Available at http://www.math.uiuc.edu/~kolount/surveys.html

Problems in the Additive Number Theory of General Sets, I

Sets with distinct sums

Mihail N. Kolountzakis[1]
October 10, 1996

§0. Introduction

0.1 In these notes I shall present some additive number-theoretic problems which refer to "general" sets. This means that I will not be interested so much in the additive properties of specific sets such as the primes (e.g., the Goldbach Conjecture that every large even integer can be written as the sume of two primes) or the n-th powers (e.g., the Waring problem in which one wants to find information about the number of ways to represent an integer as a sum of k n-th powers).

The typical question that we will be interested in is "how large a subset of $\{1, \ldots, N\}$ can one find with a given additive property".

I will try to draw the reader's attention to a few open problems among the many that appear implicitly or explicitly here. My reasons for pointing out those problems and not some others are that I believe that they stand a good chance to be solved soon, the fact that I have given them some thought myself and, most important, that I like them more than other problems.

0.2 $B_h[g]$ **sets.** In the first lecture we shall deal with sets of the type $B_h[g]$. For h, g positive integers, $h \ge 2$, we call a set $A \subseteq \mathbf{N}$ a $B_h[g]$ set if $r_h(x) \le g$ for all

[1]Department of Mathematics, University of Illinois at Urbana-Champaign, 1409 W. Green St., Urbana, IL 61801. E-mail: kolount@math.uiuc.edu

positive integers x. The *representation function* $r_h(x)$ of a set A is defined by

$$r_h(x) = r_{h,A}(x) = \#\Big\{(a_1,\ldots,a_h) \in A^h : \ a_j \le a_{j+1} \ \ \& \ \ x = a_1 + \cdots + a_h\Big\}.$$

$B_h[1]$ sets are called just B_h sets and B_2 sets are sometimes called *Sidon* sets (although the term has an entirely different meaning in harmonic analysis). Notice that the property of being a $B_h[g]$ set is translation invariant.

We shall consider both finite (subsets of $\{1,\ldots,N\}$) and infinite sets and we shall be concerned with the question "how large can a $B_h[g]$ set be?" By far the best (though not completely) understood case are the B_2 sets. The principal reference is [15, Ch 2,3].

§1. Finite $B_h[g]$ sets

1.1 The order of magnitude. The trivial upper bound. Let $F_{h,g}(N)$ be the maximum size of a $B_h[g]$ subset of $\{1,\ldots,N\}$. We shall also write $F_h(N) = F_{h,1}(N)$. It is easy to see that

$$F_{h,g}(N) \lesssim C_{h,g} N^{1/h}, \tag{1}$$

where C_g is a constant that depends on h and g only. To see (1) observe that, if $A \subset \{1,\ldots,N\}$ is a $B_h[g]$ set, then all expressions of the type

$$a_1 + \cdots + a_h, \quad \text{with } a_1 \le \ldots \le a_h, \ \& \ a_j \in A, \tag{2}$$

fall in the interval $[1, hN]$ and at most g of them can coincide. This reasoning gives

$$F_{h,g}(N) \lesssim (g\,h\,h!)^{1/h} N^{1/h}. \tag{3}$$

1.2 A lower bound. It is not as easy to give lower bounds for $F_{h,g}(N)$. Even proving $F_2(N) \gtrsim C\sqrt{N}$ is not trivial. Erdős [26] gave such a proof (see also [15, p. 90]): Take a prime p and cosider the set $A \subset \{2p^2,\ldots,4p^2-p\}$ defined by

$$A = \Big\{2p(k+p) + (k^2)_p \ : \ k = 1,\ldots,p-1\Big\}, \tag{4}$$

where $(x)_p \in \{0,\ldots,p-1\}$ denotes the residue of the integer x mod p. It is not hard to show that A is a B_2 set. After shifting it to the interval $[0, 2p^2-p]$ it remains a B_2 set of size $p-1$. Thus $F_2(N) \gtrsim \sqrt{N/2}$.

Furthermore, as a corollary of a theorem of Singer [25], generalized by Bose and Chowla [3, 4] we know

$$F_h(N) \gtrsim N^{1/h}. \tag{5}$$

Notice that for $h = 2$ this gives a larger set than Erdős's construction (4). This was first obtained for $h = 2$ as a corollary of a theorem of Singer that guarantees the existence of so called perfect difference sets in $\mathbf{Z}_m$ (the residues mod m) for certain values of the integer m. These are subsets D of $\mathbf{Z}_m$ such that every non-zero element of $\mathbf{Z}_m$ can be represented exactly once as a difference of two elements of D. The construction is algebraic (see [15, p. 79]).

1.3 B_2 sets. There is an important difference between the B_2 case (and that of *even* h more generally) and that of the general $B_h[g]$ case (with either odd h or $g > 1$). A set A is of the type B_2 if and only if it has distinct differences, as well as distinct sums. That is, all the expressions of the type $a+b$, $a \le b$, $a, b \in A$, are distinct if and only if all the expressions of the type $a-b$, $a > b$, $a, b \in A$, are distinct. There exists an analogous statement for B_h sets for even h but not if h is odd or (worse) if $g > 1$.

We emphasize that working with a set that has distinct differences is a much easier task than working with one that has distinct sums. The vast majority of results that have been obtained (mostly the upper bounds for $F_{h,g}(N)$ but some lower bounds as well) have been obtained using information about distinct differences and are therefore restricted to either the B_2 case only or the, slightly more general, B_h case with even h.

As an example of how distinct differences help consider the determination of the best constant $C_{2,1}$ in (1). We saw in (3) that $C_{2,1} \le 2$. Let us now use the information that a B_2 set $A \subset \{1, \ldots, N\}$ has distinct differences. Let $k = |A|$. All the differences of the type $a-b$, $a > b$, $a, b \in A$, fall in the interval $[1, N]$, their number is $\binom{k}{2} \sim k^2/2$ and, therefore, $k \lesssim \sqrt{2}\sqrt{N}$. That is, we proved

$$C_{2,1} \le \sqrt{2}. \tag{6}$$

A similar improvement over (3) can be gained for h even and $g = 1$.

1.4 A connection with harmonic analysis The reason that the sets of the type B_2 or, more generally, B_h were first investigated (by Sidon, according to Erdős) was that they turn out to have interesting properties from the point of view of harmonic analysis (see also §1.6). Let us briefly mention here the fact that (infinite) sets of the type B_2 are so-called Λ_4 sets. A Λ_p set of integers $E = \{\lambda_1, \lambda_2, \ldots\}$, with $p \ge 1$, is a set for which all L^q norms, for $1 \le q \le p$, are comparable, for all trigonometric polynomials with frequencies on E only. That is, for all numerical sequences a_j we have

$$\|f\|_q \le C_q \|f\|_1, \quad \text{for any } f(x) = \sum_j a_j e^{i\lambda_j x}. \tag{7}$$

To prove that a B_2 set is of the type Λ_4 it is sufficient to show that

$$\|f\|_4 \le C_g \|f\|_2. \tag{8}$$

Inequality (7) then follows by a simple application of Hölder's inequality. To prove

(8) one simply notices that, because E is B_2, the Fourier coefficients of f^2 are of the type $2a_j a_k$, when $j \neq k$, or a_j^2.

1.5 The Erdős–Turán theorem. Erdős and Turán [11] have proved that the actual value of $C_{2,1}$ is 1 (they proved the upper bound–the lower bound is Singer's theorem). They proved

$$F_2(N) \leq N^{1/2} + O(N^{1/4}). \tag{9}$$

Erdős [10] actually believed that the error term in (9) can be made $O(N^\epsilon)$, for any positive ϵ. This inequality has been extended to the case of even $h = 2m$ by Jia [21] and myself [19]:

$$F_{2m}(N) \leq (m(m!)^2)^{1/2m} N^{1/2m} + O(N^{1/4m}). \tag{10}$$

(This contains an earlier result of Lindström [24] who had proved that $F_4(N) \leq 2^{3/4}N^{1/4} + O(N^{1/8})$.) Jia's method was a generalization of that of Erdős and Turán while the method in [19] is described in §1.6.

1.6 An analytic method for bounding $F_{2m}(N)$ from above. One first proves that if

$$M = -\min_{x\in[0,2\pi)} \sum_{j=1}^n \cos \lambda_j x, \tag{11}$$

and the distinct positive integers λ_j are all $\leq (2-\epsilon)n$, then

$$M \geq C\epsilon^2 n. \tag{12}$$

(This result was obtained while working on variants of the so called *cosine problem* in which lower bounds for M as in (11) are sought, where the λ_j can be any distinct integers. A conjectured lower bound (Chowla [5]) is $\gtrsim \sqrt{n}$.) The proof of this result is based on a very old theorem of Fejér [13] which bounds from above the value at 0 of a n-th degree, nonnegative trigonometric polynomial, of integral 1, by $n+1$ instead of the trivial $2n+1$. This is itself a consequence of a well known theorem of Fejér and Riesz which states that every nonegative trigonometric polynomial of degree n is the square of the modulus of a trigonometric polynomial of the same degree.

Assume now that $A = \{n_1, \ldots, n_k\} \subset \{1, \ldots, N\}$ is a B_{2m} set. Then one can show that all expressions of the type

$$a_1 + \cdots + a_m - b_1 - \cdots - b_m, \tag{13}$$

with

$$a_j, b_j \in A,\ a_j < a_{j+1},\ b_j < b_{j+1} \text{ and } a_i \neq b_j, \tag{14}$$

are distinct. Form a nonnegative cosine sum by

$$\begin{aligned} f(x) &= \left|\sum_{j=1}^{k} e^{in_j x}\right|^h = \left(\sum_{j=1}^{k} e^{in_j x}\right)^m \left(\sum_{j=1}^{k} e^{-in_j x}\right)^m \\ &= r(x) + 2(m!)^2 \left(\sum_{a_j,\, b_j \text{ satisfy (14)}} \cos\left(\sum a_j - \sum b_j\right)x\right). \end{aligned}$$

The "remainder" $r(x)$ is of modulus $O(k^{h-1})$ and thus the cosine polynomial in the parentheses has a "small" minimum. If the number k were large the number of frequencies $\sum a_j - \sum b_j$ of that cosine polynomial would be too large for its degree and (12) would contradict the fact that it has a "small" minimum.

> **Problem 1.** Determine the value of $C_{4,1}$ in $F_{4,1}(N) \sim C_{4,1}N^{1/4}$. What's known is that $2^{1/2} \leq C_{4,1} \leq 2^{3/4}$.

1.7 More upper bounds: the case of odd h and $g = 1$. Some progress has also been made in the case when h is odd. Graham [14] has proved

$$F_{2k-1}(N) \leq (k!)^{2/(2k-1)} N^{1/(2k-1)} + O(N^{-1/(4k-2)}). \tag{15}$$

The method is similar to that of Erdős and Turán and Jia (van der Corput's lemma is used).

For the particular case of $h = 3$ Li [23] proved

$$F_3(N) \leq \left(4 - \frac{2\log^2 2}{3\log^2 N}\right)^{1/3} N^{1/3} + O(1).$$

Graham [14] improved the estimate to

$$F_3(N) \leq (3.99561029143 \cdot N)^{1/3} + O(1).$$

1.8 A probabilistic approach. Let $A = \{a_1, \ldots, a_k\} \subset \{1, \ldots, N\}$ be a $B_h[g]$ set. Alon [2] has obtained a better bound for $F_{h,g}(N)$ exploiting the "concentration" of the sums $a_1 + \cdots + a_h$ around their mean.

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. For the variance $\sigma^2(Y)$ we have, since the X_j are independent,

$$\sigma^2(Y) = h\sigma^2(X_1) \leq hN^2.$$

Chebyshev's inequality then gives

$$\mathbf{Pr}\left[|Y-\mathbf{E}Y|\geq\lambda\sigma(Y)\right]\leq\frac{1}{\lambda^2},$$

and the bound on $\sigma(Y)$ gives

$$\begin{aligned}\mathbf{Pr}\left[|Y-\mathbf{E}Y|\leq\sqrt{h}\lambda N\right] &\geq \mathbf{Pr}\left[|Y-\mathbf{E}Y|\leq\lambda\sigma(Y)\right]\\ &\geq 1-\frac{1}{\lambda^2}.\end{aligned}\tag{16}$$

On the other hand, since A is of the type $B_h[g]$,

$$\mathbf{Pr}\left[|Y-\mathbf{E}Y|\leq\lambda\sqrt{h}N\right]\leq k^{-h}2\lambda\sqrt{h}N\,gh!.\tag{17}$$

From (16) and (17) we conclude that

$$k^h\leq\frac{\lambda^3}{\lambda^2-1}2g\sqrt{h}h!N.\tag{18}$$

Optimizing for λ gives

$$F_{h,g}(N)\leq(3^{3/2}g\sqrt{h}h!)^{1/h}N^{1/h}.\tag{19}$$

The same method can be applied in the particular case of B_h sets, with h either even or odd. Let us consider the case of even $h=2m$. One then studies the random variable

$$Y=X_1+\cdots+X_m-X_1'-\cdots-X_m',$$

where, again, all the X_j, X_j' are independent and uniformly distributed in A. Using the fact that the expressions of the type (13) subject to condition (14) are all distinct and similar reasoning as above one obtains

$$F_{2m}(N)\lesssim(6^{3/2}\sqrt{m}(m!)^2)^{1/2m}N^{1/2m}.\tag{20}$$

This estimate is better than the one in (10) when $m\geq 6^3$. A similar estimate can be obtained for odd h.

> **Problem 2.** Reduce the constants $3^{3/2}$ and $6^{3/2}$ in (19) and (20) respectively.

1.9 A dense $B_2[2]$ set. Here we show how to use dense finite B_2 sets in order to construct a dense finite $B_2[2]$ set

$$B\subset\{1,\ldots,N\},\quad |B|\sim\sqrt{2N}.\tag{21}$$

We follow [19].

By Singer's theorem (5) there is a B_2 set $A \subset \{1, \ldots, \lfloor N/2 \rfloor - 1\}$, with $|A| \sim \sqrt{N/2}$. We show that $B = 2A \cup (2A+1) \subset \{1, \ldots, N\}$ is $B_2[2]$, which proves (21).

The proof is by contradiction. Assume that we have the non-trivial relations

$$x_1 + y_1 = x_2 + y_2 = x_3 + y_3, \tag{22}$$

with $x_j, y_j \in B$ and let $z = x_1 + y_1$. Look at $x_j + y_j \bmod 2$. There are three possible patterns: $0+0$, $1+1$ and $0+1$.

If z is even then only $0+0$ and $1+1$ may appear in (22) and we have either a relation of the pattern $0+0 = 0+0$ or a relation of the pattern $1+1 = 1+1$. Both cases contradict the fact that A is B_2, the first after just dividing by 2, the second after cancelling the remainders and then dividing by 2.

If z is odd then only the pattern $0+1$ appears in (22) which can be rewritten as

$$2a_1 + (2a_1' + 1) = 2a_2 + (2a_2' + 1) = 2a_3 + (2a_3' + 1) \tag{23}$$

with $a_j, a_j' \in A$. By canceling 1 and dividing by 2 we have

$$a_1 + a_1' = a_2 + a_2' = a_3 + a_3'.$$

But A is B_2 so for at least one of the above relations, say the first one, we have $a_1 = a_2$ and $a_1' = a_2'$ which contradicts the fact that the first relation in (23) is non-trivial.

Jia [22] has improved and generalized (21). By a similar method he has proved the existence of a $B_h[g]$ set $B \subset \{1, \ldots, N\}$ such that

$$|B| \sim (m(h,g))^{1-1/h} N^{1/h}, \tag{24}$$

where $m(h,g)$ is the largest integer m for which the equation $a = x_1 + \cdots + x_h$ has at most g solutions in $\mathbf{Z}_m$ (up to rearrangement) for each $a \in \mathbf{Z}_m$. For $h = g = 2$ the coefficient in (24) becomes $\sqrt{3}$, which is better than (21).

Problem 3. Find a "direct" way to construct a dense $B_2[2]$ set. It seems to me that combining two or more dense $B_2[1]$ sets as in my construction (or that of Jia) is bound to lead to a suboptimal value for $C_{2,2}$.

Problem 4. Find a non-trivial upper bound for $C_{2,2}$. We stress again that $B_2[2]$ sets do not have any property of distinct differences like the B_2 sets do. That's what makes the methods that give upper bounds for $C_{2,1}$ useless here. What's known is only the trivial upper bound $C_{2,2} \le 2\sqrt{2}$, which comes from (3).

§2. Infinite $B_h[g]$ sets with large lower density

2.1 In this section we look at the question "can one have an infinite $B_h[g]$ sequence that does not grow too fast?" Let $A = \{a_1 < a_2 < \cdots\}$ be an infinite $B_h[g]$ sequence. We write

$$A(N) = |A \cap \{1, \ldots, N\}|$$

for the *counting function* of the set A. Clearly if A is of the type $B_h[g]$ then

$$A(N) \le C N^{1/h}, \tag{25}$$

where, again, the constant C may depend on h and g only (this dependence will not concern us here). In the case of finite $B_h[g]$ sets one could always find a subset of $\{1, \ldots, N\}$ that matched the apriori maximum order of magnitude (the result of Bose and Chowla [3] mentioned in §1.2). The question here is whether one can do so with an infinite set. For example, in the case of B_2 sets is there an infinite B_2 set A with

$$A(N) \ge C\sqrt{N}? \tag{26}$$

The answer turns out to be negative as we shall see in §2.3.

The question can stated otherwise. Can one find a B_2 sequence $A = \{a_1 < a_2 < \cdots\}$ with

$$a_j \le C j^2 \tag{27}$$

for all j? We should point out that this is really a question about finite sequences and that what distinguishes this question from the investigations of §1 is that in §1 one only asked to find a large B_2 subset of, say, $\{1, \ldots, N\}$ where (26) was required to hold at *one* value of N only, whereas here it is required to hold for all values of N. Indeed, if one could, for each N, provide a finite B_2 subset of $\{1, \ldots, N\}$ for which (27) held for all $j \le N$ and with the same constant C for all different N, then one could use an easy diagonal argument to construct an infinite B_2 sequence for which (27) holds for all j.

2.2 Lower bounds. It is not hard to see that given any set of integers $B = \{b_1 < b_2 < \cdots\}$ one can always find a subsequence $A = \{b_{i_1} < b_{i_2} < \cdots\}$ which is B_2 and with $i_k \le k^3$. The way to do this is to pick the elements of A one by one and in increasing order, and always to pick the least possible element of B that will not destroy the B_2 property that A has up to that step (this is called the *greedy method*). One can easily prove $i_k \le k^3$ inductively. When B is the set of all positive integers we get a set A with $A(N) \ge C N^{1/3}$.

This result has only been improved slightly (and it is a difficult improvement) by Ajtai, Komlós and Szemerédi [1] who proved that there exists a B_2 set A with

$$A(N) \ge C(N \log N)^{1/3}. \tag{28}$$

Erdős conjectured that for any $\epsilon > 0$ one can have $A(N) \geq CN^{1/2-\epsilon}$.

2.3 An upper bound for the counting function of infinite B_2 sets. Here we sketch Erdős' proof [26] (we follow [15, p. 89]) that if $A = \{a_1 < a_2 < \cdots\}$ is an infinite B_2 set then

$$\liminf_{N\to\infty} \frac{A(N)\log^{1/2} N}{\sqrt{N}} \leq C. \tag{29}$$

Let N be large and define

$$\tau = \inf_{n\geq N} \frac{A(N)\log^{1/2} N}{\sqrt{N}}$$

and

$$D_l = |A \cap [(l-1)N, lN)|,$$

for $l = 1, \ldots, N$. Remember that a B_2 set has distinct differences as well as distinct sums and observe that for each interval $[(l-1)N, lN)$ the number of positive differences one can form with the elements of $A \cap [(l-1)N, lN)$ is $\binom{D_l}{2}$ which is $\sim D_l^2/2$. When we let $l = 1, \ldots, N$ and form all such differences they fall in the interval $[1, N]$ and they are all distinct. Therefore

$$\sum_{l=1}^{N} D_l^2 \leq CN. \tag{30}$$

We also have by the Cauchy-Schwarz inequality

$$\sum_{l=1}^{N} D_l l^{-1/2} \leq \left(\sum_{l=1}^{N} D_l^2\right)^{1/2} \left(\sum_{l=1}^{N} \frac{1}{l}\right)^{1/2}. \tag{31}$$

On the other hand, by summation by parts, we can get

$$\begin{aligned}
\sum_{l=1}^{N} D_l l^{-1/2} &\geq C\sum_{l=1}^{N} A(lN) l^{-3/2} \\
&\geq C\tau \sum_{l=1}^{N} \left(\frac{lN}{\log lN}\right)^{1/2} l^{-3/2} \\
&\geq C\tau \left(\frac{N}{\log N}\right)^{1/2} \sum_{l=1}^{N} \frac{1}{l}
\end{aligned}$$

Using now the last inequality together with (30) and (31) we get the desired $\tau \leq C$.

> **Problem 5.** Prove that one cannot have an infinite B_2 set with $A(N) \geq C\sqrt{N}$ by using an analytic method, perhaps in the spirit of §1.6. The similarity between the original combinatorial proof of the Erdős-Turán theorem (9) and the proof we gave in this section makes the existence of an analytic proof very likely.

2.4 Infinite B_h sets for even h. In the proof of §2.3 the fact that B_2 sets had distinct differences (as well as sums) was heavily used in (30). Some similar property of distinct differences is also present for B_h sets if h is even and a result similar to (29) is now known. Consult [18] for a list of references to the papers that led to the proof of the fact that

$$\liminf_{N\to\infty} \frac{A(N)}{N^{1/h}} = 0, \tag{32}$$

for any infinite B_h set A, and for any even h.

2.5 Odd h and the case $h = 3$. The situation is different when h is odd. The result (32) is not known to hold for any odd h. Some partial results are known in the case $h = 3$. We mention the following result of Helm [17] and indicate its proof. Let $A = \{a_1 < a_2 < \cdots\}$ be an infinite set. Then, Helm proves, if

$$\lim_{N\to\infty} \frac{A(N)}{N^{1/3}} = L > 0, \tag{33}$$

the set A cannot be of the type B_3.

We try to prove this in a way similar to that of §2.3. Observe that if a set A is of the type B_3 then all expressions of the type

$$a + b - c, \quad \text{with } a > b > c, \text{ and } a, b, c \in A, \tag{34}$$

are distinct. As in §2.3 let N be a large integer and define

$$D_l = |A \cap [(l-1)N, lN)|,$$

for $l = 1, \ldots, N$. The fact that the expressions (34) are all distinct implies (in a way similar to that in §2.3) that, for any k,

$$\sum_{i\ge j} D_i D_j D_{k+i+j} \le CN. \tag{35}$$

This is the analogue of (30) that was the principal ingredient in the proof of §2.3. Notice that here we have a family of inequalities, one for each value of k.

We are not going to show Helm's proof. Rather we show that assuming a "reasonable" behavior about the sequence a_j is enough to prove the result. We shall make the *assumption* that the sequence D_i is non-increasing. Helm's proof is very similar to the one described here. Furthermore we use (35) only for $k = 0$.

Since D_i is non-increasing we have (all the summation indices run from 1 to N unless otherwise restricted)

$$CN \ \ge\ \sum_{i\ge j} D_i D_j D_{i+j}$$

$$
\begin{aligned}
&\geq \sum_{i\geq j} D_{i+j}^3 \\
&\geq C\sum_i iD_i^3. && (36)
\end{aligned}
$$

We also have

$$
\begin{aligned}
S &= \sum_i i^{-1/3}D_i \\
&\leq \left(\sum_i iD_i^3\right)^{1/3}\left(\sum_i \frac{1}{i}\right)^{2/3} \quad \text{by Hölder's inequality,} \\
&\leq CN^{1/3}\log^{2/3} N, \quad \text{by (36),} && (37)
\end{aligned}
$$

as well as

$$
\begin{aligned}
S &\geq C\sum_i A(iN)i^{-4/3} \quad \text{by summation by parts,} \\
&\geq \sum_i (iN)^{1/3}i^{-4/3} \quad \text{by assumption (33),} \\
&\geq CN^{1/3}\log N. && (38)
\end{aligned}
$$

We get a contradiction from (37) and (38) which finishes the proof.

We should mention that Helm [18] has shown that use of the inequality (35) with $k = 0$ alone cannot prove that there are no infinite B_3 sequences A with $A(N) \geq CN^{1/3}$. That is because he has shown the existence of a sequence with this growth that satisfies (35) with $k = 0$ for infinitely many N. Thus, use of (35) for a single arbitrarily large N is not going to prove the general theorem, but the possibility is still open that one might exploit the fact that (35) actually holds *for all* N and k.

> **Problem 6.** Prove that there is no infinite B_3 set A with $A(N) \geq CN^{1/3}$. Erdős had put a high prize ($500 according to [6]) on this but I think he should have promised less. Of course, this is of no importance any more.

2.6 Infinite $B_2[g]$ sets of almost quadratic growth. Here we mention the following result of Erdős and Rényi [12]. For every $\delta > 0$ there is an integer g and an infinite $B_2[g]$ sequence $A = \{a_1 < a_2 < \cdots\}$ such that

$$a_j \leq Cj^{2+\delta},$$

for all $j > 0$. This is, of course, the same as saying that for any given $\delta > 0$ there exist such a set A with

$$A(N) \geq CN^{1/2-\delta}. \qquad (39)$$

We shall give the proof of this in a later section as its proof is probabilistic and is more suitable for presentation together with the results of Erdős on additive bases of the positive integers.

§3. Infinite $B_h[g]$ sets with large upper density.

3.1 Constructing infinite $B_h[g]$ sets with large *upper* density, that is sets which are dense infinitely often, seems to be a considerably easier problem. One wants to construct a $B_h[g]$ set $A = \{a_1 < a_2 < \cdots\}$ such that

$$K_{h,g} = \limsup_{N\to\infty} \frac{A(N)}{N^{1/h}} \tag{40}$$

is as large as possible. Most constructions build the set A by somehow putting together dense *finite* sets of the type $B_{h'}[g']$ (h' and g' need not be the same as h and g). The existence of those finite sets is guaranteed by the Singer or Bose-Chowla theorem (see §1.2). Typically, the constant $K_{h,g}$ that one gets thus will not be as large as the constant $C_{h,g}$ of §1.1 and there are no results at all to show that these two constants have to differ.

3.2 Infinite B_2 sequences with large upper density. Here we sketch the construction of an infinite B_2 set A with large upper density

$$A(N) \gtrsim \frac{1}{\sqrt{2}}\sqrt{N}. \tag{41}$$

This is a very simple construction of Erdős [26] and of Krückeberg [20] (see also [15, p. 90]). Compare with the result of Singer (5) that there exist B_2 subsets of $\{1,\ldots,N\}$ of size $\sim \sqrt{N}$.

Assume that we have constructed already our set A in the interval $[1,N]$. We show how to extend the set in an interval $[1,M]$, where M is very large compared to N, so that $A(M) \gtrsim \frac{1}{\sqrt{2}}\sqrt{M}$. The B_2 property is translation-invariant so we can find a B_2 subset B of $[M/2,M]$, of size $\sim \frac{1}{\sqrt{2}}\sqrt{M}$ and we add that set to the already existing part of A. We still have to eliminate some elements of B though, since relations of the type

$$b' - b = a' - a, \quad a, a' \in A,\ b, b' \in B, \tag{42}$$

are still possible (relations of a different type have been taken care of because of the size of the elements of B). For that, we remove from B the b' elements from any relation of the type (42). Since the number of those relations is $O(N)$ this removal will not affect the asymptotics of $A(M)$, as long as M is large enough.

Problem 7. Prove (41) with a constant larger than $1/\sqrt{2}$ (perhaps 1). My comment about the likely suboptimality of constructions that use ready-made building blocks (see Problem 3) applies here as well. In any case, a "direct" construction of a B_2 set that is dense infinitely often would be interesting in itself.

3.3 Infinite $B_2[2]$ sequences with large upper density. We show here how to construct an infinite $B_2[2]$ sequence A with

$$\limsup_{N\to\infty} \frac{A(N)}{\sqrt{N}} = 1. \tag{43}$$

Thus, as was the case in finite sets, going from the class B_2 to the class $B_2[2]$ helps construct denser sets. We follow [19] again.

It suffices to show that any $B_2[2]$ sequence $1 \le n_1 < \cdots < n_k$ can be extended to a sequence $1 \le n_1 < \cdots < n_k < n_{k+1} < \cdots < n_l$, such that $n_l \sim l^2$.

Write $A = \{n_1, \ldots, n_k\}$ and $x = n_k$. Take $B \subseteq \{2x+1, \ldots, x^4\}$ to be a B_2 set with $|B| \sim x^2$. In what follows $a_j \in A$, $b_j \in B$ and $d_j \in D$ (to be defined below).

Consider the relations of the form

$$a_1 + b_1 = a_2 + b_2. \tag{44}$$

Such a relation may be written as $a_1 - a_2 = b_2 - b_1$. But B is a B_2 set, so all differences $b_2 - b_1$ are distinct, which implies that a pair $a_1, a_2 \in A$ may appear in (44) only once. Thus there are $O(k^2) = O(x)$ of these relations which may involve $O(x)$ elements of B. Let then

$$D = \{b \in B : b \text{ does not appear in any relation of the form (44)}\} \tag{45}$$

and $E = A \cup D$. Obviously $|E| \sim x^2$. We show that E is a $B_2[2]$ set.

First note that the relations of the form

$$a_1 + a_2 = a_3 + d_1 \quad \text{or} \quad a_1 + a_2 = d_1 + d_2$$

are not possible (the left hand side is too small) and A is itself $B_2[2]$. This proves $r_E(a_1 + a_2; 2) \le 2$ for all $a_1, a_2 \in A$.

It remains to be checked that $r_E(a_1 + d_1; 2) \le 2$ and $r_E(d_1 + d_2; 2) \le 2$. By passing from B to D we eliminated all relations of the form (44) and so the only remaining non-trivial relations that we have to check are of the form

$$a_1 + d_1 = d_2 + d_3. \tag{46}$$

These are indeed possible. Assume $y = a_1 + d_1 = d_2 + d_3$. We have to show that these are the only ways that y can be written as a sum of two elements of E. But this is obvious since $y = d_2' + d_3'$ is impossible (this would mean $d_2 + d_3 = d_2' + d_3'$ which contradicts D in B_2), $y = a_1' + a_2'$ is impossible because of size and $y = a_1' + d_1'$ would mean that $a_1' + d_1' = a_1 + d_1$ which we took care to eliminate in (45).

> **Problem 8.** Prove (43) with a right-hand side larger than 1. The comment in Problem 7 applies here as well.

§4. Bibliography

[1] M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, Europ. J. Comb. **2** (1981), 1-11.

[2] N. Alon, Personal Communication, January 1995.

[3] R.C. Bose and S. Chowla, Theorems in the additive theory of numbers, Comment. Math. Helv. **37** (1962-63), 141-147.

[4] S. Chowla, Solution of a problem of Erdős and Turán in additive number theory, Proc. Nat. Acad. Sci. India **14** (1944) 1-2.

[5] S. Chowla, Some applications of a method of A. Selberg, J. Reine Angew. Math. **217** (1965), 128-132.

[6] R.K. Guy, *Unsolved problems in Number Theory*, Springer-Verlag, New York (1994).

[7] P. Erdős, On a problem of Sidon in additive number theory, Acta Sci. Math. (Szeged) **15** (1953-54), 255-259.

[8] P. Erdős, Problems and results in additive number theory, Colloque sur la Théorie des Nombres (CBRM, Bruxelles), 127-137.

[9] P. Erdős, Extremal problems in number theory, Proceedings of the Symp. Pure Math. VIII, AMS (1965), 181-189.

[10] P. Erdős, Some applications of probability theory to Number Theory. Successes and limitations, in *Sequences*, Capocelli (editor), Springer-Verlag.

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

[12] P. Erdős and A. Rényi, Additive properties of random sequences of positive integers, Acta Arith. **6** (1960), 83-110.

[13] L. Fejér, Über trigonometrische Polynome, J. Reine Angew. Math. **146** (1915), 53-82.

[14] S.W. Graham, Upper bounds for Sidon sequences, Analytic Number Theory, a conference in honor of H. Halberstam, Birkhäuser, 1996.

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

[16] M. Helm, Some Remarks on the Erdős-Turán conjecture, Acta Arith. **63** (1993), 373-378.

[17] M. Helm, On the distribution of B_3 sequences, J. Number Th. **58** (1996), 1, 124-129.

[18] M. Helm, On B_3 sequences, Analytic Number Theory, a conference in honor of H. Halberstam, Birkhäuser, 1996.

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

[20] F. Krückeberg, B_2 Folgen und verwandte Zahlenfolgen, J. Reine Angew. Math. **106** (1961), 53-60.

[21] X.-D. Jia, On finite Sidon sequences, J. Number Th. **44** (1993), 1, 84-92.

[22] X.-D. Jia, $B_h[g]$-sequences with large upper density, J. Number Th. **56** (1996), 2, 298-308.

[23] An Ping Li, On B_3 sequences (Chinese), Acta Math. Sinica **34** (1991), 67-71.

[24] B. Lindström, A remark on B_4 sequences, J. Comb. Theory **7** (1969), 276-277.

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

[26] A. Stöhr, Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe II, J. Reine Angew. Math. **194** (1955), 111-140.