Some Problems and Results on Combinatorial Number Theory

PAUL ERDÖS

AT&T Bell Laboratories
Murray Hill, New Jersey 07974
and
Mathematical Institute
Hungarian Academy of Sciences
Budapest, Hungary

I have written many papers with similar titles during my long life. I will try to write this paper in such a way that it will not entirely be contained in the union of the set of my previous papers and that at least some of the open problems I state will not be entirely hopeless. Perhaps the most interesting and significant results are those connected with van der Waerden's and Szemerédi's theorem, but since I and others have written a great deal about these questions, I will include only a short discussion of these problems at the end of the paper.

For a rich source of solved and unsolved problems in combinatorial number theory, see [10, 20].

First of all I mention a few old problems and results. First a very simple old well-known result of mine.

1. In December 1933 I observed that if $1 \leq a_1 < \cdots < a_{n+1} \leq 2n$, then for at least two indices i and j $a_i | a_j$, that is, among any $n+1$ integers $\leq 2n$, at least one divides the other. My first proof was not as simple as it should have been, but M. Wachsberger (Dr. M. Svéd) and E. Vázsonyi independently argued as follows: $a_i = 2^{d_i} b_i$, b_i odd. Thus two a's belong to the same b, and therefore one divides the other. The integers $n+1, \ldots, 2n$ show that this result is best possible.

Perhaps one of my first theorems, which I found in 1932, states: There is an absolute constant C so that if $a_1 < a_2 \cdots$ is a finite or infinite sequence of integers no one dividing the other, then

$$\sum_{i=1}^{\infty} \frac{1}{a_i \log a_i} < C. \tag{1}$$

The proof is simple but not entirely trivial. I conjectured that the sum in (1) is maximal if the a's are the primes. I expect that this problem can be decided with the help of a computer, some patience, and perhaps not too many new ideas. For more related problems and results of this type, see the survey paper of Sárközy, Szemerédi, and myself. (Selfridge convinced me that the proof will be more difficult than I thought.)

In 1936 I observed that if $1 < a_1 < \cdots < a_{k_n} \leq n$ is a sequence of integers for

which the products $a_i a_j$ are all distinct, then

$$\pi(n) + c_1 \frac{n^{3/4}}{(\log n)^{3/2}} < \max k_n < \pi(n) + c_2 \frac{n^{3/4}}{(\log n)^{3/2}}, \tag{2}$$

where $\pi(n)$ denotes the number of primes not exceeding n and c_1, c_2 are positive absolute constants. Probably there is a positive absolute constant c for which

$$\max k_n = \pi(n) + (c + o(1))\, n^{3/4}/(\log n)^{3/2}. \tag{3}$$

I was never able to prove (3). I should perhaps make a remark about the proof of (2), since it is basically graph-theroetical. It depends on the following result of E. Klein and myself: Denote by $T(n; C_4)$ the Turán number of C_4, that is, the smallest integer for which every graph of n vertices and $T(n;C_4)$ edges contains a C_4 (a cycle of length 4). We proved

$$c_1 n^{3/2} < T(n; C_4) < c_2 n^{3/2}. \tag{4}$$

Curiously we failed to formulate the general problems on extremal graphs—this was done 2 or 3 years later by Turán, who did not even know of our paper, and as well deserved punishment for our failure, these numbers are now called Turán numbers. For references on extremal graph theory, see the excellent book of Bollobás or the excellent survey paper of Simonovits, and [4, 5, 18, 21].

2. Here is another elementary problem in number theory that nevertheless leads to a problem that, as far as I know, is still open. Observe that if $1 < a_1 < \cdots < a_l \le n$, $l > n/2$, then at least two of the a's must be relatively prime. Denote by $f(n; t)$ the largest set of integers $1 \le a_1 < \cdots < a_{l_t} \le n$, for which there is no set of $t + 1$ a's that is pairwise relatively prime. I expect that one gets $f(n; t_n)$ by taking the t smallest primes $2, 3, \ldots, p_t$ and taking the set of all their multiples not exceeding n. This has been proved for $k \le 4$, but for $k = 4$ the proof is already quite long and cumbersome. The conjecture would follow if one could prove that t is obtained by taking the set of multiples not exceeding n of a set of t primes. If this is proved, then denote $\psi(q_1, \ldots, q_t; n)$ the set of multiples of $q_1, \ldots, q_t$ not exceeding n. To complete the proof it suffices to show that $\psi(q_1, \ldots, q_t)$ is maximal if the q's are the first t primes. In my lecture at Yenan Teachers College I stated that I think this is simple but I do not quite see how to do it. Fan Chung, who translated my lecture, said this must be trivial. I had another look at it and saw that she is indeed right. Let $Q_1, Q_2, \ldots$ be the complement of the q's, that is, the set of the other primes. Clearly $\psi(q_1, q_2, \ldots, q_t; n)$ equals n minus the number of integers not exceeding n composed of the Q's. The number of integers $\le n$ composed of the Q's is clearly minimal if the Q's are as large as possible or if $q_i = p_i$, *Q.E.D.*

3. R. L. Graham stated about 15 years ago the following very nice conjecture: Let $1 \le a_1 < a_2 < \cdots < a_n$ be any set of n integers. Then

$$\max_{1 \le i < j \le n} \frac{a_j}{(a_i, a_j)} \ge n.$$

Recently M. Szegedy proved this conjecture for sufficiently large n. His proof is not too simple and is very ingenious [25].

4. A little-known conjecture of Hegyvári states: Let $G(n)$ be the graph whose vertices are the integers $i \leq n$, and i is joined to j if $i|j$ (or if $[i, j] \leq n$). Denote by $f(n)$, respectively, $F(n)$ the longest path of this graph. Prove $f(n)$ and $F(n)$ are both $o(n)$. These conjectures were recently proved by Pomerance, but the proof is surprisingly difficult [24]. Is it true that $F(n)/f(n) \to \infty$ or at least $F(n) - f(n) \to \infty$? The second conjecture must certainly be true. It would perhaps be worth while to get an asymptotic formula or at least good upper and lower bounds for these functions.

5. Now perhaps I should discuss some problems and results on additive number theory that occupied me a great deal over the last few years. The classical problems (Goldbach, Waring) will be ignored, not because I do not consider them important, but because they perhaps cannot be attacked by combinatorial methods. Let $A = \{1 \leq a_1 < a_2 < \cdots\}$ be an infinite sequence of integers, and denote by $f_2(n)$ the number of solutions of $n = a_i + a_j$. If $f(n) > 0$ for all n, then A is called a basis of order 2. If $f(n) > 0$ for all $n > n_0$, then A is called an asymptotic basis of order 2. Sidon asked me more than 50 years ago: Does there exist a basis for which for every $\epsilon > 0$

$$\frac{f(n)}{n^{\epsilon}} \to 0 \tag{5}$$

as $n \to \infty$?

I first thought that (5) is easy and the answer is affirmative. Twenty years later I was able to prove (5) by probabilistic methods. In fact, I showed that there is a sequence $A = \{a_1, a_2, \ldots\}$ for which

$$c_1 \log n < f(n) < c_2 \log n. \tag{6}$$

Inequality (6) seems to be the natural boundary of the probabilistic methods, and in fact I conjectured long ago that there is no sequence A for which

$$\lim_{n\to\infty} \frac{f(n)}{\log n} = c (0 < c < \infty). \tag{7}$$

I offer 500 dollars for a proof or disproof of (7) and I offer 100 dollars for a proof of (5) by an explicit construction. Sárközy and I proved that

$$\frac{(f(n) - c \log n)}{(\log n)^{1/2}}$$

cannot tend to 0. Sárközy and I further proved that if $g(n)$ is an increasing function of regular and slow growth, then there is a sequence A for which

$$\frac{f(n)}{g(n) \log n} \to 1.$$

Perhaps (7) can be strengthened in the following way: There is an absolute constant C so that in (6) $c_2 > (1 + C)c_1$, that is, c_2 and c_1 cannot be too close. Further, perhaps if

one only assumes that $a_k < ck^2$ for every k, then

$$\limsup \frac{f(n)}{\log n} > c_1, \qquad \text{for some } c_1 > 0. \tag{8}$$

If true, (8) is probably very deep. An old conjecture of Turán and myself (for which I often offered 500 dollars) states that if A is a basis of order 2, then

$$\limsup f(n) = \infty \tag{9}$$

that is, if $f(n) > 0$ for all $n > n_0$, then $f(n)$ cannot be bounded. Perhaps (9) follows already if we only assume that $f(n) > 0$ holds for a sequence of density (or perhaps only upper density) 1. The following question is perhaps of some interest: Is it true that if $f(n) > 0$ for a sequence of positive density, then $\limsup f(m) = \infty$? The following finite form of this question seems interesting to me: Let $a_1 < a_2 < \cdots < a_t \le n$ be a sequence of integers for which the number of integers $m < n$ for which $f(m) = 0$ is less than δn. Is it then true that for some u, $f(u) > r$. If this is true we probably can also assume $u < n$. I have to apologize to the reader if I overlooked a trivial counterexample, but I didn't have time to consider these questions seriously. This conjecture seems difficult and it certainly is much weaker than (4). The analogous conjecture for multiplicative representation is true and is not very hard to prove. Rényi and I proved by probabilistic methods that there is a sequence A satisfying $a_k < ck^2$ for which

$$\sum_{n=1}^{X} f(n)^2 < CX. \tag{10}$$

On the other hand, perhaps if A is a basis of order 2, then

$$\frac{1}{X}\sum_{n=1}^{X} f(n)^2 \to \infty. \tag{11}$$

If it is true, no doubt (11) is very difficult; it is very much sharper than (9), and even (9) is probably deep. Equation (11) has recently been disproved by I. Ruzsa.

By the way, it is a simple exercise to construct a sequence A for which every positive integer n can be written uniquely in the form $a_j - a_i$. It is not known if there is such a sequence for which $a_k/k^{2+\epsilon}$ or even $a_k/k^3 \to 0$. The greedy algorithm easily gives such a sequence for which $a_k < ck^3$. It is not difficult to prove that if $a_k < ck^2$ for every k, then the number of solutions of $n = a_j - a_i$ cannot be bounded. This last statement follows from the fact that if $a_k < ck^2$ and $F(n)$ denotes the number of solutions of $a_j - a_i < n$, then $F(n)/n \to \infty$.

Denote by $A_r(n)$ the largest integer l for which there is a sequence $1 \le a_1 < a_2 < \cdots < a_l \le n$ for which all the sums

$$\sum d_i a_i, \sum_i d_i \le r, \qquad d_i \ge 0$$

are all distinct. Turán and I proved

$$A_2(n) < n^{1/2} + cn^{1/4}. \tag{12}$$

Let

$$1 \le a_1 < a_2 < \cdots < a_r \le n,\ t \ge c_r n,\ n > n_0(r). \tag{12'}$$

Determine the smallest c_r for which (12') implies that there is some m for which $f(m) \ge r$. Or one could also ask for the smallest c_r' for which $t_r \ge c_r' n^{1/2}$ implies that for some m, $m = a_i - a_j$ has at least r solutions. It is easy to see that both c_r and c_r' are $O(r^{1/2})$. Perhaps the following question is more interesting: Let $1 \le a_1 < \cdots < a_k \le n$ be such that the number of distinct integers of the form $a_i + a_j$ is $(1 + o(1))\binom{k}{2}$. I can show that $k \ge (1 + o(1))(2/\sqrt{3})n^{1/2}$ is possible. To see this, let $1 \le b_1 < b_2 < \cdots < b_l \le n/3$ be such that the sums $b_i + b_j$ are all distinct and l is maximal. Add to this sequence the numbers $n - b_1 > \cdots > n - b_l$. The b's and $n - b$'s clearly have the required properties. I do not see whether $k > (1 + \epsilon)n^{1/2}\ (2/3^{1/2})$ is possible. Lindstrom improved (12) to $A_2(n) < n^{1/2} + n^{1/4} + 1$. Chowla and I observed that Singer's perfect difference set implies $A_2(n) > n^{1/2}$, if $n = p^2 + p + 1$, p prime or a power of a prime and $A_2(n) > (1 + o(1))n^{1/2}$ for all n.

Bose and Chowla by an extension of Singer's construction proved that for every r

$$A_r(n) > (1 + o(1))n^{1/r}.$$

Bose and Chowla observed that our proof with Turán does not seem to give

$$A_r(n) < (1 + \epsilon)n^{1/r} \tag{13}$$

for $r > 2$. Inequality (13) very likely holds for every r, but as far as I know is still open. The difficulty in proving (13) is that in the proof of (12) one could replace $a_i + a_j$ by $a_i - a_j$ and as stated, if $a_k < ck^2$, then $F(n)/n \to \infty$ if $F(n)$ denotes the number of solutions of $a_i - a_j < n$. It seems quite impossible to use this method for odd r, and there are difficulties for even r if $r > 2$. We all believe that (13) remains true even if we only allow sums for which $d_i = 1$ (i.e., sums $a_{i_1} + a_{i_2} + \cdots + a_{i_r}$), but at the moment I do not see how to attack this question.

Further I cannot prove that if A is an infinite sequence satisfying $a_k < ck^3$ for every k, then the triple sums $a_i + a_j + a_l$ cannot all be distinct. Inequality (13) and this are perhaps the most outstanding open questions here. Turán and I conjectured

$$A_2(n) = n^{1/2} + O(1). \tag{14}$$

I was always convinced that (14) holds, but H. Taylor and I. Ruzsa independently of each other convinced me that the upper bound in (14) is probably false. If $n = p^2 + p + 1$, Singer showed that there are $p + 1$ residues (mod $p^2 + p + 1$), say $a_1, a_2, \ldots, a_{p+1}$, for which every nonzero residue can be uniquely expressed in the form $a_i - a_l$. If for every C there is such a set for which for some i we have $a_{i+1} - a_i > Cp^{1/2}$, $p > p_0(C)$, then the upper bound in (14) fails. Perhaps (14) should for the time being be replaced by the more modest conjecture: for every $\epsilon > 0$

$$A_2(n) = n^{1/2} + O(n^{\epsilon}). \tag{14'}$$

When I first met Sidon in 1932 he also asked me the following problem: A sequence $a_1 < a_2 < \cdots$ is called a $B_r^{(0)}$ sequence if the number of solutions

$$\sum_i \epsilon_i a_i = n, \qquad \sum_i \epsilon_i \le r$$

is $\leq t$. In particular, in a B_2 ($B_2^{(1)}$) sequence the sums $a_i + a_j$ are all distinct. How slowly can the terms of a B_2 (or more generally of a $B_r^{(1)}$) sequence increase? It is easy to see that (by the greedy algorithm) there is a B_2 sequence for which $a_k < ck^3$; $a_k = o(k^3)$ was open for a long time, and finally a few years ago Ajtai, Komlós, and Szemerédi proved that there is a B_2 sequence for which $a_k < ck^3/\log k$. This seems to be the natural boundary of their very ingenious method. Very likely there is a B_2 sequence for which $a_k < k^{2+\epsilon}$ for $k > k_0(\epsilon)$. At the moment this seems rather hopeless. Rényi and I proved by the probability method that for $t = t(\epsilon)$ there is a $B_2^{(t)}$ sequence for which $a_k < k^{2+\epsilon}$ holds. I proved that $\liminf a_k/k^2$ can be finite for a B_2 sequence, but that for infinitely many k we must have

$$a_k > ck^2 \log k. \tag{15}$$

It would be very interesting to decide whether (15) can be improved. In several forthcoming (I hope not posthumous) papers Sárközy, V. T. Sós, and I hope to discuss many further related problems and results. Here I want to mention just one more problem that was first mentioned in a joint paper with Nathanson and that we "rediscovered" later: Let A be any sequence of integers, and denote by $B(X)$ the number of integers $n < X$ for which $f(n) \neq 1$, that is, for which $f(n) = 0$ or $f(n) > 1$. It is not hard to show that for every $\epsilon > 0$ there is a sequence A for which $B(X) = o(X^{1/2+\epsilon})$. We conjectured

$$B(X)/X^{1/2} \to \infty. \tag{16}$$

Ruzsa stated that he proved $B(X) > X^{1/3}$ and Szemerédi claimed $B(X) > X^{1/2-\epsilon}$ for every $\epsilon > 0$ if $X > X_0(\epsilon)$. Unfortunately none of these proofs have been published.

For many of the results stated see the excellent book [21]; also [1] and [13].

For many references on these and related problems see the very interesting papers of [3], [15], [17].

6. Now I will discuss some of my work with M. Nathanson [11, 12]. A sequence A is called an asymptotic basis of order r if every $n > n_0$ is the sum of r or fewer of the a's. We will mostly restrict ourselves to the case $r = 2$. An asymptotic basis is called minimal if the omission of any of its elements destroy the basis property, that is, if for every i there are infinitely many integers n for which n cannot be represented as the sum of 2 (or more generally r) terms from $A - a_i$. Nathanson proved that there are minimal asymptotic bases for every r and that there are asymptotic bases that do not contain minimal asymptotic bases. We proved that if $f(n) > c \log n$, $c > \log 4/3$, then A contains a minimal asymptotic basis of order 2. On the other hand, we proved that for every k there is a sequence A for which $f(n) \geq k$ for all $n > n_0$, but A does not contain a minimal asymptotic basis. Our most interesting open problems are: Is there an A for which $f(n) \to \infty$ and that does not contain a minimal asymptotic basis? Is it true that if $f(n) > \epsilon \log n$ for some $\epsilon > 0$, then A contains a minimal asymptotic basis?

In a forthcoming paper we show that if A is such that $f(n) > C \log n$ where C is a sufficiently large constant, then A can be decomposed as the union of two disjoint sequences $A_1 \cup A_2$ so that both A_1 and A_2 are asymptotic bases. Perhaps this remains true if $f(n) > C \log n$ is replaced by $f(n) > \epsilon \log n$, but perhaps there is an A for which

$f(n) \to \infty$, but A cannot be decomposed as the disjoint union of two bases A_1 and A_2. (We know that $f(n) > k$ does not imply such a decomposition.)

Perhaps if $A_1 \cap A_2$ is empty and A_1 and A_2 are both asymptotic bases of order 2, then $A_1 \cup A_2$ always contains a minimal asymptotic basis. Unfortunately, at the moment we can neither prove nor disprove this attractive conjecture.

Define a basis to be thin if $A(n) < cn^{1/2}$ for some c. Clearly, if a basis is not thin, then (**11**) holds. We hoped that every thin minimal basis can be decomposed as the union of r (perhaps $r = 2$) disjoint sets A_i, $\bigcup_{i=1}^{r} A_i = A$, so that for every i, $i = 1, 2, \ldots, r$ the set of integers of the form $x + y$, $x, y \in A_i$ has density 0. Originally I stated this conjecture without the assumption that the basis is minimal. Volkmann pointed out that there are trivial counterexamples, but the conjecture can perhaps be saved if we assume that the basis is minimal. (For a similar error, see [10, p. 47].) This conjecture is perhaps completely wrongheaded. For the proof of (**11**) it would suffice to show that every thin basis A has a subset A' for which $A'(X) > cX^{1/2}$ for every X and such that the density of the integers of the form $x + y$, $x, y \in A'$ is 0. This conjecture could also be completely wrong, and our excuse for making it is that we know so few thin bases (in fact, I. Ruzsa disproved this conjecture).

7. Now I will discuss some special problems on additive number theory, some of which have been partially settled during our recent meeting at Hakone.

Several years ago Silverman and I asked: Let $1 \le a_1 < a_2 < \cdots < a_{t_n} \le n$ be a sequence of integers such that none of the sums $a_i + a_j$ are squares. We at first thought that perhaps $\max t_n = n/3$, and that this is reached if $a_i = 1 + 3i$. Massias soon showed that in fact $\max t_n \ge (11/32)n$ by finding 11 residue classes mod 32, no two of which add up to a quadratic residue. I then conjectured that $\max t_n = (11n/32) + O(1)$. Lagarias, Odlyzko, and Shearer [22] proved this if the a's are residue classes mod d for some d. Their proof is quite difficult. They further proved that if $1 \le a_1 < a_2 < \cdots < a_{t_n} \le n$ is any sequence of integers for which $a_i + a_j$ is never a square, then

$$\max t_n < 0.475n. \tag{17}$$

Thus t_n is definitely less than the trivial bound $n/2$, but we are very far from the perhaps too optimistic conjecture $(11/32)n + O(1)$. Then I posed the following perhaps interesting question: Let $n_1 < n_2 < \cdots$ be an infinite sequence of integers and $a_1 < a_2 < \cdots$ an infinite sequence of integers so that $a_i + a_j \ne n_k$ for all i, j, k. For which $n_1 < n_2 < \cdots$, does it then follow that

$$\limsup A(x)/x < \frac{1}{2}? \tag{18}$$

If the n's increase sufficiently fast, then $\limsup A(x)/x = \frac{1}{2}$ is certainly possible. Perhaps $n_{k+1}/n_k \to 1$ implies (**18**). (If the n's are uniformly distributed in the residue classes, all n_1 odd are an obvious counterexample.)

I recently asked: How many integers $1 \le a_1 < a_2 < \cdots < a_{t_n} \le n$ can one give so that none of the subset sums $\sum_{i=1}^{t_n} \epsilon_i a_i$, $\epsilon_i = 0$, or 1 are squares? It is easy to see that

$$t_n > (1 + o(1))2^{1/3}n^{1/3} \tag{19}$$

is possible. To see this, let p be the least prime greater than $n^{2/3}2^{-1/3}$ and let $a_i = ip$, $i < 2^{1/3}n^{1/3}$. The sum of these a's is clearly never a square since all subset sums are

multiples of p and none of p^2. I could never get a better lower bound than (**18**), but could not even prove $t_n - o(n)$. Noga Alon [2] showed

$$t_n < \frac{cn}{\log n}. \tag{20}$$

Both (**18**) and (**19**) are probably far away from the truth. Freud and I found an infinite sequence $a_n < n^\alpha$, α about 10, so that none of the subset sums are squares. Ten is very far from the truth. We would have liked to show at least $a_n < cn^3$, but were unsuccessful. (Lipkin, Noga Alon, and Freiman have many results on these problems.)

Freud and I then posed the following two seemingly trivial problems, which as far as I know are still open: Let $1 \le a_1 < a_2 < \cdots < a_{x+1} \le 3x$. Is it true that some subset sum $\Sigma_{i=1}^{x+1} \epsilon_i a_i$, $\epsilon_i = 0$ or 1 must be a power of 2. The multiples of 3 show that if true, this is best possible. (This was proved by Freiman.)

Let $1 \le a_1 < \cdots < a_{y+1} \le 4y$. Then some subset sum must be squarefree. The multiples of 4 show that if true the result is best possible, and perhaps the sum of 2 or at most three of the a's will already have the required properties. Clearly, many related problems can be stated, but we had no success with any of them and perhaps we overlook a trivial point. (A very simple proof was found by Filaseta.)

R. L. Graham and I very recently asked: Is it true that if $n > n_0(c)$ and

$$1 \le a_1 < a_2 < \cdots < a_t \le n, \qquad t > (\tfrac{1}{2} - \epsilon)n, \tag{21}$$

then $2n$ is a subset sum of the a's? That is, $2n = \Sigma_{i=1}^{t} \epsilon_i a_i$ $\epsilon_i = 0$ or 1. We thought that perhaps $t = (n/3) + 1$ (if true, this is clearly best possible). During the meeting in Hakone, Noga Alon proved (**21**) with $t = 2/5n$ (he now improved this to $t = (1/3 + \epsilon)n$). Graham and I also asked: Let $f(n)$ be the largest integer so that for every m there is a set of $f(n)$ integers $1 \le a_1 < a_2 < \cdots < a_t \le n$, $t = f(n)$ so that m is not a subset sum of the a's. We conjectured

$$f(n) - \left(\frac{1}{2} + o(1)\right)\frac{n}{\log n}. \tag{22}$$

We showed

$$f(n) > \left(\frac{1}{2} + o(1)\right)\frac{n}{\log n}. \tag{23}$$

To show (**23**) observe that we can assume $m < 1 + 2 + \cdots + n - (n^2 + n)/2$. Let p be the least prime that does not divide m. From the prime number theorem, we have $p < (2 + o(1))\log n$ and if we put $a_i = ip$, $1 \le i < n/(2 \log n)$, then m is not the subset sum of the a's, which proves (**23**). During our meeting Noga Alon proved $f(n) < (cn/\log n)$; thus (**23**) is not very far from the truth. It would also be of interest to know how to choose $m = m(n)$ to obtain the value of $f(n)$ [2].

We also briefly discussed the following more general problem. Let $f_r(n)$ be the largest integer so that for every choice of $m_1, \ldots, m_r$ there is a set of $f_r(n)$ integers $1 \le a_1 < a_2 < \cdots < a_{t_n} \le n$, $t_n = f_r(m)$ so that none of the subset sums $\Sigma_{i=1}^{t_n} \epsilon_i a_i$, $\epsilon_i = 0$ or 1 is one of the m_i, $i = 1, 2, \ldots, r$. It seems interesting to determine how $f_r(n)$ decreases as

r increases and how we have to choose $m_1, \ldots, m_r$ to get $f_r(n)$. Our old $f(n)$ is of course $f_1(n)$.

8. Now I would like to discuss a few older problems that seem to have been perhaps undeservedly neglected. Not to make the paper too long, I will restrict myself to a few of them.

I. Sárközy and I [14] considered the following question: Let $A = \{a_1 < a_2 < \cdots\}$ be an infinite sequence of integers, and assume that no a_i divides the sum of two larger a's. What can be said about $A(x) = \sum_{a_i < x} 1$? The reason that we restricted ourselves to $a_i \nmid (a_j + a_l), i < j, i < l$ was that we wanted to exclude the case $2a_i = a_j + a_l$, which leads to the well-known problem of three-term arithmetic progressions. We proved

$$A(x) = o(x) \tag{24}$$

and that (24) is best possible, that is, if $f(x) \to \infty$ as slowly as we please, then $A(x) = o(x/f(x))$ does not have to hold. On the other hand, we conjectured that for infinitely many x, perhaps $A(x) = o(x/\log x)$ or even $A(x) < x^{1-\epsilon}$ will hold, and we further conjectured that $\sum 1/a_i$ must converge. It is rather annoying that we could not prove this rather attractive conjecture. Perhaps even more annoying is the finite version of our problem: Let $a_1 < a_2 < \cdots < a_{t_n} \leq n$ be a sequence of integers and assume that $a_i \nmid (a_j + a_l)$ for every $i < j < l$. One would expect that then

$$\max t_n = \frac{n}{3} + 1. \tag{25}$$

Equation (25), if true, is best possible. This can be seen by the integers $(2n/3) < a_i \leq n$. This conjecture, if true, should not be hard to prove and perhaps we overlooked a simple argument. The following related problems should perhaps be mentioned. Assume that the finite sequence A is such that no a_i divides the sum of r or fewer a_j, $a_j > a_i$. Here one would guess that

$$\max A_r(n) = \frac{n}{r+1} + O(1),$$

where the extremal sequence is given by the integers $n, n-1, \ldots, n - t_r$, where t_r is the largest integer for which our property is satisfied. The same conjecture can be made if we assume that no a divides any sum of larger a's. Here one would assume that the largest set of integers with this property is given by the integers in $\{n - t, n\}$ with $\binom{t}{2} < n$ if $A(n) = \sqrt{2n} + O(1)$. $A(n) < c\sqrt{n}$ can probably be proved by the methods of Szemerédi and Olson. See also [14] and [26].

II. Szemerédi and I [16] considered the following problem: Let $a_1 < a_2 < \cdots < a_n$ be any set of n distinct integers. Denote by $b_1 < b_2 < \cdots < b_{m_n}$ the set of all integers that can be written in the form $a_i + a_j$ or $a_i a_j$. Determine or estimate as accurately as possible $\min m_n$. It seemed to us that if there are few distinct integers of the form $a_i + a_j$, then there must be many distinct integers of the form $a_i a_j$ and vice versa. We conjectured $m_n > n^{2-\epsilon}$ for every $\epsilon > 0$ if $n > n_0(\epsilon)$, and proved that for a certain $c > 0$

$m_n > n^{1+c}$, but that there is a sequence $a_1 < a_2 < \cdots < a_n$ for which

$$m_n < n^{2-c/\log\log n}.$$

The upper bound seemed to us to be close to the truth.

For further problems of a related nature, see [16].

III. Divide the integers $1, 2, \ldots, 2n$ into two disjoint sets $a_1, a_2, \ldots, a_n; b_1, b_2, \ldots, b_n$ with n elements in each class. Denote by M_k the number of solutions of $a_i - b_j = k$ and put

$$M = M(n) = \min \max_k M_k$$

where the maximum is to be taken for all $-2n \le k \le 2n$ and the minimum for all the $\binom{2n}{n}$ divisions of the integers into two disjoint classes with both having n elements. I asked for the determination or estimation of M more than 30 years ago. The best upper bound is still $M < 0.4n$. The best lower bound is due to L. Moser [23]

$$M > \sqrt{4 - \sqrt{15}}(n-1) > 0.3570(n-1).$$

As far as I can tell, the problem has been completely (perhaps undeservedly) forgotten.

IV. Let $a_1, a_2, \ldots, a_n$ be n real numbers, all different from 0. Denote by $f(n)$ the largest integer so that for every sequence $a_1, a_2, \ldots, a_n$ one can select $k = f(n)$ of them $a_{i_1}, \ldots, a_{i_k}$ so that the sum of two of these a's never equals a third. I proved [7]

$$f(n) \ge \frac{n}{3}. \tag{26}$$

Very likely (26) is not best possible, perhaps $f(n) \ge 3n/7$. If in the representation $a_u + a_v = a_w$ we insist that $u \ne v$, then we could hope that

$$f(n) \ge \left[\frac{n+2}{2}\right]. \tag{27}$$

It is surprising that this problem has been completely forgotten, even by myself. A few years ago Ruzsa independently rediscovered the conjecture; I told him what a nice problem, how silly that I did not think of it myself.

For many other problems on combinatorial number theory, see [6], [8], [9].

9. Finally I present a very short discussion of some of the questions connected with van der Waerden's theorem. Denote by $W(n)$ the smallest integer for which if we divide the integers from 1 to $W(n)$ into two classes, at least one of the classes contains an arithmetic progression of n terms. van der Waerden proved about 60 years ago the now classic result that $W(n)$ exists for every n. van der Waerden's upper bound for $W(n)$ is probably very far from the truth; it tends to infinity as fast as the Ackermann's function. The first and most important task would be to prove or disprove that $W(n)$ does not tend to infinity so fast, that is, to give better upper bounds for $W(n)$. This has recently been accomplished by Shelah; his upper bound is probably still far too large.

More than 50 years ago this problem lead Turán and me to conjecture that van der Waerden's theorem is really a density theorem and that in fact every sequence of positive density contains arbitrarily long arithmetic progressions. More precisely: Denote by $r_k(n)$ the largest integer for which there is a sequence $1 \leq a_1 < a_2 < \cdots < a_t \leq n, t = r_k(n)$ so that the a's do not contain an arithmetic progression of k terms. We conjectured that for every k

$$r_k(n) = o(n). \tag{28}$$

More than 30 years ago K. F. Roth proved

$$r_3(n) < \frac{cn}{\log \log n}. \tag{29}$$

Heath-Brown certainly proved $r_3(n) < en/(\log n)^{\epsilon}$, $\epsilon > 0$ is a small positive number, and Szemerédi and Pintz proved $r_3(n) < n/(\log n)^{1/4}$.

Some time ago I offered 1000 dollars for a proof or disproof of (**28**). About 15 years ago Szemerédi proved (**28**). His proof, which is a masterpiece of combinatorial reasoning, already had many applications for many other combinatorial problems. A few years later Fürstenberg proved (**28**) by using methods of ergodic theory; no doubt his method will have many more applications in combinatorial number theory and other subjects.

About 30 years ago I conjectured that if $\sum_{n=1}^{\infty} 1/a_n = \infty$, then the a's contain arbitrarily long arithmetic progressions. This conjecture would in particular imply that the primes contain arbitrarily long arithmetic progressions.

I offer 3000 dollars for a proof or disproof of my conjecture. It is not even known that if $a_1 < a_2 < \cdots$ contains no three-term arithmetic progression, then $\sum_{i=1}^{\infty} 1/a_i < \infty$. This would of course follow if the following were in fact true

$$r_3(n) < \frac{n}{(\log n)^{1+\epsilon}}.$$

Perhaps, in fact, for every k and l we have for $n > n_0(k, l)$

$$r_k(n) < \frac{n}{(\log n)^l}. \tag{30}$$

If true, (**30**) would of course imply my 3000 dollar conjecture. Unfortunately, none of the results so far give an improvement of the upper bound of $W(n)$ given by van der Waerden. The first exponential lower bound for $W(n)$ was given by Rado and myself, and our bound was later improved by W. Schmidt. The current best lower bounds are due to Berlekamp, Lovász, and myself. Berlekamp proved that for every prime p we have

$$W(p + 1) > p2^p, \tag{31}$$

and Lovász and I proved that for every n

$$W(n) > \frac{c2^n}{n}. \tag{32}$$

The first step (which perhaps will not be too hard) is to prove

$$\frac{W(n)}{2^n} \to \infty. \tag{33}$$

We all believe that

$$W(n)^{1/n} \to \infty,$$

but even $W(n) > (2 + \epsilon)^n$ for $n > n_0(\epsilon)$ seems to be beyond our reach at present.

Define $W(k, l)$ as the smallest integer so that if we split the integers $1 \le n \le W(k, l)$ into two classes, either the first class contains an arithmetic progression of k terms or the second class contains an arithmetic progression of l terms. $W(n, n) = W(n)$. These van der Waerden numbers have, as far as I know, not been considered at all. Observe that if $a_1 < a_2 < \cdots < a_t \le n$ does not contain an arithmetic progression of k terms, then $t < r_k(n)$, and thus the complementary set of the a's contains an arithmetic progression of length at least $n/[r_k(n)]$, or

$$W\left(k, \frac{n}{r_k(n)}\right) < n. \tag{34}$$

Is it true that $W(k, l)$ is significantly less than the trivial bound given by (34)? The simplest problem is: Let $1 \le a_1 < \cdots < a_t \le n$ be such that it does not contain a three-term arithmetic progression. Is it then true that the complementary sequence contains an arithmetic progression of more than $(1 + c)n/r_3(n)$ terms?

Freud and I tried to investigate the following problem: Let $h(n)$ be the largest integer for which if we decompose the integers $1, 2, \ldots, n$ into $h(n)$ disjoint sets $A_1, \ldots, A_{h(n)}$, then there are always two of them, say A_i and A_j, whose union contains a four-term arithmetic progression. It would be of interest to determine or estimate $h(n)$ as accurately as possible. Clearly, many generalizations are possible, but we had no success even with $h(n)$.

To end this paper I would like to mention one of my oldest conjectures, which was inspired by van der Waerden's theorem and which is about 55 years old. Divide the integers into two classes in an arbitrary way, or define a function $f(n)$ that is either $+1$ or -1. Put

$$g(d, m) = \sum_{k=1}^{m} f(kd).$$

Is it true that

$$\limsup_{d,m} g(d, m) = \infty?$$

The connection with van der Waerden's theorem is clear. I restrict the arithmetic progressions much more, but we demand much less. We do not demand that all terms be in the same class, but only that one of the classes should have a majority that is unbounded. A more precise quantitative formulation of the conjecture is: There is an absolute constant $c > 0$ so that for some d and m

$$\left| \sum_{\substack{k=1 \\ md<n}}^{m} f(kd) \right| > c \log n. \tag{35}$$

It is easy to see that (**35**) if true is best possible. A weaker form of the conjecture states: Let $f(n)$ be a multiplicative function that only assumes the values ± 1 [i.e., $f(ab) - f(a)f(b)$]. Is it then true that

$$\left| \sum_{k=1}^{n} f(k) \right|$$

is unbounded? This was conjectured independently also by Chudakov. Perhaps for every e the density of integers n for which $|\Sigma_{k=1}^{n} f(k)| < c$ is 0. See [26] and [27].

REFERENCES

1. AJTAI, M., J. KOMLÓS & E. SZEMERÉDI. 1981. A dense infinite Sidon sequence. Eur. J. Comb. **2:** 1–11.
2. ALON, N. 1987. Subset sums. J. Number Theory **27:** 196–205.
3. BABAI, L. & V. T. SÓS. 1985. Sidon sets in groups and induced subgraphs of Cayley graphs. Eur. J. Comb. **6:** 101–114.
4. ERDŐS, P. 1935. Note on sequences of integers no one of which is divisible by any other. J. London Math. Soc. **10:** 126–128.
5. ERDŐS, P. 1969. On some applications of graph theory to number theoretic problems. Publ. Ramanujan Inst. **1:** 131–135.
6. ERDŐS, P. 1984. On some of my problems in number theory I would most like to see solved. *In* Proceedings of Octacamunel India. Lecture Notes in Math 1122. Number Theory, K. Alladi, Ed.: 79–84. Springer-Verlag. New York/Berlin.
7. ERDŐS, P. Extremal problems in number theory. *In* Proceedings of the Symposium on Pure Mathematics, Vol. VIII: 181–189. Am. Math. Soc. Providence, R.I.
8. ERDŐS, P. Problems and results on combinatorial number theory. *In* Number Theory Day. Lecture Notes in Math 626, M. B. Nathanson, Ed.: 43–79. Springer-Verlag. New York/Berlin.
9. ERDŐS, P. Some new problems and results in number theory. *In* Mysore Conference. Lecture Notes in Math 938, K. Alladi, Ed.: 50–74. Springer-Verlag. New York/Berlin.
10. ERDŐS, P. & R. L. GRAHAM. 1980. Old and new problems and results in combinatorial number theory. Enseign. Math. **28.**
11. ERDŐS, P. & M. NATHANSON. 1979. Systems of distinct representatives and minimal bases in additive number theory. *In* Number Theory Carbondale. Lecture Notes in Math 751, M. Nathanson, Ed.: 81–107. Springer-Verlag. New York/Berlin.
12. ERDŐS, P. & M. NATHANSON. 1988. Partitions of bases into disjoint unions of bases. J. Number Theory **29:** 1–9.
13. ERDŐS, P. & A. RÉNYI. 1960. Additive properties of random sequences of positive integers. Acta Arith. **6:** 83–110.
14. ERDŐS, P. & A. SÁRKÖZY. 1970. On the divisibility properties of sequences of integers. Proc. London Math. Soc. **21**(3): 97–101.
15. ERDŐS, P. & A. SÁRKÖZY. 1985. Problems and results on additive properties of general sequences I. Pac. J. Math. **118:** 347–357.
16. ERDŐS, P. & E. SZEMERÉDI. 1976. On multiplicative representation of integers. J. Aus. Math. Soc., Ser. A **21:** 418–427.
17. ERDŐS, P., A. SÁRKÖZY & V. T. SÓS. 1984. Problems and results on additive properties of general sequences IV. *In* Proceedings of Octacamunel India. Lecture Notes in Math 1122. Number Theory, K. Alladi, Ed.: 85–104. Springer-Verlag. New York/Berlin.
18. ERDŐS, P., A. SÁRKÖZY & E. SZEMERÉDI. 1970. Divisibility properties of sequences of integers. Number Theory Colloquium, Journal of the Bolyai Mathematical Society: 35–49. North-Holland. Amsterdam.
19. FÜRSTENBERG H. 1981. Recurrence in ergodic theory and combinatorial number theory. *In* M. B. Porter Lectures. Princeton Univ. Press. Princeton, N.J.

20. GUY, R. 1981. Unsolved Problems in Intuitive Mathematics, Vol. 1, Number Theory. Springer-Verlag. New York/Berlin.
21. HALBERSTROM, H. & K. F. ROTH. 1983. Sequences. Springer-Verlag. New York/Berlin.
22. LAGARIAS, J. C., A. M. ODLYZKO & J. B. SHEARER. 1982. On the density of sequences of integers the sum of no two of which is a square, I: Arithmetic progressions. J. Comb. Theory, Ser. A **33:** 167–185; also, ———. 1983. II: General sequences. J. Comb. Theory, Ser. A **34:** 123–139.
23. MOSER, L. 1959. On the minimal overlap problem of Erdös. Acta Arith. **5:** 117–119.
24. POMERANCE, C. 1983. On the longest simple path in the divisor graph. Congr. Numer. **40:** 291–304.
25. SZEGEDY, M. 1986. The solution of Graham's greatest common division problem. Combinatoria **6:** 67–79.
26. SZEMERÉDI, E. 1970. On a conjecture of Erdös and Heilbronn. Acta Arith. **17:** 227–229.
27. SZEMERÉDI, E. 1975. On sets of integers containing no k elements in arithmetic progression. Acta Arith. **27:** 199–245.