$B_h[g]$ **SEQUENCES**

Javier Cilleruelo and Jorge Jiménez-Urroz

Abstract. We give new upper and lower bounds for $F_h(g,N)$, the maximum size of a $B_h[g]$ sequence contained in $[1,N]$. We prove

$$F_h(g,N) \le (\sqrt{3h}h!gN)^{1/h},$$

and for any $\epsilon > 0$ and $g > g(\epsilon,h)$,

$$F_h(g,N) \ge \left((1-\epsilon)\sqrt{\frac{\pi}{6}}\sqrt{h}gN\right)^{1/h} + o(N^{1/h}).$$

1. Introduction

Given a sequence of integers A, we define $R_h(A;k)$ as the number of representations of k as the sum of h elements of A,

$$R_h(A;k) = \#\{k = a_1 + \cdots + a_h;\ a_1 \le \cdots \le a_h, a_i \in A\},$$

and we say that a sequence of integers A is a $B_h[g]$ sequence if $R_h(A;k) \le g$ for any integer k. Sidon was led to consider such sequences in connection with the theory of Fourier series. $B_2[1]$ sequences are also called Sidon sequences.

It is a major problem giving good estimates for $F_h(g,N)$, the maximum size of $B_h[g]$ sequences contained in $\{1,\dots,N\}$. See [H-R] for a classical reference about this topic, and [S-S] and [K] for recent surveys.

By a trivial counting argument we obtain the upper bound

$$F_h(g,N) \le (hh!gN)^{1/h}.$$

1991 *Mathematics Subject Classification.* 11B13.

Key words and phrases. B_h sequences, Sidon sets.

Typeset by $\mathcal{AMS}$-TeX

For $g = 1$, Erdős and Turan [E-T] proved $F_2(1, N) \leq N^{1/2} + O(N^{1/4})$. See [C1] for new upper bounds for $F_h(1, N), h \geq 3$.

On the other hand, Erdős observed in an addendum to [E-T] that a construction of J. Singer in [S] gives $F_2(1, N) \geq N^{1/2} + o(N^{1/2})$. (See [R] for an easy construction). Later R. C. Bose and S. Chowla [B-Ch] were able to prove $F_h(1, N) \geq N^{1/h} + o(N^{1/h})$ for any integer h.

When $g > 1$ is more difficult to obtain good estimates for $F_h(g, N)$. The first author [C2] and M. Helm [H], independently proved $F_2(2, N) \leq \sqrt{6N} + 1$. In [C-R-T] nontrivial upper bounds were proved for $F_h(g, N)$:

$$F_2(g, N) \leq \left(\frac{4}{1 + \frac{1}{(\pi/2+1)^2}} gN \right)^{1/2} + o(N^{1/2}),$$

$$F_h(g, N) \leq \frac{1}{(1 + \cos^h(\pi/h))^{1/h}} (hh!gN)^{1/h} + o(N^{1/h}), \qquad h \geq 3.$$

Also N. Alon (see [K]) has obtained non trivial upper bounds for large h by exploiting the "concentration" of the sums $a_1 + \cdots + a_h$ around their mean, using the Chebyshev inequality. He gets

$$F_h(g, N) \leq \left(3^{3/2} \sqrt{h} h! gN \right)^{1/h} + o(N^{1/h}).$$

In section 2 we get new upper bounds which improve the previous ones for $h \geq 7$. In particular we prove

Theorem 1.1.
$$F_h(g, N) \leq (\sqrt{3h} h! gN)^{1/h}.$$

In relation to the lower bounds, a construction of M. Kolountzakis [K] of $B_2[2]$ sequences gives $F_2(2, N) \geq \sqrt{2N} + o(N^{1/2})$. In [C-R-T] it is proved that $F_2(g, N) \geq \sqrt{\frac{g+[g/2]}{g+2[g/2]}} N^{1/2} + o(N^{1/2})$.

Recently Lindstrom [L] generalized the argument of [K] to prove $F_h(g, N) \geq (gN)^{1/h} + o(N^{1/h})$ for $g = m^{h-1}$, $m \geq 2$.

In section 3 we obtain new lower bounds, improving the previous ones for $h \geq 3$.

Theorem 1.2. *Let h be an integer fixed. For any $\epsilon > 0$, and for any $g > g(\epsilon, h)$*

$$F_h(g,N) \geq \left((1-\epsilon)\sqrt{\frac{\pi}{6}}\sqrt{h}gN\right)^{1/h} + o(N^{1/h}),$$

when $N \to \infty$.

In our construction it is needed g to be big enough. This situation already appears in [L, (1.6)] and in fact our construction recover Lindstrom result.

2. Upper bounds

In this section we note that the use of Chebyshev inequality is wasteful in Alon's argument. Instead of it we obtain a lower bound for the variance by observing that it cannot be smaller than the case where the integers are as compressed as possible.

Proof of Theorem 1.1.

Suppose that $A \subset [1, N]$ is a $B_h[g]$ sequence. Let the random variable Y be defined by $Y = X_1 + \cdots + X_h$, where the X_j are independent random variables uniformly distributed in A. We can obtain an upper bound in an easy way:

$$E\left((Y-\bar{Y})^2\right) = hE\left((X-\bar{X})^2\right) \leq hE\left((X-(N+1)/2)^2\right) \leq h\frac{(N-1)^2}{4}.$$

In order to estimate $E\left((Y-\bar{Y})^2\right)$ we consider the multiset $hA = \{a_1 + \cdots + a_h; a_i \in A\} = \{s_i : \quad i = 1, \dots, k\}$, where $k = |A|^h$.

$$|A|^h E\left((Y-\bar{Y})^2\right) = \sum_{s_i \in hA} (s_i - \bar{Y})^2.$$

The minimum value of the variance of a set happens when the elements are as close as possible. We observe that the s_i's take integer values which appear, at most, $gh!$ times (because A is a $B_h[g]$ sequence of integers). Then, the variance of hA is not less than the variance of the multiset $L = \{\overbrace{1, \dots, 1}^{h!g \text{ times}}, 2, \dots, 2, \dots, l, \dots, l\}$, where $l = [k/gh!]$. Hence,

$$|A|^h E\left((Y-\bar{Y})^2\right) \geq \sum_{x \in L} (x - \bar{x})^2 = \sum_{x \in L} x^2 - |L|\bar{x}^2$$

$$= gh!\sum_{k=1}^{l} k^2 - gh!l\left(\frac{l+1}{2}\right)^2 = gh!\frac{l(l+1)(2l+1)}{6} - gh!\frac{l(l+1)^2}{4}$$

$$= gh!\frac{l(l+1)(l+2)}{12} \geq \frac{k^3}{12(gh!)^2} - \frac{k}{12} \geq |A|^h\left(\frac{|A|^{2h}}{12(gh!)^2} - \frac{1}{12}\right).$$

Then we have proved

$$\frac{1}{12}\left(\frac{|A|^{2h}}{(gh!)^2} - 1\right) \leq h\frac{(N-1)^2}{4},$$

which implies

$$|A| \leq \left((gh!)^2(3h(N-1)^2+1)\right)^{1/2h} \leq (\sqrt{3h}h!gN)^{1/h}.$$

□

3. Lower bounds

Now we are interested in $B_h[g]$ sequences as dense as possible. We will establish a generalization of Theorem 2.1 of [C-R-T] for any integer h.

The proof will go as Theorem 2.1 in [C-R-T]. So, first of all we will need the analogous definitions as 2.1 and 2.2 in [C-R-T] for this general context.

Definition 3.1. *We say that A satisfies the $B_h^*[g]$ condition if the equation $a_1+\cdots+a_h = k$ has at most g solutions for any k, counting different those in distinct order.*

Definition 3.2. *We say that a sequence of integers $C = \{c_i\}$ is a B_h $\pmod m$ sequence if $c_{i_1} + \cdots + c_{i_h} = c_{j_1} + \cdots + c_{j_h} \pmod m$ implies $\{c_{i_1}, \ldots, c_{i_h}\} = \{c_{j_1}, \ldots, c_{j_h}\}$.*

Now we can establish the lemma generalizing Lemma 2.2 in [C-R-T].

Lemma 3.1. *If $A = \{a_i\}$ satisfies the $B_h^*[g]$ condition, and C is a B_h $\pmod m$ sequence, then $B = \cup_{i=0}^{k}(C + ma_i)$ is a $B_h[g]$ sequence.*

Proof. Suppose $b_{1,1} + \cdots + b_{h,1} = \cdots = b_{1,g+1} + \cdots + b_{h,g+1}$ for $b_{i,j} \in B$. We can write $b_{i,j} = c_{i,j} + ma_{i,j}$ for some $c_{i,j} \in C$ and $a_{i,j} \in A$. Let us order $b_{i,j}$ such that, for any i and j, $c_{i,j} \leq c_{i+1,j}$. Then, since

$c_{1,1}+\cdots+c_{h,1} \equiv c_{1,j}+\cdots+c_{h,j} \pmod m$ for any $1 \leq j \leq g+1$ we have that all the sets $\{c_{1,j},\dots,c_{h,j}\}$ are the same. Moreover the elements are ordered and so $c_{i,j} = c_{i,1}$ for every i,j. This implies further that all the $g+1$ sums $a_{1,j}+\cdots+a_{h,j}$ are equal hence, for some j,j' we have $a_{i,j} = a_{i,j'}$ for any $1 \leq i \leq h$. Both together give us, for these j,j', that $b_{i,j} = b_{i,j'}$ for any $1 \leq i \leq h$. □

In order to use Lemma 3.1, we have to find convenient sequences C and A on those conditions.

It is known [p. 81, H-R], that for $m = p^h - 1$, p prime, there exists a $B_h \pmod m$ sequence $C_m \subset [1,m]$ with cardinal $|C_m| = p$.

On the other hand we can choose the trivial $A_n = \{0,1,\dots,n-1\}$ to get our bounds. Our next step is to find the greatest n so that A_n satisfies the $B_h^*[g]$ condition. Let us call $n(g,h)$ to this n. We have

Proposition 3.1. $F_h(g,N) \geq n(g,h)^{1-1/h}N^{1/h} + o(N^{1/h})$.

Proof. Let us take a prime p such that $n(g,h)(p^h-1) = N + o(N)$. Now we apply Lemma 3.1 with $m = p^h - 1$, $A = A_{n(g,h)}$. Then $B \subset [1, n(g,h)m]$ and $|B| = n(g,h)p$. □

In order to estimate $n(g,h)$ we define

$$r_h(n,k) = \#\{k = a_1+\cdots+a_h : 0 \leq a_i \leq n-1\}$$

and $M_h(n) = \max_k r_h(n,k)$. Then $n(g,h)$ is the greatest n such that $M_h(n) \leq g$.

Proposition 3.2. $M_h(n) \sim n^{h-1}\frac{2}{\pi}\int_0^\infty \left(\frac{\sin t}{t}\right)^h dt$

Proof. The obvious inequality $M_h(2m-1) \leq M_h(2m) \leq M_h(2m+1)$ allows us to reduce the proof to $n = 2m+1$, odd. In this case we can write $r_h(n,k) = \#\{k-hm = a_1+\cdots+a_h : -m \leq a_i \leq m\}$. It is now trivial to deduce

$$r_h(n,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{j=-m}^{m} e^{ij\theta}\right)^h e^{-i(k-hm)}d\theta,$$

by expanding the h-power of the Dirichlet kernel. Hence, by a simple change of variables

$$r_h(n,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\sin((m+1/2)\theta)}{\sin(\theta/2)}\right)^h e^{-i(k-hm)\theta}d\theta$$

$$= \frac{1}{\pi}(2m+1)^{h-1}\int_{-\pi(m+1/2)}^{\pi(m+1/2)}\left(\frac{\sin t}{t}\right)^h\left(\frac{t/(2m+1)}{\sin(t/(2m+1))}\right)^h e^{-i\frac{k-hm}{m+1/2}t}dt$$

$$= \frac{1}{\pi}(2m+1)^{h-1}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^h e^{-i\frac{k-hm}{m+1/2}t}dt + o(m^{h-1}).$$

Now observe that $\frac{\sin t}{t}$ is the fourier transform of the characteristic function of the interval $[-1,1]$. Then $\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^h e^{-ixt}dt$ is the value of the h-convolution of $\chi_{[-1,1]}$ at x, whi ch is maximum at $x=0$ and so $M_h(n) = \max_k r_h(n,k) \sim n^{h-1}\frac{2}{\pi}\int_0^{\infty}\left(\frac{\sin t}{t}\right)^h dt.$ □

Proposition 3.3. *Let h be an integer fixed. For every $\epsilon > 0$ and for every $g > g(\epsilon, h)$ we have*

$$F_h(g,N) \geq \left((1-\epsilon)\frac{g}{m_h}N\right)^{1/h} + o(N^{1/h}).$$

where $m_h = \frac{2}{\pi}\int_0^{\infty}\left(\frac{\sin t}{t}\right)^h dt$.

Proof. It is consequence of Proposition 3.1 and 3.2 □

Proof of Theorem 1.2. We only need to study the behaviour of m_h. The upper bound $m_h \leq \sqrt{6/\pi h}$ for $h \geq 100$ follows from the corresponding for J_p (1) in [3.3, L-N]. For $3 \leq h \leq 100$ we get the bound by computing the explicit formula (17), (see (15)), of [N]

$$m_h = \frac{1}{(h-1)!}\sum_{j<h/2}(h/2-j)^{h-1}\binom{h}{j}(-1)^j.$$

In particular we get for the first few values

$$m_3 = \frac{3}{4},\quad m_4 = \frac{2}{3},\quad m_5 = \frac{115}{192},\quad m_6 = \frac{11}{20},\quad m_7 = \frac{5587}{11520}$$

$$m_8 = \frac{151}{315},\quad m_9 = \frac{259723}{573550},\quad m_{10} = \frac{15619}{36288}.$$

The case $h=2$ is covered in [C-R-T]. □

Remark 1. It is possible to find an explicit formula for $r_h(n,k)$ by using its generating function $\left(\sum_{k=0}^{n-1}x^k\right)^h = \sum_k r_h(n,k)x^k$ together with $\sum x^n = \frac{1}{1-x}$. Moreover, one can prove that its maximum $M_h(n)$ is attained at the mean, $k_h = (n-1)h/2$ or $k_h = ((n-1)h+1)/2$. In this way one can obtain the explicit expression

$$M_h(n) = \sum_{j=0}^{h} \binom{\frac{n-1}{2}h + \delta - nj + h - 1}{h-1} \binom{h}{j} (-1)^j,$$

where $\delta = 0$ or $1/2$ depending on the parity of $(n-1)h$, which is useful for small values of h.

For example, for $h = 3$ we obtain $M_3(n) = \left[\frac{3n^2+1}{4}\right]$, and Proposition 3.1 gives the more precise estimate

$$F_3(g, N) \geq \left(\left[\sqrt{\frac{4g}{3}} \right]^2 N \right)^{1/3} + o(N^{1/3}).$$

Remark 2. Induction in $r_h(n,k) = \sum_{j=0}^{n-1} r_{h-1}(n, k-j)$ immediately implies $M_h(n) \leq n^{h-1}$. This and Proposition 3.1 recover Lindstrom result.

Acknowledgment.

We would like to express our gratitude to the referee for suggesting the analytical approach to prove Proposition 3.2. This has made the proof considerably shorter and more direct than our previous combinatorial method. We also would like to thank A. Córdoba and F. Chamizo for their comments in order to present the final proof of Proposition 3.2 as it appears now.

We are pleased to thank J.L. Nicolas for providing us with the very useful infornation in [L-N] and [N].

References

[B-Ch] R.C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.

[C-R-T] J. Cilleruelo, I. Ruzsa and C. Trujillo, *Upper and lower bounds for $B_h[g]$ sequences*, to appear in J. Number Theory.

[C1] J. Cilleruelo, *New upper bounds for finite B_h sequences*, preprint.

[C2] J. Cilleruelo, *An upper bound for $B_2[2]$ sequences*, Journal of Combinatorial. Series A **89** (2000), 141-144.

[E-T] P. Erdős and P. Turan, *On a problem of Sidon in additive number theory and on some related problems*, J.London Math.Soc. **16** (1941), 212–215; Addendum (by P. Erdős), ibid **19** (1944), 208.

[H-R] H. Halberstam and K.F. Roth, *Sequences*, Springer-Verlag, New York, 1983.

[H] M. Helm, *Upper bounds for $B_2[g]$-sets*, preprint.

[K] M. Kolountzakis, *Problems in the Additive Number Theory of General Sets, I. Sets with distinct sums.*, 1996, Available at http://www.math.uiuc.edu/ ∼ kolount/surveys.html.

[L-N] L. Lesieur and J.L. Nicolas, *On the Eulerian numbers $M_n = \max_{1\le k\le n} A(n,k)$*, Europ. J. Comb. **13** (1992), 379–399.

[L] B. Lindstrom, *$B_h[g]$-sequences from B_h sequences*, Proc.Amer.Math.Soc. **128** (2000), 657-659.

[N] J.L. Nicolas, *An Integral representation for Eulerian numbers*, Coll. Math. Soc. János Bolyai 60, Sets, Graphs and Numbers Budapest, 1991.

[R] I. Ruzsa, *Solving a linear equation I*, Acta Arithmetica **LXV.3** (1993).

[S-S] A. Sarközy and V.T. Sos, *The Mathematics of Paul Erdös*, vol. I, Algorithms and Combinatorics 13, Springer, 1996.

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

Departamento de Matematicas. Universidad Autónoma de Madrid. Madrid-28049.SPAIN

E-mail address: `franciscojavier.cilleruelo@uam.es`

Departamento de Matematicas. Universidad Autónoma de Madrid. Madrid-28049.SPAIN

E-mail address: `jorge.jimenez@uam.es`