AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **83(1)** (2022), Pages 129–140

Upper and lower bounds on the size of $B_k[g]$ sets

Griffin Johnston Michael Tait*

Department of Mathematics & Statistics
Villanova University
Villanova, PA, U.S.A.xa
jjohns79@villanova.edu michael.tait@villanova.edu

Craig Timmons†

Department of Mathematics and Statistics
California State University Sacramento
U.S.A.
craig.timmons@csus.edu

Abstract

A subset A of the integers is a $B_k[g]$ set if the number of k-element multisets from A that sum to any fixed integer is at most g. Let $F_{k,g}(n)$ denote the maximum size of a $B_k[g]$ set in $\{1,\dots,n\}$. In this paper we improve the best-known upper bounds on $F_{k,g}(n)$ for $g>1$ and k large. When $g=1$ we match the best upper bound of Green with an improved error term. Additionally, we give a lower bound on $F_{k,g}(n)$ that matches a construction of Lindström while removing one of the hypotheses.

1 Introduction

We will denote the set $\{1,2,\dots,n\}$ by $[n]$. Given natural numbers k and g, a subset of $\mathbb{Z}$ is called a $B_k[g]$ set if for any m there are at most g multisets $\{x_1,\dots,x_k\}$ such that $x_1+\cdots+x_k=m$ and $x_i\in A$. Determining bounds on the maximum size of a $B_k[g]$ set in $[n]$ is a difficult and well-studied problem and it is the focus of this paper. Let $F_{k,g}(n)$ be the maximum size of a $B_k[g]$ set in $[n]$.

When $k=2$ and $g=1$, $B_2[1]$ sets are called *Sidon sets*. They have been studied extensively since being introduced by Sidon [28] in the context of Fourier series, and then studied further by Erdős and Turán [14] from a combinatorial perspective. It is known [5, 14] that $F_{2,1}(n)\sim n^{1/2}$, and determining whether or not $F_{2,1}(n)=$

* The second author is partially supported by National Science Foundation grant DMS-2011553.
† Research is supported in part by Simons Foundation Grant #359419

ISSN: 2202-3518

$n^{1/2} + O(1)$ is a 500 USD Erdős problem [13]. Very recently the error term was improved by Balogh, Füredi, and Roy [2] to $F_{2,1}(n) \leq n^{1/2} + 0.998n^{1/4}$ for sufficiently large n. This represents the first improvement upon the error term in over 50 years.

For other choices of k and g, the asymptotic behavior of $F_{k,g}(n)$ has not been determined, yet there are upper and lower bounds that give the order of magnitude as a function of n. First we discuss upper bounds.

If A is a $B_k[g]$ set contained in $[n]$, then each of the $\binom{|A|+k-1}{k}$ k-element multisets from A determines a sum which is an integer in $[kn]$. Each such integer can appear as a sum at most g times so that $\binom{|A|+k-1}{k} \leq gkn$. This implies $F_{k,g}(n) \leq (gk!kn)^{1/k}$ which is known as the trivial bound. When $g = 1$, a $B_k[g]$ set is often called a B_k set. Nontrivial bounds on the size of a B_k set were proved by Jia [18] and Kolountzakis [19] for k even, and by Chen [7] for k odd. These bounds show that

$$F_{k,1}(n) \leq \left(\left\lfloor \frac{k}{2} \right\rfloor ! \left\lceil \frac{k}{2} \right\rceil ! kn \right)^{1/k} + O_k(1). \tag{1.1}$$

When k is large, these bounds were improved by Green [17], who proved

$$F_{k,1}(n) \leq \left(\left\lceil \frac{k}{2} \right\rceil ! \left\lfloor \frac{k}{2} \right\rfloor ! \sqrt{\frac{\pi k}{2}} (1 + \epsilon_k) n \right)^{1/k}. \tag{1.2}$$

It is noted (see page 379 of [17]) that ϵ_k can be taken to be $O(k^{-1/8})$.

For $g > 1$, Cilleruelo, Ruzsa, and Trujillo [10] improved the trivial bound by showing

$$F_{k,g}(n) \leq \left(\frac{k!kgn}{1 + \cos^k(\pi/k)} \right)^{1/k}. \tag{1.3}$$

Cilleruelo and Jiménez-Urroz [8], using an idea attributed to Alon, showed

$$F_{k,g}(n) \leq \left(\sqrt{3k} k! gn \right)^{1/k}. \tag{1.4}$$

When $3 \leq k \leq 6$, (1.3) is a better bound while (1.4) is better for large k. Currently the best general upper bound, proved by the third author [31], is

$$F_{k,g}(n) \leq (1 + o(1)) \left(\frac{x_k k! kgn}{\pi} \right)^{1/k}. \tag{1.5}$$

Here x_k is the unique real number in $(0, \pi)$ that satisfies $\frac{\sin x_k}{x_k} = \left(\frac{4}{3 - \cos(\pi/k)} - 1 \right)^k$. In [31] it is shown that this upper bound is better than both (1.3) and (1.4), but that $\frac{x_k k! kg}{\pi} \to \sqrt{3k} k! g$ as $k \to \infty$.

Our first main theorem is an upper bound for large k that improves (1.5) and matches (1.2). Our theorem improves the error term in (1.2) from $O(k^{-1/8})$ [17] to $O(k^{-1/3})$.

Theorem 1.6. *Let $A \subset [n]$ be a $B_k[g]$ set.*

(i) When $g = 1$

$$|A| \leq \left(\left\lceil \frac{k}{2} \right\rceil ! \left\lfloor \frac{k}{2} \right\rfloor ! \sqrt{\frac{\pi k}{2}} (1 + O(k^{-1/3}))n \right)^{\frac{1}{k}},$$

and

(ii) when $g > 1$

$$|A| \leq \left(\sqrt{\frac{\pi k}{2}} k! g (1 + O(k^{-1/3}))n \right)^{\frac{1}{k}}.$$

The proof of Theorem 1.6 uses a Berry-Esseen type theorem [4, 15] and is inspired by the recent work of Dubroff, Fox, and Xu [12] who used a similar technique applied to the Erdős distinct subset sums problem.

Next we turn to lower bounds. Bose and Chowla [5] constructed a B_k set of size q in $\mathbb{Z}_{q^k-1}$ for q an odd prime power. This implies that $F_{k,1}(n) \geq (1 - o(1))n^{1/k}$. Lindström [20] showed that

$$F_{k,g}(n) \geq (1 - o(1))(gn)^{1/k}, \tag{1.7}$$

when $g = m^{k-1}$ for some integer m. Cilleruelo and Jiménez-Urroz [8] proved that for any $\epsilon > 0$, there exists a constant $g(k, \epsilon)$ so that for $g > g(k, \epsilon)$

$$F_{k,g}(n) \geq \left((1 - \epsilon) \sqrt{\frac{\pi k}{6}} gn \right)^{1/k}. \tag{1.8}$$

Our second main theorem matches the bound in (1.7) and removes the requirement that $g = m^{k-1}$.

Theorem 1.9. *For any integers* $k \geq 2$ *and* $g \geq 1$*, we have* $F_{k,g}(n) \geq (1 - o(1))(gn)^{1/k}$.

Before continuing the discussion it is important to note that shortly after a preprint of this article was made available the authors were notified by Carlos Trujillo that, while not stated explicitly, Theorem 1.9 follows from results of Caicedo, Gómez, and Trujillo [6]. The proof ideas are similar with the notable difference that [6] works first in a general setting, and then specializes to known B_h-sets. In our work we focus only on Bose-Chowla B_h sets and consequently, the proof of the lower bound $F_{k,g}(n) \geq (1 - o(1))(gn)^{1/k}$ is shorter. However, we recommend [6] for details on this technique and how it can be applied to B_h sets of Bose and Chowla, Ruzsa, and Gómez and Trujillo. We leave in the details of the proof of Theorem 1.9 for completeness and because it includes a density of primes argument that gives an asymptotic lower bound for all n; the theorem in [6] applies only to an infinite sequence of n.

Previous work on bounding $F_{k,g}(n)$ is extensive and we have not included all of it. In particular, there are numerous papers that consider the case when $k = 2$ and

$g > 1$ (e.g. [9, 11, 19, 22, 23, 32, 33]). For more information, see the surveys of Plagne [26] and O'Bryant [25]. In Section 2 we prove Theorem 1.6 and in Section 3 we prove Theorem 1.9.

2 Upper bounds

For a finite set $S \subset \mathbb{N}$, we define $\mathbb{E}(S) = \frac{1}{|S|}\sum_{s\in S} s$ and $\mathrm{Var}(S) = \frac{1}{|S|}\sum_{s\in S}(s-\mathbb{E}(S))^2$. For any random variable X let f_X be its probability distribution function and let F_X be its cumulative distribution function.

Assume that A is a $B_k[g]$ set in $[n]$. To give an upper bound for the size of A, we will consider the distribution of sums of random elements of A. Define random variables X_i that are independent and identically distributed by

$$\mathbb{P}(X_i = a - \mathbb{E}(A)) = \frac{1}{|A|}$$

for every $a \in A$. Note that $\mathbb{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = \mathrm{Var}(A)$. Define δ such that $\mathrm{Var}(A) = \delta n^2$. Any set of natural numbers up to n has variance at most $\frac{(n-1)^2}{4}$ and so $\delta \leq 1/4$.

The details split into two cases depending on whether $g = 1$ or $g > 1$. When $g = 1$ we take advantage of the fact that if A is a $B_k[1]$ set, then for any $c \in \mathbb{Z}$ there is at most one solution (up to rearranging) to the equation

$$a_1 + \cdots + a_{\lceil k/2\rceil} - a_{\lceil k/2\rceil+1} - \cdots - a_k = c, \tag{2.1}$$

where $a_1, \ldots, a_k \in A$.

When $g > 1$, if A is a $B_k[g]$ set, then there are at most g solutions (up to rearranging) to the equation

$$a_1 + \cdots + a_k = c. \tag{2.2}$$

Define

$$\begin{aligned} Y_1 &= X_1 + \cdots + X_{\lceil k/2\rceil} \\ Y_2 &= X_{\lceil k/2\rceil+1} + \cdots + X_k \\ Y &= Y_1 - Y_2 \\ Z &= X_1 + \cdots + X_k. \end{aligned}$$

When $g = 1$ we will consider the random variable Y and when $g > 1$ we will consider Z.

In [8], Cilleruelo and Jiménez-Urroz give an upper bound on the size of a $B_k[g]$ set for $g > 1$ that depends on the variance of the set. Their proof is easily modified to include the $g = 1$ case. For our purposes it is more convenient to phrase the result in terms of the variance of the set. We give a short proof for completeness.

Theorem 2.3 (Theorem 1.1 in [8]). *Let $k, g \in \mathbb{N}$ be fixed. If A is a $B_k[g]$ set in $[n]$, then for $g = 1$ we have*

$$\frac{|A|^{2k}}{12\left(\lceil \frac{k}{2} \rceil! \lfloor \frac{k}{2} \rfloor!\right)^2} \leq (k + o(1))\mathrm{Var}(A),$$

and for $g > 1$ we have

$$\frac{|A|^{2k}}{12(gk!)^2} \leq (k + o(1))\mathrm{Var}(A).$$

Proof. For convenience we assume that $|A|^k$ is divisible by $\lceil \frac{k}{2} \rceil! \lfloor \frac{k}{2} \rfloor!$ when $g = 1$ and by $gk!$ when $g > 1$. If this is not the case, then we may truncate A and let the $o(1)$ terms account for the difference.

Because the X_i are independent, we have that $\mathrm{Var}(Y) = \mathrm{Var}(Z) = k\mathrm{Var}(A)$. To lower bound this quantity, observe that the variance of Y or Z is as small as possible when the values taken by the random variables are as close together as possible. By (2.1) and (2.2), when we look at all outputs of Y or Z, each output can occur at most $\lceil \frac{k}{2} \rceil! \lfloor \frac{k}{2} \rfloor!$ times for $g = 1$ and at most $gk!$ times for $g > 1$. Hence, we have that the variance is bounded below by the variance of the multiset of integers from 1 to ℓ where $\ell = \frac{|A|^k}{\lceil \frac{k}{2} \rceil! \lfloor \frac{k}{2} \rfloor!}$ when $g = 1$, and where $\ell = \frac{|A|^k}{gk!}$ when $g > 1$. Since each integer occurs the same number of times in this multiset, the variance of the multiset is the same as that of the discrete uniform distribution of integers up to ℓ. This is given by

$$\mathrm{Var}(\{1, \ldots, \ell\}) = \frac{\ell^2 - 1}{12},$$

and the result follows. □

When k gets large, we can improve Theorem 2.3 by using more precise information about Y and Z than the variance. As k goes to infinity, these distributions will be close to normal distributions, and we use a Berry-Esseen [4, 15] theorem to quantify this.

Theorem 2.4 (Berry-Esseen). *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = \mathrm{Var}(X_i)$ and $\mathbb{E}[|X_i|^3] = \rho_i < \infty$. Let $X = X_1 + \cdots + X_n$, $\sigma^2 = \mathbb{E}[X^2]$, and $\psi = \sigma^{-3} \cdot \sum_{i=1}^n \rho_i$. Then*

$$\sup_{x \in \mathbb{R}} |F_X(x) - \Phi(x)| \leq 0.56\psi,$$

where $F_X(x)$ and $\Phi(x)$ are the cumulative distribution functions for X and the normal distribution with mean zero and standard deviation σ respectively.

One brief remark before continuing is that our proof does not depend on the value of the constant 0.56; any constant here would work as it will be absorbed into an error term. The 0.56 constant we use is proved in [27].

By Theorem 2.4, for any j we can approximate $X_1+\cdots+X_j$ by a normal random variable with mean 0 and variance $j\delta n^2$ by using

$$\begin{aligned}\rho_i &= \mathbb{E}[|X_i|^3] \le n\mathbb{E}[X_i^2] = \delta n^3,\\ \sigma^2 &= \mathrm{Var}(X) = j\mathrm{Var}(X_i) = j\delta n^2,\\ \psi &= \frac{1}{\sigma^3}\sum_{i=1}^{j}\rho_i \le \frac{j\delta n^3}{j^{3/2}\delta^{3/2}n^3} = \frac{1}{\sqrt{j\delta}}.\end{aligned} \tag{2.5}$$

We will approximate Y_1 by a normal distribution $\mathcal{N}(0, \lceil\frac{k}{2}\rceil\delta n^2)$ that has probability distribution $\varphi_1(x)$ and cumulative distribution function $\Phi_1(x)$. Similarly, let $\mathcal{N}(0, \lfloor\frac{k}{2}\rfloor\delta n^2)$ and $\mathcal{N}(0, k\delta n^2)$ have probability distribution functions $\varphi_2(x)$ and $\varphi(x)$ and cumulative distribution functions $\Phi_2(x)$ and $\Phi(x)$, respectively. Since F_{Y_1} and F_{Y_2} are close to Φ_1 and Φ_2 by Theorem 2.4, we have that F_Y is close to Φ, quantified by the following lemma.

Lemma 2.6. *For Φ the cumulative distribution function of $\mathcal{N}(0, k\delta n^2)$, we have*

$$\sup_x |F_Z(x) - \Phi(x)| \le \frac{0.56}{\sqrt{k\delta}},$$

and

$$\sup_x |F_Y(x) - \Phi(x)| \le 4\cdot\frac{0.56}{\sqrt{\lfloor\frac{k}{2}\rfloor\delta}}.$$

Proof. The first inequality follows from Theorem 2.4 and (2.5). Now we prove the second. Since $Y = Y_1 - Y_2$ we have that $f_Y(x) = (f_{Y_1} * f_{-Y_2})(x)$. We also have $\varphi = \varphi_1 * \varphi_2$. Let x be arbitrary. Then

$$\begin{aligned}|F_Y(x) - \Phi(x)| &= \left|\int_{-\infty}^{x} f_{Y_1} * f_{-Y_2} - \varphi_1 * \varphi_2\right|\\ &= \left|\int_{-\infty}^{x} \varphi_2 * (f_{Y_1} - \varphi_1) + \varphi_1 * (f_{-Y_2} - \varphi_2) + (f_{Y_1} - \varphi_1) * (f_{-Y_2} - \varphi_2)\right|\\ &\le \left|\int_{-\infty}^{x} \varphi_2 * (f_{Y_1} - \varphi_1)\right| + \left|\int_{-\infty}^{x} \varphi_1 * (f_{-Y_2} - \varphi_2)\right|\\ &+ \left|\int_{-\infty}^{x} (f_{Y_1} - \varphi_1) * (f_{-Y_2} - \varphi_2)\right|.\end{aligned}$$

We use Theorem 2.4 to show that each of the final three terms are small.

$$\begin{aligned}\left|\int_{-\infty}^{x} \varphi_2 * (f_{Y_1} - \varphi_1)\right| &= \left|\int_{-\infty}^{x}\int_{-\infty}^{\infty} \varphi_2(z-y)\Big(f_{Y_1}(y) - \varphi_1(y)\Big)dzdy\right|\\ &= \left|\int_{-\infty}^{x} f_{Y_1}(y) - \varphi_1(y)\, dy \int_{-\infty}^{\infty} \varphi_2(z-y)dz\right|\end{aligned}$$

$$\begin{aligned}
&= \left| \int_{-\infty}^{x} f_{Y_1}(y) - \varphi_1(y)\, dy \right| \left| \int_{-\infty}^{\infty} \varphi_2(z-y) dz \right| \\
&\leq \sup_{x \in \mathbb{R}} |F_{Y_1}(x) - \Phi_1(x)| \left| \int_{-\infty}^{\infty} \phi_2(z-y) dz \right| \\
&\leq \frac{0.56}{\sqrt{\delta \lceil \frac{k}{2} \rceil}}
\end{aligned}$$

where the last inequality follows because φ_2 is a probability distribution function and from (2.5). Similarly, noting that φ_2 is symmetric around 0, we have that

$$\left| \int_{-\infty}^{x} \varphi_1 * (f_{-Y_2} - \varphi_2) \right| \leq \frac{0.56}{\sqrt{\delta \lfloor \frac{k}{2} \rfloor}},$$

and

$$\begin{aligned}
\left| \int_{-\infty}^{x} (f_{Y_1} - \varphi_1) * (f_{-Y_2} - \varphi_2) \right| &\leq \left| \int_{-\infty}^{x} f_{Y_1} * (f_{-Y_2} - \varphi_2) \right| + \left| \int_{-\infty}^{x} \varphi_1 * (f_{-Y_2} - \varphi_2) \right| \\
&\leq 2 \frac{0.56}{\sqrt{\delta \lfloor \frac{k}{2} \rfloor}}
\end{aligned}$$

□

We now have everything that we need to prove Theorem 1.6.

Proof of Theorem 1.6. Let k and g be fixed and assume that A is a $B_k[g]$ set in $[n]$. As before, let $\mathrm{Var}(A) = \delta n^2$. If $\delta < \frac{\pi}{24}$, then we may apply Theorem 2.3 to obtain the claimed upper bound. Hence, for the remainder of the proof we may assume that $\frac{\pi}{24} \leq \delta \leq \frac{1}{4}$. Since the X_i are independent, we have that the standard deviations of the random variables Y and Z are the same and we will denote this quantity by σ. We will consider the probability of the events that $-t < Y \leq t$ and $-t < Z \leq t$ where t is an integer that will be chosen later. For $g = 1$, by Lemma 2.6 and the assumption that $\delta \geq \pi/24$, we have that

$$|F_Y(x) - \Phi(x)| \leq \frac{4 \cdot 0.56}{\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}}$$

for all x. Hence,

$$\begin{aligned}
\mathbb{P}[-t < Y \leq t] &= F_Y(t) - F_Y(-t) \\
&= \Phi(t) - \Phi(-t) - \Big((F_Y(-t) - \Phi(-t)) - (F_Y(t) - \Phi(t)) \Big) \\
&\geq \Phi(t) - \Phi(-t) - \Big| F_Y(-t) - \Phi(-t) \Big| - \Big| F_Y(t) - \Phi(t) \Big|
\end{aligned}$$

$$\begin{aligned}
&\geq \Phi(t) - \Phi(-t) - \frac{8 \cdot 0.56}{\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}} \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-t}^{t} \exp(-\frac{x^2}{2\sigma^2}) dx - \frac{4.48}{\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}} \\
&\geq \frac{(2t) \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}} - \frac{4.48}{\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}}.
\end{aligned}$$

On the other hand, by (2.1), we have that for any fixed x

$$\mathbb{P}[Y = x] \leq \left(\left\lceil \frac{k}{2} \right\rceil !\right) \left(\left\lfloor \frac{k}{2} \right\rfloor !\right) |A|^{-k}.$$

Combining these two inequalities yields

$$\frac{(2t) \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}} - \frac{4.48}{\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}} \leq \mathbb{P}[-t < Y \leq t] \leq (2t) \left(\left\lceil \frac{k}{2} \right\rceil !\right) \left(\left\lfloor \frac{k}{2} \right\rfloor !\right) |A|^{-k}.$$

Using the inequality $1 - x \leq e^{-x}$ for all x and $\sigma^2 \geq k\pi n^2/24$, we have

$$e^{\frac{-t^2}{2\sigma^2}} \geq 1 - \frac{t^2}{2\sigma^2} \geq 1 - \frac{12t^2}{\pi k n^2}.$$

Applying this inequality, dividing both sides by $2t$ and using $\sigma \leq \frac{\sqrt{k}n}{2}$ leads to

$$\frac{1 - \frac{12t^2}{\pi k n^2}}{n\sqrt{\pi k/2}} - \frac{4.48}{2t\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}} \leq \left(\left\lceil \frac{k}{2} \right\rceil !\right) \left(\left\lfloor \frac{k}{2} \right\rfloor !\right) |A|^{-k}.$$

Setting $t = k^{1/3}n$, we find that s

$$\frac{1 - \frac{12}{\pi k^{1/3}}}{n\sqrt{\pi k/2}} - \frac{\frac{4.48}{2k^{1/3}n\sqrt{\lfloor \frac{k}{2} \rfloor \frac{\pi}{24}}} \cdot n\sqrt{\pi k/2}}{n\sqrt{\pi k/2}} \leq \left(\left\lceil \frac{k}{2} \right\rceil !\right) \left(\left\lfloor \frac{k}{2} \right\rfloor !\right) |A|^{-k}.$$

Rearranging gives the result for $g = 1$. For $g > 1$ we use (2.2) and have that

$$\mathbb{P}[Z = x] \leq gk!|A|^{-k}$$

for any x. Performing a similar calculation on $\mathbb{P}(-k^{1/3}n < Z \leq k^{1/3}n)$ gives the result and we omit these details.

□

3 Lower Bounds

In this section we prove Theorem 1.9. The idea is to begin with a known construction of a $B_k[1]$ set and then consider the image of that set in a quotient group. This idea has been used in other extremal graph theory and combinatorial number theory problems before [1, 2, 16, 21, 24, 30]. In particular, and as noted in the discussion following Theorem 1.9, Caicedo, Gómez, and Trujillo [6] contains a more general approach in comparison to what is done here.

Let g and k be fixed positive integers with $k \geq 2$. Assume that q is a power of prime such that g divides $q-1$. Let $\mathbb{F}_{q^k}$ be the finite field with q^k elements and let θ be a generator of the multiplicative group $\mathbb{F}_{q^k}^*$ of nonzero elements of $\mathbb{F}_{q^k}$. Bose and Chowla [5] proved that

$$A = \{a \in \mathbb{Z}_{q^k-1} : \theta^a - \theta \in \mathbb{F}_q\}$$

is a $B_k[1]$ set in $\mathbb{Z}_{q^k-1}$. Let $\mu = \frac{q^k-1}{g}$ and let H be the subgroup of $\mathbb{Z}_{q^k-1}$ generated by μ. Thus, H is the unique subgroup of $\mathbb{Z}_{q^k-1}$ with g elements.

Next we prove a lemma that is crucial to the construction. This lemma is essentially known (Lemma 2.2 of [29] or Lemma 2.1 of [3]). A short proof is included for completeness.

Lemma 3.1. *If g, k, q, H and A are given as above and $A-A := \{a-b : a, b \in A\}$, then $(A-A) \cap H = \{0\}$.*

Proof. Suppose $a, b \in A$ and $a-b \in H$. There is an element $s \in \{0, 1, \ldots, g-1\}$ such that $a - b \equiv s\mu \pmod{q^k-1}$, so $\theta^{a-b} = \theta^{s\mu}$ in $\mathbb{F}_{q^k}$. Let $\alpha, \beta \in \mathbb{F}_q$ satisfy $\theta^a = \theta + \alpha$ and $\theta^b = \theta + \beta$. Observe that $(\theta^{s\mu})^{q-1} = (\theta^{q^k-1})^{\frac{s(q-1)}{g}} = 1$ and so $\theta^{s\mu} \in \mathbb{F}_q$. From $\theta^{a-b} = \theta^{s\mu}$ it follows that $\theta + \alpha = \theta^{s\mu}(\theta + \beta)$ so

$$(\theta^{s\mu} - 1)\theta + \theta^{s\mu}\beta - \alpha = 0.$$

The minimal polynomial of θ over $\mathbb{F}_q$ has degree $k \geq 2$ and so we must have $\theta^{s\mu} - 1 = 0$ and $\theta^{s\mu}\beta - \alpha = 0$. In particular, the first equation implies $s = 0$ and so $a \equiv b \pmod{q^k-1}$. □

In the quotient group $\Gamma := \mathbb{Z}_{q^k-1}/H$, let $A_H = \{a + H : a \in A\}$. If $a + H = b + H$ for some $a + H, b + H \in A_H$, then $a - b \in H$ which, by Lemma 3.1, implies $a \equiv b \pmod{q^k-1}$. Hence, $q = |A| = |A_H|$. Next we prove that A_H is a $B_k[g]$ set in Γ. Suppose $c + H \in \Gamma$ and we have

$$a_1 + H + \cdots + a_k + H = c + H \tag{3.2}$$

for some $a_i + H \in A_H$. We will show that there are at most g such solutions up to the ordering of the terms on the left hand side of (3.2). Indeed, (3.2) implies $a_1 + \cdots + a_k \equiv c + h \pmod{q^k-1}$ for some $h \in H$. There is at most one multiset $\{a_1, \ldots, a_k\}$ from A that is a solution to this equation. As there are g choices for h,

there will be at most g multisets $\{a_1 + H, \ldots, a_k + H\}$ from A_H that are solutions to (3.2). Therefore, A_H is a $B_k[g]$ set in Γ.

We now finish the proof using a density of primes argument. For positive integers x, c, and m, let $\pi(x; c, m)$ be the number of primes p for which $p \leq x$ and $p \equiv c (\mathrm{mod}\ m)$. Writing ϕ for the Euler phi function, the Prime Number Theorem in Arithmetic Progressions states that if $\gcd(c, m) = 1$, then

$$\pi(x; c, m) = \frac{x}{\phi(m) \ln x} + O\left(\frac{x}{\ln^2 x}\right).$$

Let $\alpha = \lfloor (1-\epsilon)^{1/k}(gn)^{1/k} \rfloor$ and $\beta = \lfloor (gn)^{1/k} \rfloor$. We then have

$$\pi(\beta; 1, g) - \pi(\alpha; 1, g) \geq \frac{1}{\phi(g)} \left(\frac{\beta}{\ln \beta} - \frac{\alpha}{\ln \alpha} - O\left(\frac{\beta}{\ln^2 \beta}\right)\right). \tag{3.3}$$

For large enough n depending on ϵ, g, and k, the right hand side of (3.3) is positive since $\frac{x}{\log x}$ is strictly increasing for $x > e$. Thus, there is a prime q with $q \equiv 1 (\mathrm{mod}\ g)$ and $\alpha \leq q \leq \beta$. We can then choose a $B_k[g]$-set A in the group $\mathbb{Z}_{q^k-1}/H$ where $|A| = q$. This group is isomorphic to the cyclic group $\mathbb{Z}_{(q^k-1)/g}$ and we let A' be a $B_k[g]$-set in this cyclic group. Since $q \leq \beta$, we have $\frac{q^k-1}{g} \leq n$. Therefore, we can view A' as a subset of $\{1, 2, \ldots, n\}$ and under integer addition, A' is still a $B_k[g]$-set. It remains to show that $q \geq (1 - o(1))(gn)^{1/k}$, but this follows from the definition of α and the inequality $q \geq \alpha$. We conclude that for all positive integers g and k with $k \geq 2$,

$$F_{k,g}(n) \geq (1 - o(1))(gn)^{1/k}.$$

Acknowledgments

The authors would like to thank Carlos Trujillo for bringing [6] to our attention.

References

[1] N. Alon, L. Rónyai and T. Szabó, Norm-graphs: variations and applications, *J. Combin. Theory Ser. B* **76**(2) (1999), 280–290.

[2] J. Balogh, Z. Füredi and S. Roy, An upper bound on the size of Sidon sets, `arXiv:2103.15850` (2021).

[3] F. A. Benavides, D. F. Daza and C. A. Trujillo, Sidon sets and C_4-saturated graphs, `arXiv:1810.05262` (2019).

[4] A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Amer. Math. Soc.* **49** (1941), 122–136.

[5] R. C. Bose and S. Chowla, Theorems in the additive theory of numbers, *Comment. Math. Helv.* **37**(1) (1962), 141–147.

[6] Y. Caicedo, J. Gómez and C. Trujillo, $B_h[g]$ modular sets from B_h modular sets, *JP J. Algebra Number Theory Appl.* **37**(1) (2015), 1–19.

[7] S. Chen, On the size of finite Sidon sequences, *Proc. Amer. Math. Soc.* **121**(2) (1994), 353–356.

[8] J. Cilleruelo and J. Jiménez-Urroz, $B_h[g]$ sequences, *Mathematika* **47**(1-2) (2000), 109–115.

[9] J. Cilleruelo, I. Ruzsa and C. Vinuesa, Generalized Sidon sets, *Adv. Math.* **225**(5) (2010), 2786–2807.

[10] J. Cilleruelo, I. Z. Ruzsa and C. Trujillo, Upper and lower bounds for finite $B_h[g]$ sequences, *J. Number Theory* **97**(1) (2002), 26–34.

[11] J. Cilleruelo and C. Vinuesa, $B_2[g]$ sets and a conjecture of Schinzel and Schmidt, *Combin. Probab. Comput.* **17**(6) (2008), 741–747.

[12] Q. Dubroff, J. Fox and M. W. Xu, A note on the Erdős distinct subset sums problem, *SIAM J. Discrete Math.* **35**(1) (2021), 322–324.

[13] P. Erdős, Some old and new problems on additive and combinatorial number theory, In *Combinatorial Mathematics: Proceedings of the Third International Conference* (1989), 181–186.

[14] P. Erdős and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, *J. Lond. Math. Soc. (2)* **1**(4) (1941), 212–215.

[15] C. G. Esseen, On the Liapounoff limit of error in the theory of probability, *Ark. Mat. Astr. Fys.* **28A**(9) (1942), 19.

[16] Z. Füredi, New asymptotics for bipartite Turán numbers, *J. Combin. Theory Ser. A* **75**(1) (1996), 141–144.

[17] B. Green, The number of squares and $B_h[g]$ sets, *Acta Arith.* **100**(4) (2001), 365–390.

[18] X. D. Jia, On finite Sidon sequences, *J. Number Theory* **44** (1993), 84–92.

[19] M. N. Kolountzakis, The density of $B_h[g]$ sequences and the minimum of dense cosine sums, *J. Number Theory* **56**(1) (1996), 4–11.

[20] B. Lindström, $B_h[g]$-sequences from B_h-sequences, *Proc. Amer. Math. Soc.* (2000), 657–659.

[21] I. Livinsky, A construction for bipatite Turán numbers, `arXiv:2101.06726` (2021).

[22] G. Martin and K. O'Bryant, The symmetric subset problem in continuous Ramsey theory, *Exp. Math.* **16**(2) (2007), 145–165.

[23] G. Martin and K. O'Bryant, Constructions of generalized Sidon sets, *J. Combin. Theory Ser. A* **113**(4) (2006), 591–607.

[24] Z. L. Nagy, Supersaturation of C_4: From Zarankiewicz towards Erdős–Simonovits–Sidorenko, *European J. Combin.* **75** (2019), 19–31.

[25] K. O'Bryant, A complete annotated bibliography of work related to Sidon sequences, *Electron. J. Combin.* **DS**(11) (2004), 39pp.

[26] A. Plagne, Recent progress on finite $B_h[g]$ sets, *Congr. Numer.* (2001), 49–64.

[27] I. G. Shevtsova, An improvement of convergence rate estimates in the Lyapunov theorem, *Dokl. Math.* **82**(3) (2010), 862–864.

[28] S. Sidon, Ein satz über trigonometrische polyome und seine anwendung in der theorie der Fourier-reihen, *Math. Ann.* **106** (1932), 536–539.

[29] M. Tait and C. Timmons, Sidon sets and graphs without 4-cycles, *J. Comb.* **5**(2) (2014), 155–165.

[30] M. Tait and C. Timmons, The Zarankiewicz problem in 3-partite graphs, *J. Combin. Des.* **27**(6) (2019), 391–405.

[31] C. Timmons, Upper bounds for $B_h[g]$-sets with small h, *Integers* **16** paper A83 (2016), 1–12.

[32] G. Yu, An upper bound for $B_2[g]$ sets, *J. Number Theory* **122**(1) (2007), 211–220.

[33] G. Yu, A note on $B_2[g]$ sets, *Integers* **8**(1) paper A58 (2008), 1–5.

(Received 1 July 2021; revised 1 Mar 2022)