta)(1-\delta/3)^t\binom{n}{t} \geq (1-\delta)^t\binom{n}{t}, \tag{27}$$

as required. $\square$

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. *For every integer $h \geq 2$, there is an integer m_h such that for all $m \geq m_h$, there exists a B_h-set $Y \subset \mathbb{Z}_m$ with $|Y| = \Omega(m^{1/h})$.* $\square$

Let us now fix some n and m with $n \geq m$ such that, letting $p = m/n$, the numbers $1/(hp)$ and pn/h are integers. Theorem 5.2 implies the existence of a B_h-set $Y \subset \mathbb{Z}_m$ with $|Y| = \Omega(m^{1/h})$, provided that m is sufficiently large. We will show that there is a subset $U \subset [n]$ and a projection $\pi\colon U \subset [n] \to \mathbb{Z}_m$ such that

(*a*) any set $S \subset \pi^{-1}(Y)$ with $|S \cap \pi^{-1}(x)| \leq 1$ for all $x \in Y$ is a B_h-set;
(*b*) $|\pi^{-1}(x)| \geq 1/(hp)$ for $s = \Omega(|Y|)$ elements $x \in Y$.

We first show that the existence of π and U satisfying conditions (*a*) and (*b*) above implies Proposition 2.3(*i*).

Proof of Proposition 2.3(i). Note that, choosing c'_h appropriately small (see (4)), we may suppose that $t \leq \varepsilon n^{1/h}$ for any given $\varepsilon > 0$. Therefore, let us assume that $t \leq \varepsilon n^{1/h}$ for a suitably small constant ε for our estimates below to hold. Choose $m = O(t^h) \leq n$ so that $s = \Omega(|Y|) = \Omega(m^{1/h})$ in condition (*b*) is at least t. Let $Y' \subset Y$ be a set of t numbers x such that $|\pi^{-1}(x)| \geq 1/(hp)$ for each $x \in Y'$. Condition (*a*) implies that each set $T \subset \pi^{-1}(Y') \subset [n]$ satisfying $|T \cap \pi^{-1}(x)| = 1$ for every $x \in Y'$ is a B_h-set. Since $m = O(t^h)$, we have $|\pi^{-1}(x)| \geq 1/(hp) = n/(hm) = \Omega(n/t^h)$, and hence there are $\big(\Omega(n/t^h)\big)^t$ such sets T, proving the bound in (4). $\square$

Next, we show that the existence of π and U as above also yields the claimed lower bound in Theorem 2.6.

Lemma 5.3. *For any $1 \ll m \leq n$, we* **a.a.s.** *have $F_h([n]_m) = \Omega(m^{1/h})$.*

Proof. In the view of Lemma 5.1, we may assume that $m \gg n^{1/(2h)}$. It will be convenient for us to use the model $[n]_p$ with $p = m/n$ rather than $[n]_m$ (recall Remark 3.1). Without loss of generality we assume that n is sufficiently large and that $1/(hp)$, pn, $pn/h \in \mathbb{N}$. Fix some π and U satisfying conditions (*a*) and (*b*) above. Define a set S by selecting the smallest element from $[n]_p \cap \pi^{-1}(x)$ for each $x \in Y$, whenever this set is non-empty. By (a), the set S is a B_h-set. It suffices to show that **a.a.s.** $|S| = \Omega(m^{1/h})$.

Using (b), let $Y' \subset Y$ be a family of $s = \Omega(|Y|) = \Omega(m^{1/h})$ elements $x \in Y$ satisfying $|\pi^{-1}(x)| \geq 1/(hp)$. For any $x \in Y'$, the probability that $[n]_p \cap \pi^{-1}(x) = \emptyset$ is $q = (1-p)^{|\pi^{-1}(x)|} \leq (1-p)^{1/(hp)} \leq e^{-p/(hp)} = e^{-1/h} < 1$. It follows from the fact that the sets $\{\pi^{-1}(x)\}_{x \in Y'}$ are disjoint that the number of elements $x \in Y'$ for which $[n]_p \cap \pi^{-1}(x) = \emptyset$ is a random variable following the binomial distribution with parameters $|Y'|$ and $q < 1$. Consequently, by the Chernoff's bound,

$$\mathbf{P}\left[\left|\{x \in Y \colon [n]_p \cap \pi^{-1}(x) \neq \emptyset\}\right| < \frac{1-q}{2}|Y'|\right] \leq \exp\{-c\,|Y|\},$$

for some constant $c > 0$. Therefore, with probability at least $1 - \exp\big(-\Omega(m^{1/h})\big)$ there are at least $\frac{1-q}{2}|Y'|$ elements $x \in Y$ which satisfy $[n]_p \cap \pi^{-1}(x) \neq \emptyset$, thus proving that **a.a.s.** $F_h([n]_m) \geq \Omega\big(m^{1/h}\big)$. $\square$

Finally, we define the projection π and its domain $U \subset [n]$. We first partition $[hn]$ into intervals

$$I_j = \left[\frac{j}{p} + 1, \frac{j+1}{p}\right], \quad j = 0, \ldots, hpn - 1.$$

Furthermore, we subdivide each of the intervals above into h subintervals of equal lengths, namely,

$$I_{j,k} = \left[\frac{j}{p} + 1 + \frac{k}{hp}, \frac{j}{p} + \frac{k+1}{hp}\right], \quad j = 0, \ldots, hpn - 1 \text{ and } k = 0, \ldots, h-1. \tag{28}$$

The domain of π is defined as

$$U = \bigcup_{j=0}^{pn-1} I_{j,0}. \tag{29}$$

Note that $U \subset [n]$ since $j < pn$ in the union above. The projection π is then defined by $\pi(x) = j \in \mathbb{Z}_{pn}$ whenever $x \in I_{j,0}$. Clearly, condition (b) is satisfied.

Let us now prove that condition (a) is satisfied. Let $S \subset \pi^{-1}(Y)$ be a set satisfying $|S \cap \pi^{-1}(x)| \leq 1$ for all $x \in Y$. This ensures that $\pi|_S$ is a one-to-one map. Moreover, $\pi(S) \subset Y$ is a B_h-set. Let $(a_1, \ldots, a_h)$ be an arbitrary h-tuple such that $a_1, \ldots, a_h \in S$ with $a_1 \leq \cdots \leq a_h$ and let $0 \leq \ell \leq hpn - 1$ be such that $a_1 + \cdots + a_h \in I_\ell$. We claim that $\pi(a_1) + \cdots + \pi(a_h) = \ell \mod pn$. Indeed, for each $i \in [h]$, let j_i be such that $a_i \in I_{j_i,0}$ and observe that by (28), we have $a_i \in \left[\frac{j_i}{p} + 1, \frac{j_i}{p} + \frac{1}{hp}\right]$. Therefore,

$$a_1 + \cdots + a_h \in \left[\frac{j_1 + \cdots + j_h}{p} + h, \frac{j_1 + \cdots + j_h}{p} + h \times \frac{1}{hp}\right] \subset I_{j_1 + \cdots + j_h}.$$

Hence $\ell = j_1 + \cdots + j_h$ and since $\pi(a_i) = j_i \mod pn$, it follows that $\pi(a_1) + \cdots + \pi(a_h) = \ell \mod pn$. Since $\pi(S)$ is a B_h-set and $\pi|_S$ is one-to-one, it follows that no other h-tuple $(b_1, \ldots, b_h)$ with $b_1, \ldots, b_h \in S$ and $b_1 \leq \cdots \leq b_h$ can satisfy $\pi(b_1) + \cdots + \pi(b_h) = \ell \mod pn$. In other words, no other h-tuple $(b_1, \ldots, b_h)$ satisfies $b_1 + \cdots + b_h \in I_\ell$ and hence S must be a B_h-set.

6. Proofs of Theorems 2.4 and 2.9

We need some preparations for the proofs of Theorems 2.4 and 2.9. For the remainder of this section, we fix an integer $h \geq 2$ 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 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 (6) in Section 2.2). For every $1 \le \ell \le h$ and $\lambda > 0$ and $S \subset \mathbb{Z}_n$, let

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

Note that $r_{S,1}(z) = \mathbf{1}[z \in S]$ and therefore

$$E_{S,1}(\lambda) = n - |S| + |S|e^{\lambda} = n + (e^{\lambda} - 1)|S|. \tag{30}$$

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. *With the assumptions above, for any $S \neq \emptyset$, we have*

$$\mathbf{E}_{y \in \mathbb{Z}_n}\bigl[E_{S\cup\{y\},\ell}(\lambda) - E_{S,\ell}(\lambda)\bigr] \le \frac{1}{n} E_{S,\ell}(\lambda)\left(E_{S,\ell-1}(\ell\lambda) - n\right). \tag{31}$$

Proof. Note first that

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

Hence,

$$\sum_{y\in\mathbb{Z}_n} E_{S\cup\{y\},\ell}(\lambda) \le \sum_{z\in\mathbb{Z}_n}\Bigl[\exp\bigl(\lambda\, r_{S,\ell}(z)\bigr) \sum_{y\in\mathbb{Z}_n} \exp\bigl(\lambda \mathbf{1}[z=\ell y]\bigr) \prod_{i=1}^{\ell-1} \exp\bigl(\lambda\, r_{S,\ell-i}(z-iy)\bigr)\Bigr].$$

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

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

Consequently, recalling that we suppose that $h!$ and n are co-prime and thus that each $i \in [\ell]$ is co-prime with n, we have

$$\sum_{y\in\mathbb{Z}_n} E_{S\cup\{y\},\ell}(\lambda) \le E_{S,\ell}(\lambda)\left((n + e^{\ell\lambda} - 1)\prod_{i=i}^{\ell-1} E_{S,\ell-i}(\ell\lambda)\right)^{1/\ell}. \tag{32}$$

Observe that if $S \neq \emptyset$, then for all $\ell \ge \ell'$,

$$E_{S,\ell}(\lambda) \ge E_{S,\ell'}(\lambda) \ge n + e^{\lambda} - 1. \tag{33}$$

To see this, note that for every $\ell \in [h-1]$, every $x \in S$, and every $z \in \mathbb{Z}_n$, we have $r_{S,\ell+1}(z) \ge r_{S,\ell}(z-x)$. Inequalities (32) and (33) imply that for every non-empty S and all $\lambda > 0$,

$$\sum_{y\in\mathbb{Z}_n} E_{S\cup\{y\},\ell}(\lambda) \le E_{S,\ell}(\lambda)\, E_{S,\ell-1}(\ell\lambda). \tag{34}$$

Inequality (31) follows from (34) and the claim is proved. □

We now set
$$\lambda_\ell = \frac{h!\,\log(2n)}{\ell!\,g}$$
for each $\ell \in [h]$. We shall call $y \in \mathbb{Z}_n \setminus S$ a *good extension* of a set S if for all $2 \le \ell \le h$,
$$E_{S\cup\{y\},\ell}(\lambda_\ell) \le E_{S,\ell}(\lambda_\ell)\left(1 + \frac{2h}{\varepsilon}\,\frac{E_{S,\ell-1}(\lambda_{\ell-1}) - n}{n}\right). \tag{35}$$

Claim 6.2. *With the assumptions above, for any $S \neq \emptyset$ with $|S| \le \varepsilon n/6$, at least $(1-2\varepsilon/3)n$ elements $y \in \mathbb{Z}_n$ are good extensions of S.*

Proof. Inequality (31) in Claim 6.1 and Markov's inequality, together with the fact that $\ell\lambda_\ell = \lambda_{\ell-1}$, tell us that the number of $y \in \mathbb{Z}_n$ that violate (35) is at most $(\varepsilon/2h)n$. Summing over all ℓ and recalling that $|S| \le \varepsilon n/6$, we obtain that the number of $y \in \mathbb{Z}_n$ that fail to be good is at most $(2\varepsilon/3)n$. □

We are now in position to prove Theorem 2.9.

Proof of Theorem 2.9. Fix $\varepsilon > 0$ and assume that $1 \le m \le (\varepsilon/3h)\big(n^{1-h!/g}\big)^{1/h}$. We may and shall assume that $m \ge \log n$, since otherwise the random set $[n]_m$ is **a.a.s.** a B_h-set and we are done. Therefore, we have $m \to \infty$.

Let $R = (x_1, \ldots, x_m)$ be an ordered random subset of $\mathbb{Z}_n$. We construct a subset $S \subset R$ as follows. Let $S_1 = \{x_1\}$ and for $1 < j \le m$, let
$$S_j = \begin{cases} S_{j-1} \cup \{x_j\}, & \text{if } x_j \text{ is a good extension of } S_{j-1};\\ S_{j-1}, & \text{otherwise.}\end{cases}$$
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.

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

Proof. We shall first prove by induction that for every $1 \le \ell \le h$ and every $1 \le j \le m$, the following inequality holds
$$\varphi(\ell, j)\colon \quad E_{S_j,\ell}(\lambda_\ell) \le n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} |S_j|^\ell.$$
Observe that regardless of x_1, for every $\ell \in [h]$,
$$E_{S_1,\ell}(\lambda_\ell) = E_{\{x_1\},\ell}(\lambda_\ell) = (n-1) + e^{\lambda_\ell} \le n + e^{\lambda_1}$$
and hence $\varphi(\ell,1)$ holds for all ℓ. Moreover, it follows from (30) that $\varphi(1,j)$ holds for all j. Thus, it is enough to prove that if $\ell \ge 2$, 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 a good extension of S_{j-1}. In this case, letting $s = |S_{j-1}|$, we have
$$\begin{aligned}
E_{S_j,\ell}(\lambda_\ell) &\le E_{S_{j-1},\ell}(\lambda_\ell)\left(1 + \frac{2h}{\varepsilon}\,\frac{E_{S_{j-1},\ell-1}(\lambda_{\ell-1}) - n}{n}\right)\\
&\le \left(n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell\right)\left(1 + \frac{2h}{\varepsilon}\,\frac{(2h/\varepsilon)^{\ell-2} e^{\lambda_1} s^{\ell-1}}{n}\right)\\
&= n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^{\ell-1} + \frac{(2h/\varepsilon)^{2\ell-3} e^{2\lambda_1} s^{2\ell-1}}{n}\\
&\le n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} (s+1)^\ell.
\end{aligned}$$

To see the last inequality above, note that $(s+1)^\ell \ge s^\ell + 2s^{\ell-1}$ and that

$$(2h/\varepsilon)^{\ell-1} s^\ell e^{\lambda_1} \le (2h/\varepsilon)^{h-1} m^h e^{\lambda_1} \le n, \tag{36}$$

since $(2hm/\varepsilon)^h \le n^{1-h!/g} \le e^{-\lambda_1} n$.

In particular, $\varphi(h,m)$ holds and therefore, by (36), for every $z \in S$,

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

and hence $r_{S,h}(z) \le \lambda_h^{-1} \log(2n) = g$. In other words, S is a $B_h[g]$-set. □

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 a good extension of S_{j-1}. For each j, at least $(1-2\varepsilon/3)n$ elements of $\mathbb{Z}_n \setminus \{x_1, \dots, x_{j-1}\}$ are good extensions of S_{j-1}. Since x_j is a uniformly chosen random element of $\mathbb{Z}_n \setminus \{x_1, \dots, x_{j-1}\}$, letting $\mathrm{Bin}(N,p)$ be a binomial random variable with parameters N and p, we have

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

for some constant $c_\varepsilon > 0$, and hence $|S| \ge (1-\varepsilon)m$ with probability $1-o(1)$. This completes the proof of Theorem 2.9. □

We now derive Theorem 2.4 from Theorem 2.9 in the same way that we deduced Proposition 2.3(ii) from Lemma 5.1.

Proof of Theorem 2.4. Fix $\delta > 0$. Let $0 < \beta \le 1/6$ be such that $(1-2\beta)(1-\delta/3) \ge 1-\delta$ and $\binom{(1+\beta)t}{\beta t} \le (1+\delta/3)^t$. Now let $m = (1+\beta)t$, and note that we may suppose that $m \le (\beta/6h)\left(n^{1-h!/g}\right)^{1/h}$. It follows from Theorem 2.9 that $F_{h,g}([n]_m) \ge (1-\beta/2)m \ge t$ with probability at least $1-\beta$. We conclude that

$$Z_n^{h,g}(t) \ge (1-\beta)\binom{n}{(1+\beta)t}\binom{n}{\beta t}^{-1}. \tag{37}$$

The lower bound in (8) follows from (37) by the calculations given in (27). □

7. Concluding remarks

We close with two conjectures.

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

$$|\mathcal{Z}_n^h(t)| \le \left(\frac{n}{t^{h-\varepsilon}}\right)^t. \tag{38}$$

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

Conjecture 7.2. *Let $h \ge 3$ be an integer. Suppose $0 \le a \le 1$ 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* (10).

It is worth mentioning that an argument following the lines of the proof of the upper bound in Theorem 2.6 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 $h = 3$.

References

1. N. Alon, J. Balogh, R. Morris, and W. Samotij, *Counting sum-free sets in Abelian groups*, Israel J. Math, to appear.
2. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.
3. P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79.
4. S. Chen, *On the size of finite Sidon sequences*, Proc. Amer. Math. Soc. **121** (1994), no. 2, 353–356.
5. S. Chowla, *Solution of a problem of Erdős and Turán in additive-number theory*, Proc. Nat. Acad. Sci. India. Sect. A. **14** (1944), 1–2.
6. J. Cilleruelo, *New upper bounds for finite B_h sequences*, Adv. Math. **159** (2001), no. 1, 1–17.
7. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Submitted, 70pp, 2010.
8. A. G. D'yachkov and V. V. Rykov, *B_s-sequences*, Mat. Zametki **36** (1984), no. 4, 593–601.
9. P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208.
10. P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215.
11. B. Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390.
12. H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
13. S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.
14. X. D. Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), no. 1, 84–92.
15. D. J. Kleitman and D. B. Wilson, *On the number of graphs which lack small cycles*, manuscript, 15 pp, 1996.
16. D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), no. 2, 167–172.
17. Y. Kohayakawa, S. Lee, V. Rödl, and W. Samotij, *The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers*, Random Structures Algorithms, to appear.
18. Y. Kohayakawa, S. Lee, and V. Rödl, *The maximum size of a Sidon set contained in a sparse random set of integers*, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA), SIAM, 2011, pp. 159–171.
19. M. N. Kolountzakis, *The density of $B_h[g]$ sequences and the minimum of dense cosine sums*, J. Number Theory **56** (1996), no. 1, 4–11.
20. F. Krückeberg, *B_2-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. **206** (1961), 53–60.
21. B. Lindström, *A remark on B_4-sequences*, J. Combinatorial Theory **7** (1969), 276–277.
22. K. O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic).
23. D. Saxton and A. Thomason, *Hypergraph containers*, arXiv:1204.6595, April 2012.
24. M. Schacht, *Extremal results for random discrete structures*, Submitted, 27pp, 2009.
25. I. E. Shparlinskiĭ, *On B_s-sequences*, Combinatorial analysis, No. 7 (Russian), Moskov. Gos. Univ., Moscow, 1986, pp. 42–45, 163.
26. J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Transactions of the American Mathematical Society **43** (1938), 377–385.

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA (D. Dellamonica Jr., Y. Kohayakawa and V. Rödl)

E-mail address: domingos.junior@gmail.com, rodl@mathcs.emory.edu

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil (Y. Kohayakawa)

E-mail address: yoshi@ime.usp.br

Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea (S. J. Lee)

E-mail address: sjlee242@gmail.com

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, and Trinity College, Cambridge CB2 1TQ, UK (W. Samotij)

E-mail address: samotij@post.tau.ac.il