Annals of Discrete Mathematics 6 (1980) 89–115

A SURVEY OF PROBLEMS IN COMBINATORIAL NUMBER THEORY

Paul ERDÖS

Hungarian Academy of Science, Budapest, Hungary

During my long life I published and stated many conjectures and wrote several papers which consisted entirely of stating old and new problems. In this paper I hope to survey some of them — not to make the paper too long I restrict myself to problems on the borderline of combinatorics and number theory fields which are — as stated in a previous paper — closest to my heart or rather than to my brain. Again, to avoid excessive length, I restrict myself as much as possible to problems which either were raised by my collaborators and myself, or we worked on them. I do not want to imply that these problems are more interesting or important than the ones I neglect, but I am likely to know more about my own problems than the reader and finally, despite my great age, my memory and mind are still more or less intact. Thus I remember how and why I or we came to consider these problems, and what are the connections with other questions.

First, I list some of my papers in which relevant problems are stated.

(i) *Problems and results on combinatorial number theory* I, II, II′ *and* III. Paper I was given at the first conference at Fort Collins (in honor of Professor Bose) also edited by Srivastava: A survey of combinatorial theory (North Holland Amsterdam, 1973) 117–138; the pair II and II′ is due to my mistake: II′ should have been III, II is in Journée Arithmetique de Bordeaux, Astérisque 24–25 (Juin 1974), 295–210, and II′ is in J. Indian Math. Soc. 40 (1976) 285–298. Paper III is in Number theory day, held at Rockerfeller Univ., edited by M. Natanson, Lecture Notes in Mathematics 626 (Springer Verlag, Berlin) 43–72.

(ii) *Some older papers of mine:* Quelques problèmes de la théorie des nombres, Monographies de l'Enseignement Mathématique No. 6, Univ. de Geneva (1963) 81–135. Graham and I will soon publish a long paper in the same journal which brings this paper up to date and give also many new problems.

P. Erdös, Some unsolved problems, Michigan Math. J. 4 (1957) 291–300 and Publ. Math. Inst. Hung. Acad. Sci. 6 (1961) 2216–254. The paper in Michigan Math. Journal was my first paper on unsolved problems.

P. Erdös, Extremal problems in number theory, Proc. Symp. in Pure Math. VIII, Theory of Numbers (Amer. Math. Soc., Providence, RI, 1965) 181–189. Several of the results stated in this paper were improved and extended in various papers by S.L.R. Choi.

P. Erdös, Some recent advances and current problems in number theory,

Lectures on Modern Mathematics, edited by L. Saaty, (Wiley, New York, 1965) III. 196–244.

P. Erdös, Some extremal problems in combinatorial number theory, Math. essays dedicated to A.J. Macyntyre, edited by L. Shankar, (Ohio Univ. Press, Athens, OH, 1970) 123–133.

P. Erdös, Some problems in number theory, Computers in number theory, conference held in Oxford 1969 (Academic Press, London, 1971) 406–414.

P. Erdös, Remarks on number theory IV and V. Extremal problems in number theory I and II. Mat. Lapok 13 (1962) 28–38; 17 (1966) 135–166 (in Hungarian).

I divide the present paper into several sections, one long one is for problems which do not seem to fit into any of the classifications. I try as much as possible to indicate what happened to the problems stated in previous papers. So as not to make the paper too long I do not give proofs and I discuss some of the old problems only if there is some recent progress or if I find them very attractive and feel that perhaps they have been neglected.

1. Van der Waerden's and Szemerédi's theorem

In this Section I discuss problems connected with Van der Waerden's and Szemerédi's theorem. These questions have been discussed adequately in my previous papers. Thus I give details only if there is some important new development.

According to Alfred Brauer, Schur conjectured more than 50 years ago that if we divide the integers into two classes at least one of them contains arbitrarily long arithmetic progressions. Van der Waerden proved this conjecture in the following stronger form: There is an $f(n)$ so that if we divide the integers $1 \leqslant t \leqslant f(n)$ into two classes, at least one of them contains an arithmetic progression of n terms. There is no satisfactory upper bound for $f(n)$ — in view of the recent surprising results of Paris and his colleagues on a modification of Ramsey's theorem. I am no longer absolutely certain that Van der Waerden's original bound can be very much improved. The best lower bounds are due to Berlekamp and Lovász and myself: $f(p) > cp2^p$ if p is a prime and $f(n) > c2^n$ for all n. I am of course sure that

$$\lim_{n=\infty} f(n)/2^n = \infty. \tag{1}$$

(1) will perhaps be not too difficult to establish, but I offer 500 dollars for the proof or disproof of

$$\lim_{n=\infty} f(n)^{1/n} = \infty. \tag{2}$$

(2) seems very difficult. In view of the difficulty in dealing with $f(n)$ I modified the problem as follows: Let $\frac{1}{2}n < u \leqslant n$. Let $f(u, n)$ be the smallest integer such that if

we divide the integers $1 \leqslant t \leqslant f(u, n)$ into two classes there always is an arithmetic progression of n terms in which one of the classes has at least u terms. the probability method easily gives $f(u; n) > (1+c(\varepsilon))^n$ for $u > \frac{1}{2}n(1+\varepsilon)$. J. Spencer determined $f(n+1; 2n)$. Otherwise nothing is known. Clearly very many interesting problems remain here; e.g. determine or estimate $f(n+k; 2n)$ for fixed k if n tends to infinity.

Analogously to the Ramsey numbers we can define the Van der Waerden numbers as follows: $f_{u,v}$ is the smallest integer so that if we divide the integers $1 \leqslant t \leqslant f_{u,v}$ into two classes either class I contains an arithmetic progression of u terms or class II contains a progression of v terms. Very little is known about these Van der Waerden numbers — in particular it is not known if $f_{3,v}$ tends to infinity polynomially or faster. Another analogy with Ramsey numbers is this: Denote by $f_{(4,3),v}$ the smallest integer so that if we divide the integers $1 \leqslant t' \leqslant f_{(4,3),v}$ into two classes then either the first class contains three numbers of an arithmetic progression of four terms or the second contains an arithmetic progression of v terms. Clearly many related questions can be asked whose formulation can be left to the reader, unfortunately so far there are practically no non-trivial bounds for any of these problems.

We can extend these functions to more than two variables — in fact this was already done in the original paper of Van der Waerden. Here we only state one problem: Let $g(l)$ be the smallest integer so that if we divide the integers $1 \leqslant t \leqslant g(l)$ into l classes at least one of them contains an arithmetic progression of 3 terms. The bound $g(l) < \exp\exp cl$ follows from Roth's result but as far as I know it has not yet been shown that $g(l)$ tends to infinity faster than polynomially.

In view of the difficulty of obtaining a satisfactory upper bound for $f(n)$ Turán and I conjectured more than 45 years ago that every sequence of positive upper density contains arbitrarily long arithmetic progressions. More precisely: Let $r_k(n)$ be the smallest integer l so that if

$$1 \leqslant a_1 < \cdots < a_l \leqslant n \tag{3}$$

then the a's contain an arithmetic progression of k terms. Our conjecture with Turán states

$$r_k(n) = o(n) \tag{4}$$

We first believed $r_k(n) < n^{1-\varepsilon_k}$, but Salem and Spencer showed $r_3(n) > n^{1-\varepsilon}$ for every $\varepsilon > 0$ and $n > n_0(\varepsilon)$. I realised the real difficulty of (4) only after this result of Salem and Spencer. Behrend and Roth proved

$$n \exp(-c_1(\log n)^{\frac{1}{2}}) < r_3(n) < \frac{c_2 n}{\log\log n} \tag{5}$$

Yudin states without proof (Abstracts of Number Theory meeting in Vilnius 1974) that he can get $r_3(n) < cn/\log n$.

Finally Szemerédi proved (4); his proof is a masterpiece of combinatorial reasoning. I offered 1000 dollars for the proof of (4).

Recently, Fürstenberg proved (4) by using methods of ergodic theory and topological dynamics. This proof was recently simplified by Katz-Nelson, Ornstein and Varadhan. At this moment it is impossible to decide on the importance of this invasion of ergodic methods into combinatorial number theory. It is conceivable that it will be like the application of analysis to number theory, but perhaps it is too early to form an opinion — anyway the future will soon decide.

I conjectured long ago that if

$$\sum_i 1/a_i = \infty \tag{6}$$

then the a_i's contain arbitrarily long arithmetic progressions. If true this of course implies that there are arbitrarily long arithmetic progressions all whose terms are primes. I offer 3000 dollars for a proof or disproof of (6). In fact it would be very desirable to obtain an asymptotic formula for $r_k(n)$. I would be very pleased to have reasonably good upper and lower bounds for it.

Szemerédi remarks that we can not even prove that $r_k(n)/r_{k+1}(n) \to 0$ as $n \to \infty$. As far as I know this has not been proved for $k = 3$.

There is an interesting finite version of the 3000 dollar conjecture: Put

$$A_k = \max \sum_i 1/a_i$$

where the maximum is taken over all the sequences which contain no arithmetic progressions of k terms. It is not clear that $A_k < \infty$, I can not even prove $A_3 < \infty$. Gerver recently proved $A_k \geqslant (1+o(1))k \log k$. For further related questions see III.

Here I state only one more related problem which seems interesting: Is it true that to every $\varepsilon > 0$ and k there is an $n_0 = n_0(\varepsilon, k)$ so that if $n_0 < a_1 < a_2 \cdots$ is a sequence of integers which do not contain an arithmetic progression of k terms then $\sum_i 1/a_i < \varepsilon$?

J. Spencer observed that there is an $h(a)$ which increases very slowly (like the inverse function of the Van der Waerden function) so that we can divide the integers into three classes so that for every a all arithmetic progressions with first term a and all whose terms belong to the same class have length less than $h(a)$. He asked what happens for two classes? I can divide the integers into two classes so that all arithmetic progressions all whose terms are in the same class and whose first term is a are shorter than $c_1 a^{1-c_2}$, very likely this remains true for a^ε, unfortunately I have no non trivial lower bound.

A problem in measure theory very recently led Mauldin and myself to ask: Is it true that to every $c > 0$ there is a $t = t(c)$ so that if $1 \leqslant a_1 < \cdots < a_k \leqslant n$, $k > cn$ is a sequence of integers and $1 \leqslant b_1 < \cdots < b_t \leqslant n$ is an arbitrary set of t integers then the a's always contain an arithmetic progression of three terms and difference $b_i - b_j$. Probably the same result holds with arithmetic progressions of k terms.

Let $l(d)$ be an increasing function. More than 25 years ago Cohen asked: Divide the integers into two classes. Is there for some an arithmetic progression of

$l(d)$ terms and difference d? I showed that if $l(d) > cd$, the answer is negative and Petruska and Szemerédi proved that if $l(d) > cd^{\frac{1}{2}}$ the answer is also negative. They expect a negative answer for $l(d) > d^{\varepsilon}$ and think that their method of proof might settle this question. Unfortunately no lower bound for $l(d)$ is in sight.

Denote by $A(n; k)$ the largest integer so that if we divide the integers $1 \leq t \leq n$ into two classes there are at least $A(n; k)$ k-term arithmetic progressions all whose terms are in the same class. It is easy to see that

$$c_k n^2 < A(n; k) < (1 + o(1)) \frac{n^2}{2(k-1)2^{k-1}}. \tag{7}$$

The lower bound follows from Van der Waerden's theorem and the upper bound from the probability method. Perhaps one can get an asymptotic formula in (7). For Ramsey's theorem, A. Goodman and others obtained related and in some cases sharp results.

Denote by $f_k(n, \alpha)$ the largest integer so that every set of αn integers not exceeding n contains at least $f_k(n, \alpha)$ arithmetic progressions of k terms. The inequality $f_k(n, \alpha) > c(\alpha, k)n^2$ follows from Szemerédi's theorem (for $k = 3$ this was shown by Varnavides, this was before Szemerédi). It would be interesting to have an asymptotic formula for $f_k(n, \alpha)$.

Is it true that if we divide the integers into two classes then there always is a three term arithmetic progression all whose elements are in the same class and whose difference is larger than its first term? If true this is best possible. To see this put in the first class the integers $3^{2k} \leq t < 3^{2k+1}$, $t = 1, 2, \ldots$ and in the second class the other integers. Clearly none of the classes contains a four term arithmetic progression whose difference is larger than its first term.

Van der Waerden's theorem can be formulated in terms of hypergraphs as follows: Consider the hypergraph whose vertices are the integers and whose edges are the k term arithmetic progressions. This hypergraph has chromatic number infinity.

Does this remain true if we restrict ourselves to arithmetic progressions whose difference is a prime number or all whose terms are primes? Clearly many further questions could be posed.

On final question which was already stated in I: Let $h(n; k, l)$ $(l > k)$ be the smallest integer so that if the sequence $\{a_1, \ldots, a_n\}$ contains $h(n; k, l)$ arithmetic progressions of k terms then it contains at least one progression of l terms. Estimate $h(n; k, l)$ as well as possible. In particular prove

$$h(n; k, l) = o(n^2), \tag{8}$$

for every k and l, (8) is open for $k = 3$, $l = 4$. Suppose $\{a_1, \ldots, a_n\}$ contains cn^2 progressions of three terms. Perhaps it must then contain a progression of length $\varepsilon \log n$, $\varepsilon = \varepsilon(c)$, but as I just stated we can not even prove that it contains an arithmetic progression of four terms. Below some papers in this general field are listed.

E.R. Berlekamp, A construction for partitions which avoid long arithmetic progressions, Canad. Math. Bull. 11 (1968) 409–419.

E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299–315. For further literature and history of the problem see I, II, II′ and III and the paper of Szemerédi. Here I only make a historical remark which can not entirely be documented. E. Rothe in 1944 told me that his wife Dr. Hildegard Ille was given the problem of estimating $r_k(n)$ by I. Schur sometime in the 1930's. Thus perhaps Schur conjectured $r_k(n) = o(n)$ before Turán and myself.

J. Spencer, Bull. Canad. Math. Soc. 16 (1973) 464.

H. Fürstenberg, Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. Analyse Math. 31 (1977) 204–265. See also the preprint H. Fürstenberg and B. Weiss Topological dynamics and combinatorial number theory and the lecture of Jean Paul Thouvenot, La démonstration de Fürstenberg du théorème de Szemerédi sur les progressions arithmétiques, Sem. Bourbaki, Vol. 1977–78 (Février 1978) 518, 1–11.

J.L. Gerver, The sum of the reciprocals of a set of integers which no arithmetic progression of k terms, Proc. Amer. Math. Soc. 62 (1977) 211–214.

A.W. Goodman, On sets of acquaintences and strangers at any party, Amer. Math. Monthly 69 (1962) 114–120.

P. Varnavides, On certain sets of positive density, J. London Math. Soc. 34 (1959) 358–360.

G.J. Simmons and H.L. Abbott, How many three-term arithmetic progressions can there be if there are no longer ones? Amer. Math. Monthly 84 (1977) 633–635.

J. Paris, Independence result for Peano arithmetic using inner models, to appear; see also J. Paris and L. Harrington, A mathematical incompleteness in Peano arithmetic, Handbook of Mathematical logic, edited by J. Barwise, Studies in Logic and Foundation of Math., Vol. 90, (North Holland, Amsterdam, 1977) 1133–1142.

H. Fürstenberg and Y. Katz-Nelson, An ergodic theorem for commuting transformations, J. Analyse Math. 34 (1978) 275–291. See also the forthcoming book of Fürstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory (Princeton Univ. Press).

2. Covering congruences and related questions

These problems are very adequately discussed in the papers quoted in the introduction and in my forthcoming paper with R.L. Graham. Thus I make this section short.

A system of congruences

$$a_i \pmod{n_i}, \quad n_1 < \cdots < n_k \tag{1}$$

is called a covering if every integer satisfies at least one of the congruences (1). The basic unsolved problem which I formulated in 1934 states: Can n_1 be chosen arbitrarily — in particular can it be arbitrarily large? I offer 500 dollars for a proof or disproof. The current record is still $n_1 = 20$ due to S.R.L. Choi. Another question: Is it true that for every $d > 1$ there is a system (1) with $(n_i, d) = 1$?

A set of integers $1 < n_1 < \cdots < n_k$ is called a covering set if they can be the moduli of the system (1). Such a covering set is called irreducible if no subset of it is a covering. It is easy to see that there are only a finite number of irreducible covering sets of size k. How large is their number and how small (resp. large) can n_k be for an irreducible covering set $1 \leqslant n_1 < \cdots < n_k$? Also how many irreducible covering systems $1 \leqslant n_1 < \cdots < n_l \leqslant x$ (l is variable) are there? Let $1 < n_1 < \cdots < n_k \leqslant x$ be an irreducible covering system. Determine or estimate $\max \sum 1/n_i$.

Are there infinitely many integers n so that the set of divisors $d \geqslant 2$ of n form an irreducible covering system? $n = 12$ is such an n.

I conjectured that for every C there is an integer N with $\sigma(N)/N > C$ ($\sigma(N)$ is the sum of the divisors of N) so that the divisors of N do not form a covering set. J. Haight recently proved this conjecture, his paper will soon appear in Mathematica. The following extremal problem can now be posed: Put

$$f(x) = \max_{m<x} \sigma(m)/m$$

where the maximum is to be taken over all the m for which the divisors of m do not form a covering system. By Haight's theorem $f(x)$ tends to infinity as $x \to \infty$. Is it true that $f(x) = o(\log \log x)$. In other words does $f(x)$ tend to infinity much slower than $\max \sigma(m)/m$.

Some more extremal problems: Determine or estimate the maximum number j_x of covering systems $\{n_1^{(i)} < n_2^{(i)} < \cdots < n_{k_i}^{(i)}\}$ where all the n's are distinct and less than x. We do not even know that j_x tends to infinity with x since this would imply that there is a covering system for which n_1 is arbitrarily large. I expect that j_x tends to infinity very slowly.

Determine or estimate

$$\max \sum 1/m_i$$

where the maximum is taken over all the sequences $\{m_i\}$ which do not form a covering system and all terms of which are less than x. This maximum is perhaps greater than $\log x - C$; if not then again there is a covering system for which n_1 is as large as we please.

Romanoff proved in 1934 that the density of integers of the form $2^k + p$ is positive. In a letter he then asked me if there are infinitely many odd numbers not of this form. Using covering congruences (see my paper in Summa Bras. Math.). I proved that there are arithmetic progressions of odd numbers $u \pmod v$ and a finite set of primes $P = \{p_1, \ldots, p_k\}$ so that for every $n = u \pmod v$ every number $n - 2^l$ is divisible by one of the primes P. Consider all the arithmetic progressions

no term of which is of the form 2^k+p — they probably are all included in the progressions determined by covering congruences. Is it true that almost all odd integers not contained in any of these progressions are of the form 2^k+p?

If j_x tends to infinity — or what is the same there are systems (1) with n_1 as large as we please then for every r there is an arithmetic progression no term of which is the form 2^k+Q_r where Q_r has at most r prime factors. Surely every large n is of the form 2^k+Q_r, $r<\log\log n$, but I have not been able to prove this.

Are there infinitely many odd integers n for which $n-2^u$, $1\leqslant 2^u\leqslant n$ is never squarefree? In fact is there any such an integer?

One more question which can be formulated in terms of covering congruences. Is there a finite set $p_1,\ldots,p_k$ of primes and an infinite sequence of integers $\{n_i\}$ so that all positive numbers of the form $n-2^u-2^v$ are multiples of the least one of the p_i $(i=1,\ldots,k)$? I conjecture with some trepidation that the answer is negative. With even more trepidation, I conjecture that almost all odd numbers are of the form $p+2^u+2^v$ and that for some r every integer is of the form $p+2^{u_1}+\cdots+2^{u_r}$. A theorem of Gallagher asserts that for $r>r_0(\varepsilon)$ the lower density of the integers of this form is $>1-\varepsilon$; Crocker proved that there are infinitely many odd integers not of the form $p+2^u+2^v$. Covering congruences can clearly be generalised for groups and many interesting problems can be posed, Schinzel used covering congruences to study the irreducibility of polynomials, he and Selfridge made interesting observations on the existence of covering congruences.

A system of congruences

$$a_i \pmod{n_i}, \quad n_1<\cdots<n_k \tag{2}$$

is called disjoint if there is no integer which satisfies two of the congruences (2). It would be of some interest to find necessary and sufficient conditions on the $\{n_i\}$ which would imply that a disjoint system (2) exists for suitable a_i. Stein and I asked: Determine or estimate $\max k=f(x)$ where the maximum is extended over all the disjoint systems (2). Szemerédi and I proved that

$$x\exp(-(\log x)^{\frac{1}{2}+\varepsilon})<f(x)<\frac{x}{(\log x)^c}.$$

The lower bound we obtained with the help of Stein. We believe that the lower bound is closer to the true order of magnitude of $f(x)$.

One of the few theorems on covering congruences was proved by L. Mirsky and D. Newman. A disjoint system can never be covering, or (2) implies $\sum 1/n_i<1$. Put

$$\varepsilon_m=\max\sum 1/n_i$$

where the maximum is taken over all disjoint systems for which $n_1>m$. Determine or estimate ε_m as well as possible. I could not even decide whether $\varepsilon_m\to 0$ as $m\to\infty$.

Some further work is contained in the following papers:

P. Erdös and E. Szemerédi, On a problem of Erdös and Stein, Acta Arithmetica 15 (1968) 85–90.

A. Schinzel, Reducibility of polynomials and covering systems of congruences, Acta Arithmetica 13 (1967) 91–101.

For a further extremal problem see: B.B. Crittenden and C.L. Van der Eynden, Any k arithmetic progressions covering the first 2^k integers cover all integers, Proc. Amer. Math. Soc. 24 (1970) 475–481.

P. Erdös, On systems of congruences (in Hungarian), Mat. Lapok 3 (1952) 122–128; On the integers of the form $p+2^k$ and some related problems, Summa Basil. Math. 11 (1950) 113–123.

S.L.G. Choi, Covering the set of integers by congruence classes of distinct moduli, Math. Comp. 25 (1971) 885–895.

R. Crocker, On the sum of a prime and of two powers of two, Pacific J. Math. 36 (1971) 103–107.

J.A. Haight, Covering systems of congruences. A negative result, Mathematika 26 (1979) 53–61.

3. Some problems in additive number theory

Here are some of my oldest problems and conjectures. Unfortunately not much progress has been made with these questions in the last few years.

Let $1 \leqslant a_1 < \cdots < a_k \leqslant x$ be a sequence of integers. Assume that the sums $\sum_{i=1}^{k} \varepsilon_i a_1$, $\varepsilon_i = 0$ or 1 are all different. Put max $k = F(x)$. I conjectured more than 45 years ago that

$$F(x) = \frac{\log x}{\log 2} + O(1) \tag{1}$$

A simple counting argument gives

$$F(x) < \frac{\log x}{\log 2} + \frac{\log \log x}{\log 2} + O(1)$$

and L. Moser and I using second moments improved this to

$$F(x) < \frac{\log x}{\log 2} + \frac{\log \log x}{2 \log 2} + O(1) \tag{2}$$

As far as I know (2) has never been improved.

I asked: Is it true that $f(2^k) \geqslant k+2$ for $k > k_0$? (Both this question and (1) were posed independently also by L. Moser.) I offer 500 dollars for a proof or disproof of (1).

Conway and Guy proved $F(2^l) \geqslant k+2$ for $k > k_0$ and it has been conjectured that for $k > k_0 F(2^k) = k+2$. I have no opinion.

In 1932 or 1933, S. Sidon defined a sequence $1 \leqslant a_1 < a_2 < \cdots$ to be a B_r sequence if the sums $\sum \varepsilon_i a_i$, $\sum \varepsilon_i \leqslant r$ are all distinct ($\varepsilon_i = 0$ or 1). In other words the sums taken r (or fewer) at a time should be all different. He wanted to estimate the slowest possible growth of a B_r sequence. He was led to these problems from his well known work on lacunary trigonometric series.

Also he asked the following very fruitful question: Let $1 \leqslant a_1 < a_2 < \cdots$ be an infinite sequence of integers. Denote by $f_2(n)$ the number of solutions of $n = a_i + a_j$. Is there a sequence A for which $f_2(n) > 0$ for $n > n_0$ but $f_2(n) m^{-\varepsilon} \to 0$ for all $\varepsilon > 0$?

Sidon mentioned these problems to me when we first met in 1932 or 1933. By the greedy algorithm I easily constructed a B_2 sequence satisfying $a_k < ck^3$. We both conjectured that this is very far from the truth and probably there are B_2 sequences with $a_k < k^{2+\varepsilon}$ for every $\varepsilon > 0$ if $k > k_0(\varepsilon)$.

We are still very far from being able to settle this question. Using the probabilistic method Rényi and I proved the existence of a sequence with $a_k < k^{2+\varepsilon}$ and $f_2(n) < C_\varepsilon$. The problem whether a B_2 sequence exists with $a_k = o(k^3)$ is still open and I offer a hundred dollars for a proof or disproof. (*Added in proof.* Ajtai, Komlos and Szemerédi proved this conjecture; their proof will appear in Europ. Comb. J.)

Using the probability method I proved the existence of a sequence $A = \{a_1 < \cdots\}$ satisfying

$$c_1 \log n < f_2(n) < c_2 \log n \tag{3}$$

which answers affirmatively Sidon's question on whether $f_2(n) n^{-\varepsilon} \to 0$. Is (3) the best possible? Turán and I conjectured that if $f_2(n) > 0$ for all $n > n_0$, then $f_2(n)$ can not be bounded. I offer 500 dollars for a proof or disproof. Perhaps $f_2(n) > 0$ for all n already implies that $f_2(n) > c \log n$ for infinitely many n. Is there a sequence A for which

$$f_2(n)/\log n \to 1? \tag{4}$$

The probability method easily gives that if $h(n) \to \infty$ monotonically, then there is a sequence A satisfying $f_2(n)(h(n) \log n)^{-1} \to 1$. I expect that there is no sequence satisfying (4).

I proved that for a B_2 sequence

$$\limsup a_k/k^2 \log k > 0. \tag{5}$$

Perhaps in (5) the lim sup is in fact infinity. On the other hand perhaps there is a B_2 sequence satisfying (for all k)

$$a_k < c_1 k^2 (\log k)^{c_2}. \tag{6}$$

I offer a thousand dollars for clearing up the problems raised by (5) and (6).

These problems change character completely if we restrict ourselves to finite sequences. Denote by $F_r(n)$ the largest integer l for which there is a sequence $1 \leqslant a_1 < \cdots < a_l \leqslant n$ so that all the sums $\sum \varepsilon_i a_i$, $\sum \varepsilon_i \leqslant r$, $\varepsilon_i = 0$ or 1 are different.

Turán and I conjectured

$$F_2(n) = n^{\frac{1}{2}} + O(1). \tag{7}$$

We only could prove $F_2(n) < n^{\frac{1}{2}} + cn^{\frac{1}{4}}$, $F_2(n) < n^{\frac{1}{2}} + n^{\frac{1}{4}} + 1$ was proved by Lindstrom. I offered (and offer) 500 dollars for a proof or disproof of (7). Bose and Chowla showed that $F_r(n) \geqslant (1+o(1))n^{1/r}$ follows by using finite geometries. Bose called attention to the fact that the proof of

$$F_r(n) < (1+\varepsilon)n^{1/r} \tag{8}$$

presents great difficulties for $r > 2$. Our proof with Turán for $r = 2$ does not work and at the moment this attractive problem seems intractable. Perhaps $F_r(n) = n^{1/r} + O(1)$ holds for every r.

I was not able to prove that if $1 \leqslant a_1 < a_2 < \cdots$ is an infinite B_3 sequence, then

$$\limsup a_k/k^3 = \infty. \tag{9}$$

The same problem arises for every $r > 3$.

A further generalisation: Let $a_k < ck^r$ for every k. Is it true that

$$\limsup f_r(n) = \infty?$$

This is open even for $r = 2$ and perhaps here the real difficulty occurs already for $r = 2$.

Rényi and I proved by the probabilistic method that there is a sequence $a_k < ck^r$ for which

$$\sum_{n=1}^{x} f_r(n)^2 < Cx. \tag{10}$$

Probably (10) holds even for a basis of order r. In other words there is a sequence A satisfying (10) and $f_r(n) > 0$ for every n.

I proved that there is an infinite B_2 sequence for which

$$\limsup A(n)n^{\frac{1}{2}} \geqslant \tfrac{1}{2},$$

where

$$A(n) = \sum_{a_i < n} 1.$$

this has been improved to $1/\sqrt{2}$ by Krickeberg and I conjectured that it can be improved to 1, which of course would be best possible.

This would follow from one of my favourite recent conjectures: Let $1 \leqslant a_1 < \cdots < a_k$ be a finite B_2 sequence. Prove that it can be imbedded into a perfect difference set, i.e. there is a prime p and a set of $p+1$ residues $u_1, \ldots, u_{p+1} \bmod p^2+p+1$ so that all the differences $u_i - u_j$ are incongruent mod p^2+p+1 and the a's all occur amongst the u's. I offer a thousand dollars for a proof or disproof of this conjecture.

Recently Nathanson and I published several papers on bases and we will soon (I hope) publish a survey paper on this subject — here I only state one of our most attractive problems: Is there an infinite sequence $a_1 < a_2 < \cdots$ of integers so that

$f_2(n)>0$ for all $n>n_0$ but if we omit any a_i, then the number of $n<x$ with $f_2(n)=0$ is $>c_i x$ for all (or perhaps only infinitely many) x?

A recent conjecture of D. Newman and myself states as follows: There is a sequence $a_1<a_2<\cdots$ of integers for which $f_2(n)$ is bounded but which is not the union of a finite number of B_2 sequences. (*Added May 1980;* I proved this conjecture in 1979 and my proof will soon appear in the first issue of Europ. Combinatorial J.)

Let $g(n)>0$ be a non-decreasing function of n. I conjecture that the lower density of the integers n for which $f_2(n)=g(n)$ is 0. The upper density can be positive but I believe it is bounded away from 1.

It is easy to construct a sequence of integers $1\leqslant a_1<\cdots$ so that every integer n can be uniquely written in the form a_i-a_j. It is easy to see that

$$\limsup_{n\to\infty} a_i^{(n)}/n=\infty \quad \text{where } n=a_i^{(n)}-a_j^{(n)}. \tag{11}$$

I can not determine how fast (11) must increase.

In general the problems here can at present be attacked only by the probability method and where these methods do not apply we have not been able to make much progress.

Many of the problems stated here are discussed in the excellent book of Halberstam and Roth Sequences, (Oxford Univ. Press, Oxford, 1966). The book has very extensive references.

For many problems and results on bases and related questions see A. Stöhr, Gelöste and ungelöste Fragen über Basen der natürlichen Zahlenreihe I and II, J. Reine u. Angew. Math. 194 (1955) 40–64, 110–140.

B. Lindström, An inequality for B_2 sequences, J. Combinatorial Theory, 6 (1969) 211–212.

P. Erdös, Problems and results in additive number theory, Colloque sur la Théorie des Nombres, Bruxelles, George Thone, Liège, Manon and Cie, Paris 1959, 127–137.

In 1959, 1963 and 1972 there were meetings on number theory at the University of Colorado in Boulder, many interesting problems can be found in the conference reports which were published but unfortunately are not easily available. Also see the book H.H. Ostmann, Additive Zahlentheorie, Ergebnisse der Math, Heft 7 und 11 contains a review of the older litterature and also H. Rohrbach, Ein Beitrag zur additiven Zahlentheorie, Math. Zeitschrift 42 (1937) 1–30 and Einige neuere Untersuchungen über die Dichte in der additiven Zahlentheorie, Jahresbericht D.M.V. 48 (1938) 199–236.

4. Solutions of equations in dense sets of integers or real numbers

A sequence of real numbers is called primitive if no one divides any other. A sequence of real numbers $1<a_1<\cdots$ is called primitive if for every i, j and k

$$|ka_i-a_j|\geqslant 1. \tag{1}$$

If the a's are integers (1) means that no a divides any other. In 1934 Besicovitch made the surprising discovery that there is a primitive sequence of positive upper density. Behrend and I proved that the lower density of a primitive sequence must be 0. Behrend proved that if $1<a_1<\cdots<a_k\leqq x$ is a primitive sequence, then

$$\sum_{i=1}^{k}\frac{1}{a_i}<c\log x(\log\log x)^{-\frac{1}{2}}. \tag{2}$$

Szemerédi, Sárközi and I proved that for an infinite primitive sequence

$$\sum_{a_i<x}\frac{1}{a_i}=o(\log x(\log\log x)^{-\frac{1}{2}}). \tag{3}$$

(2) and (3) are both best possible.

Perhaps (2) and (3) hold if A is a sequence of real numbers satisfying (1). In fact I can not even prove that if (1) is satisfied, then $\limsup a_n/n=\infty$.

The only result in this direction is an unpublished one of J. Haight. He proved that if the $\{a_i\}$ are rationally independent and satisfy (1), then $a_n/n\to\infty$, or in other words $[a_n]$ is a sequence of density 0. Observe that by the example of Besicovitch this does not have to hold if the a's are integers.

A question of W. Schmidt states: Is there a set S of real numbers of infinite measure so that x/y, $x\in S$, $y\in S$ is never an integer?

J. Haight and E. Szemerédi (independently) constructed such a set. Denote by $m(S,x)$ the measure of the intersection of S with $(0,x)$. How fast can $m(S,x)$ tend to infinity? It is easy to see that $m(S,x)=o(x)$ but probably very much more is true. Immediately two questions can be posed: Let $F(x)\to\infty$; is there a set S so that x/y $x\in S$, $y\in S$ is never an integer and $m(S,x)>F(x)$ for all $x>x_0$, or only for a sequence $x_m\to\infty$? The answer to these two questions will no doubt be quite different.

A conjecture of Sárközi, Szemerédi and myself which has perhaps been neglected states: To every $\varepsilon>0$ there is a k so that if $k<a_1<a_2<\cdots$ is any primitive sequence then

$$\sum_{i=1}^{\infty}\frac{1}{a_i\log a_i}<1+\varepsilon. \tag{4}$$

The following question seems difficult and perhaps has no reasonable solution: Let $b_1<b_2<\cdots$ be an infinite sequence. What is the necessary and sufficient condition that there should exist a primitive sequence $\{a_n\}$ satisfying $a_n<Cb_n$ for every n? Perhaps there is more chance to get an answer for the following question: Let $n_1<n_2<\cdots$. What is the necessary and sufficient condition that there should exist a primitive sequence $\{a_k\}$ satisfying $A(2^{n_i})>c2^{n_i}$ for every i and an absolute constant c?

Now we discuss some multiplicative extremal problems. Let $1<a_1<\cdots<a_k\leqq x$ be a sequence of integers. Assume that the products a_ia_j are all distinct. I

proved

$$\pi(x)+c_1x^{\frac{3}{4}}(\log x)^{-\frac{3}{2}}<\max k<\pi(x)+c_2x^{\frac{3}{4}}(\log x)^{-\frac{3}{2}}. \tag{5}$$

where $\pi(x)$ is the number of primes $\leqslant x$.

(5) is perhaps unexpectedly accurate I am sure that there is an absolute constant c so that

$$\max k=\pi(x)+(c+o(1))x^{\frac{3}{4}}(\log x)^{-\frac{3}{2}}. \tag{6}$$

I do not at the moment see how to prove (6) and I do not believe that there is a simple explicit formula for max k.

Let $F(x)$ be the largest l for which there is a sequence of real numbers $1\leqslant a_1<\cdots<a_l\leqslant x$ for which

$$|a_ia_j-a_ra_s|\geqslant 1 \tag{7}$$

for all choices of the indices i, j, r, s. I was sure that $l=o(x)$ and in fact I expected that an estimation like (6) will hold here too. In fact it turned out that I completely misjudged the situation. Ralph Alexander constructed a sequence satisfying (7) and $l>cx$. Here is the outline of Alexander's construction. Put $x=4e^{N+1}$. By a theorem of Turán and myself there are integers

$$1<y_1<\cdots<y_t<e^{2N},\quad t>(1-\varepsilon)e^N \tag{8}$$

so that the sums y_i+y_j, $1\leqslant i<j\leqslant l$ are all distinct. By removing at most half of the y's we obtain a subsequence

$$1\leqslant y_1'<\cdots<y_l'<e^{2N},\qquad l\geqslant\tfrac{1}{2}t,\quad y_{i+1}-y_i'>\tfrac{1}{4}e^N \tag{8'}$$

Put

$$x_i=\exp(N+y_i'/e^{2N}). \tag{9}$$

From (9) we obtain by a simple computation that

$$|x_i-x_j|>\tfrac{1}{4}\quad\text{and}\quad|x_ix_j-x_rx_p|>1.$$

Put finally $a_i=4x_i$. (8), (8') and (9) show that (7) and $l>cx$ are satisfied. After this surprising result I am no longer so sure that (1) implies lim sup $a_n/n=\infty$ as I used to be. The following problem might be still of some interest: Let $1\leqslant a_1<\cdots<a_{t_n}\leqslant n$ satisfy (7). Determine of estimate max t_n. Perhaps lim $t_n/n=C$. Determine C, this perhaps is not a hopeless task. Finally is there an infinite sequence $1\leqslant a_1<\cdots$ satisfying (7) and $A(x)>cx(\log x)^{-\frac{1}{2}}$ for every x? If true this is clearly best possible.

Some more problems. Let $1\leqslant a_1<\cdots<a_{t_n}\leqslant n$ be again a sequence of integers. Assume that all the products $\prod_i a_i^{\alpha_i}$ ($\alpha_i\geqslant 0$ integer) are distinct. Then it is easy to see that max $t_n=\pi(n)$. Assume next that the products $\prod a_i^{\varepsilon_i}$, $\varepsilon_i=0$ or 1 are all distinct. I suspect that then

$$\max t_n=\pi(n)+\pi(n^{\frac{1}{2}})+o(n^{\frac{1}{2}}/\log n)$$

I could only prove that $\max t_n < \pi(n) + Cn^{\frac{1}{2}}(\log n)^{-1}$. Perhaps the reader will forgive a very old man to tell how Pósa and I conjectured in 1963 how large t_n can be. In 1962 or 1963 I talked to high school students at the Bolyai Math. Soc. I stated that perhaps

$$\max t_n = \pi(n) + \sum_{k=1}^{\infty} \pi(n^{1/2^k}),$$

but then immediately noticed that this is wrong. Not wanting to state more nonsense I talked immediately about something else. Before the end of my talk I formulated in my mind the following conjecture:

$$\max t_n = \pi(n) + \pi(n^{\frac{1}{2}}) + \pi(n^{\frac{1}{4}}) + \pi(n^{\frac{1}{7}}) + \cdots \tag{10}$$

where in the sum (10) $\pi(n^{1/k})$ occurs if and only if $F(k) > F(k-1)$ where $F(k)$ is defined in the first problem of III (it is the largest l so that there is a sequence $1 \leqslant a_1 < \cdots < a_l \leqslant k$, all the sums $\sum_{i=1}^{l} \varepsilon_i a_i$ are different)

$$\max t_n \geqslant \sum \pi(n^{1/a_k}) \tag{11}$$

is of course easy — it is not clear if the opposite inequality is also true. After the lecture I talked to Pósa (who was then about 14) and he formulated the same conjecture during my talk and noticed of course that $\sum \pi(n^{1/2^k})$ is wrong. We do not know at present if (10) is true.

The following nice conjecture is due to Beurling. Let $1 < P_1 < \cdots$ be a sequence of real numbers. $b_1 < b_2 < \cdots$ is the set of real numbers of the form $\prod P_i^{\alpha_i}$ where the α_i are non-negative integers. Assume that

$$B(x) = \sum_{b_i < x} 1 = x + o(\log x). \tag{12}$$

Then the P's are the primes. It is not hard to see that if (12) is true it is best possible.

In this connection H.N. Shapiro asked the following question: Assume that the numbers $\prod_i P_i^{\alpha_i}$ differ by at least one. Is it then true that

$$\sum_{P_i \leqslant x} 1 \leqslant \pi(x), \tag{13}$$

The equality occurring only if the P's are the primes?[1]

To end this section I state the following interesting measure theoretic conjecture of John Haight: Let E be a set of a positive measure in $(0, \infty)$. Consider the

[1] I just notice (1978.IX.17) that I stated nearly the same conjecture in my paper "Some applications of graph theory to number theory". The many Facets of Graph Theory, Lecture Notes in Mathematics 110 (Springer Verlag, Berlin) 77–82, see p. 82. I did not conjecture though that equality holds only if the P's are the primes. In view of this I should perhaps with regret modify what I said about my memory and mind. When Shapiro told me this problem I felt foolish that I did not think of it myself.

set $E' = \bigcup_{r=1}^{\infty} rE$. In other words $z \in E'$ if for some integer r, $z/r \in E$. Is it true that for almost all x there is an $M(x)$ so that for every integer $n > M(x)$, $nx \in E'$?

For primitive sequences see Halberstam–Roth, "Sequences" and P. Erdös, A. Sárközy and E. Szemerédi, On the divisibility properties of sequences of integers, Coll. Math. Soc. J. Bolyai, Vol. 2. Number theory (North Holland Amsterdam, 1968) 35–49, this paper has extensive references.

J.A. Haight, A linear set of infinite measure with no two points having integral ratio, Mathematika 17 (1970) 133–138.

E. Szemerédi, On a problem of W. Schmidt, Studia Sci. Math. Hungar. 6 (1971) 287–288.

About Beurling primes see e.g. H.G. Diamond, When do Beurling generalised integers have a density, J. Reine Angew. Math. 295 (1977) 22–29. This paper contains many references to older results.

P. Erdös, On some applications of graph theory to number theoretic problems, Publ. Ramanujan Inst. 1 (1969) 131–136.

5. Some problems on infinite subsets

Graham and Rothschild conjectured that if the integers are split into two classes there always is an infinite sequence $1 < a_1 < \cdots$ so that all the sums $\sum \varepsilon_i a_i$, $\varepsilon_i = 0$ or 1 are in the same class. This beautiful conjecture was proved by Hindman whose proof was greatly simplified by Baumgartner. Later a different and perhaps the simplest proof was given by Glaser.

Some time ago, I thought of the following fascinating problem: Divide the integers into two classes. Is it true that there always is an infinite sequence $a_1 < \cdots$ so that all the multilinear expressions formed from the a's are all in the same class. One would perhaps guess that the answer must be "no" but no counterexample is in sight.

The following much weaker conjecture is also open: Divide the integers into two classes. Is there an infinite sequence $a_1 < \cdots$ so that all the sums $a_i + a_j$ and the products $a_i a_j$ are in the same class? A further complication arises if we also insist that $a_1, a_2, \ldots$ should also belong to the same class. In fact very little is known. Graham proved that if we divide the integers $1 \leqslant t \leqslant 252$ there are always 4 distinct numbers x, y, $x+y$, xy all in the same class. The number 252 is the best possible. Hindman showed that if we require $x > 1$, $y > 1$ then the same result holds for $t \leqslant 990$ and 990 is the best possible. As far as I know this is all that is known. (Recently Hindman found some very interesting counterexamples.)

Ewing conjectured that there always is an infinite sequence $a_1 < \cdots$ where all the sums $a_i + a_j$ ($i = j$ permitted) are in the same class. It is rather annoying that this simple and interesting question is still open. Hindman has certain preliminary results. Among others he proved that this conjecture fails certainly for three classes. In fact Hindman observes that one of his sequences can have density 0

and in fact

$$A_1(x) = \sum_{a_i < x} 1 < Cx^{\frac{1}{2}} \tag{1}$$

It is not yet clear if (1) is the best possible.

I thought of strengthening Hindman's theorem in the same way as Szemerédi strengthened Van der Waerden's theorem: Let A be a sequence of positive density. Is there an infinite subsequence $a_1 < \cdots$ and an integer t so that all the integers a_i, $a_i + a_j + t$ belong to the same class?

I first hoped that Hindman's theorem on all subsums $\sum \varepsilon_i a_i$ can be generalised for sets of positive density, but the following example of E. Straus seems to give a counterexample to all such attempts: Let $p_1 < \cdots$ be a set of primes tending to infinity fast, and let $\sum \varepsilon_i < \infty$. Consider the set of integers $a_1 < a_2 < \cdots$ so that $a_u \not\equiv \alpha \pmod{p_i}$ where $|\alpha| \leqslant \varepsilon_i p_i$. The density of this sequence is $\prod_i (1 - \varepsilon_i) > 0$, and it is not difficult to see that this sequence furnishes the required counterexample.

More than 10 years ago, Graham and I conjectured that if we split the integers into k classes, then

$$1 = \sum \frac{1}{x_i}, \quad x_1 < \cdots \quad \text{(finite sum)} \tag{2}$$

is always solvable with all the x_i in the same class. In the language of hypergraphs the conjecture states: The chromatic number of the non-uniform hypergraph whose vertices are the integers and whose edges are the sets satisfying (2) is infinite. A finite form of this problem states that if $\sum_{a_i < n} 1/a_i$ is large enough, then one can select amongst the a's a sequence satisfying (2). "Large enough" as a function of n? We really have no idea what "large enough" should mean. It could be $o((\log \log n)^{\alpha})$ or $\varepsilon \log n$.

Silverman and I conjectured that if we split the integers into k classes, then there are always two integers in the same class whose sum is an rth power (in particular a square). It would be of interest to characterise the sequences for which this conjecture holds. We also conjectured that if $1 \leqslant a_1 < \cdots < a_k \leqslant n$ is such that $a_i + a_j$ is never a square, then $k \leqslant (1 + o(1))n/3$, $n/3$ is given by the integers $\equiv 1 \pmod 3$. The exact determination of $\max k$ is perhaps not hopeless, but we certainly have not succeeded.

If we assume that $a_k - a_i$ is never a square, then the situation is radically different. Fürstenberg and Sárközi independently of each other proved that here $k = o(m)$. Sárközi obtains explicit bounds for k but no doubt they are very far from being best possible. Many further conjectures could be formulated but we must leave these to the reader.

Glazer's proof is given in the excellent survey paper of W.W. Comfort, Ultrafilters: Some old and some new results, Bull. Amer. Math. Soc. 83 (1977) 417–455, see 449–452.

N. Hindman, Finite sums from sequences within cells of a partition of N, J. Combinatorial theory 17 (1974) 1–11; Baumgartner, A short proof of Hindman's theorem, 17 (1974) 384–386.

A. Sárközy, On difference sets of integers I, II and III. II will appear in Annales Univ. Sci. Budapest, I and III, Acta Math. Acad. Sci. Hungar. 31 (1978) 125–149 and 355–386.

N. Hindman, Ultrafilters and combinatorial number theory, Number Theory Carbondale, 1979, Lecture Notes in Math. 751 (Springer-Verlag, Berlin) 119–184. This paper has a very extensive list of references.

6. Some problems on sieve methods

During my long life I often applied Brun's method but unfortunately I contributed next to nothing to the improvement of the method or its generalisations. All I can do is to state a few problems which are directly or indirectly connected with sieve methods.

1. Let $f(x)$ be the smallest integer so that there is a set of residues

$$a_p \pmod p, \quad p < f(x) \tag{1}$$

so that every integer $n < x$ satisfies one of the congruence (1). In particular must $f(x)$ be significantly larger than $x^{\frac{1}{2}}$? Also it would be very useful to estimate $F(x)$ if we only require that the number of integers $n \leqslant x$ not satisfying any of the congruences (1) is $o(x/\log x)$ (where $p < F(x)$). Is $F(x)$ significantly smaller than $f(x)$? These problems can of course be extended if more than one residue is omitted.

It is not clear who first formulated this problem — probably many of us did it independently. I offer max(1000 dollars, $\frac{1}{2}$ my total savings) for clearing up of this problem. It is clear that many important problems could be attacked if we would know a little more.

Let ε_x be the largest number for which there is a system of congruences

$$a_p \pmod p, \quad x^{\varepsilon_x} < p < x \tag{1'}$$

so that every integer $n < x$ satisfies at least one of the consequences (1'). It is not difficult to prove that

$$\varepsilon_x \geqslant \frac{c \log\log\log x}{\log\log x},$$

but perhaps ε_x is much larger.

Ruzsa and I made the following surprising conjecture. There is a constant C for which there is a set of primes $p_i < x$, $\sum 1/p_i < C$ and a system of congruences

$$a_i \pmod{p_i} \tag{1''}$$

so that every integer $n < x$ satisfies at least one of the congruences (1''). If we are right, then very likely $\varepsilon_x > c$ for some absolute constant c. (See our forthcoming paper in the Journal of Number Theory.)

2. Is it true that to every ε and η there is a k so that the density of integers n for which

$$\max_{1\leq i\leq k} P(n+i) > n^{1-\varepsilon}$$

is greater than $1-\eta$, where $P(m)$ denotes the greatest prime factor of m. I can only do this for $\varepsilon = \frac{1}{2}$.

3. Is it true that to every $\varepsilon > 0$, $\eta > 0$ there is a $k = k_0(\varepsilon, \eta)$ so that $p_1 < \cdots < p_k < n^{1-\varepsilon}$ and $a_i^{(j)}$, $1 \leq i < (p_j - 1)/2$, $j = 1, \ldots, k$ are any set of $(p_j - 1)/2$ distinct residues mod p_j, $j = 1, \ldots, k$. Then the number of integers $m \leq n$ for which

$$m \not\equiv a_i^{(j)} \pmod{p_j}, \quad 1 \leq i \leq (p_j - 1)/2, \quad j = 1, \ldots, k$$

is less than εn? This follows easily from the large sieve if $p_k < n^{\frac{1}{2}}$ but the general case seems intractable at present.

4. Denote by $V(n)$ the number of distinct prime factors of n I conjecture that for infinitely many

$$\max_{m<n} m + V(m) \leq n \tag{2}$$

(2) seems unattackable. Replace $[m + V(m)]$ by $[m + \varepsilon V(m)]$ where $\varepsilon > 0$ is sufficiently small. It seems to me that even this weakened form of (2) is just beyond what we can do at the present time.

Selfridge and I considered the following much more difficult problem

$$\max_{m<n} m + d(m) \leq n + 2 \tag{3}$$

where $d(m)$ is the number of divisors of m.

$n = 24$ satisfies (3). We convinced ourselves that if (3) has a solution $n > 24$, then n must be enormously big–far beyond the limits of our tables and computations.

A related problem can be formulated as follows. Put

$$f(n, t) = \max_{1\leq i\leq t} V(n+i)$$

Trivially $f(n, t) \geq (1 + o(1))\log t (\log\log t)^{-1}$, but I expect that usually very much more will be true. More precisely put

$$\max_t f(n, t) \log\log t \, (\log t)^{-1} = F(n). \tag{4}$$

Then $F(n) \to \infty$ as $n \to \infty$. I am very far from being able to prove this.

4. Are there infinitely many integers n for which $n - i \not\equiv 0 \pmod{p^2}$ for $i = 0, 1, \ldots, p-1$ and for all $p^2 < n$? A further question is: are there infinitely many integers n for which $n - i \not\equiv 0 \pmod{k^2}$ for every $i = 0, 1, \ldots, k-1$ and $k^2 < n$?

The density of integers in both of these classes is 0, but I have no good estimation for these numbers. Some preliminary calculations of Selfridge seem to

indicate that the number of these integers not exceeding x is $(c+o(1))x^{\alpha}$ for some $1>\alpha>0$ and c, but we are very far from being able to prove this.

5. Elliott considered the following problem: Let $0\leqslant a_1<\cdots<a_k$ be a sequence of integers which does not contain a complete set of residues mod p (p runs through the set of all primes). Clearly only the primes $p\leqslant k$ have to be considered. Elliott investigates the estimation of $\min a_k = A(k)$. He proves

$$(1+o(1))k\log k\leqslant A(k)\leqslant(2+o(1)k\log k. \tag{5}$$

The lower bound in (5) is due to Davenport. Probably the lower bound is the correct one, this seems very hard and is of course connected with the problem stated in 1. The exact determination of a_k is probably hopeless.

It would be of interest to determine A_k for small values of k. Also determine or estimate

$$B_k=\min(a_1+\cdots+a_k)/k. \tag{6}$$

There is no reason to assume that the minimum in (5) and (6) is given by the same sequence.

One can define a sequence $\{a_1,\ldots,a_k\}$ by the greedy algorithm as follows: Assume $a_1,\ldots,a_{i-1}$ is already defined let a_i be the smallest integer greater than a_{i-1} so that $a_1,\ldots,a_{i-1},a_i$ does not form a complete set of residues (mod p). Estimate this a_k as well as possible. There is no reason to assume that this sequence gives the minimum in (5) or (6).

6. Let $f(x)$ be the largest integer for which there is a system of congruences

$$a_p \pmod p \tag{7}$$

(where p runs through the primes not exceeding x) so that every integer $n<x$ should satisfy at least $f(x)$ of the congruences (1). I can not even prove that $f(x)\geqslant 2$ for $x>x_0$. On the other hand perhaps $f(x)$ tends to infinity together with x.

Denote by $F(x)$ the largest integer for which there is a system of congruences

$$a_n \pmod n \tag{8}$$

(where n runs through the integers not exceeding x) so that every integer $t\leqslant x$ satisfies at least $F(x)$ of the congruences (8). Now it is a simple exercise to prove that $F(x)\to\infty$ as $x\to\infty$. The only problem is to determine or estimate how fast $F(x)$ tends to infinity.

Denote by $\pi(x)$ the number of primes not exceeding x. $\pi(x)=(1+o(1))x/\log x$ is the prime number theorem. The number of primes in short intervals is very difficult to estimate. I conjectured

$$\pi(x)-\pi(y)<c\frac{x-y}{\log x}\quad\text{for every } y<x-(\log x)^C \tag{9}$$

where C is a sufficiently large absolute constant.

The proof or disproof of (9) is probably hopeless at present.

P.D.T.A. Elliott, On sequences of integers, Quarterly J. Math. 16 (1965) 35–45.

About sieve methods in general see Halberstam–Roth, "Sequences" and the more recent Cambridge Tract of Hooley and the book of Halberstam–Richert on sieve methods.

7. Miscellaneous problems

This section is necessarily incomplete. I have to concentrate mainly on problems stated in previous papers. I discuss problems only if they were in my opinion neglected or if some progress has been made.

Riddell defines $g_k(n)$ as the largest integer so that among any n real numbers one can always find $g_k(n)$ of them which do not contain an arithmetic progression of k terms. Clearly $g_k(n) \leqslant r_k(n)$ and Riddell showed that $g_k(n)$ can be less than $r_k(n)$, $g_3(5)=3$, $r_3(5)=4$ (from the set 1, 3, 4, 5, 7) one can select only 3 integers not containing an arithmetic progression). Riddell also shows $g_3(14)=7<r_3(14)=8$. Perhaps for large n, $g_k(n)=r_k(n)$.

Riddell also investigated the following problem: Let $1 \leqslant a_1 < \cdots < a_n$ be any set of n integers (or real numbers). Denote by $F_2^*(n)$ the minimum of the largest r for which $a_{i_1} < \cdots < a_{i_r}$ is a B_2 sequence. The minimum is to be taken for all the sequences $\{a_1, \ldots, a_n\}$. I conjectured $F_2^*(n)=(1+o(1))n^{\frac{1}{2}}$ and perhaps $F_2^*(n)=F(n)$ for all large n. Komlós, Sulyok and Szemerédi proved a general theorem on the solution linear equations which implies $g_k(n) > cr_k(n)$ and $F_2^*(n) > cF_2(n)$. They proved that apart from a multiplicative constant the worst set for all such problems is to take the integers $1, 2, \ldots, n$. This remarkable theorem still leaves many questions unanswered. Very likely c can be chosen to be $1-\varepsilon$ for every $\varepsilon > 0$ if $n > n_0(\varepsilon)$ and perhaps $g_k(n)=r_k(n)$, $F_2^*(n)=F_2(n)$ for $n > n_0$. If this conjecture seems too optimistic perhaps one should only expect $g_k(n) > r_k(n) - C$, $F_2^*(n) > F_2(n) - C$. (Recall the functions $r_k(n)$ and $F_k(n)$ defined in Sections 1 and 3 respectively.)

Also, what happens for non linear equations? Here are some of the questions I have in mind: Let $A_n=\{a_1, \ldots, a_n\}$ be a sequence of n integers. Denote by $g_k(A_n)$ the largest integer l for which there is a subsequence $\{a_{i_1}, \ldots, a_{i_l}\}$ so that $\{a_{i_l}^k\}$ is a B_2 sequence. Put $\min g_k(A_n)=G(n, k)$ where the minimum is to be taken over all the sequences A_n. Is it true that $G(n, k)$ is attained if A_n is $1, \ldots, n$? Is it true that $G(n, 2) > n^{1-\varepsilon}$ for every $\varepsilon > 0$ if $n > n_0(\varepsilon)$? $G(n, k) > c_k n$ for $k \geqslant 3$? Perhaps these conjectures are completely wrongheaded.

We can ask questions for infinite sequences which are perhaps both interesting and fruitful: Is there an infinite sequence $a_1 < \cdots, a_k < k^{1+\varepsilon}$, so that $\{a_i^2\}$ is a B_2 sequence? Further is there a sequence $a_t < Ct$ so that the sums $a_i^3 + a_j^3$ are all distinct?

Riddell denotes by $P(n, k)$ the largest integer so that amongst n points in k-dimensional space one can always find $P(n, k)$ points which do not contain an isosceles triangle. It is not hard to prove that $P(n, k) > n^{\varepsilon_k}$, $\varepsilon_k \to 0$ as $k \to \infty$. Perhaps $P(n, 2) < n^{1-c}$. Here again perhaps the worst set of points is if they are lattice points in a sphere of as small radius as possible. By the way, $P(n, 1)$ reduces to having no three points in an arithmetic progression. Instead of no three points forming an isosceles triangle we could require that all the distances be distinct. We can leave the exact formulation of the problems to the reader.

One of my oldest unsolved problems states as follows: Let $f(n)$ be $+1$ or -1 for every n. Is it true that to every c there is a d and an m so that

$$\left| \sum_{k=1}^{m} f(kd) \right| > c? \tag{1}$$

I offer 500 dollars for a proof or disproof of (1). Perhaps (1) can be strengthened as follows: For suitable d and m, is it true that

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

It is not difficult to prove that (1'), if true, is the best possible.

Non-averaging sets. A set of integers $a_1 < a_2 < \cdots < a_k \leqslant n$ is called not averaging by E. Straus if no a_i is the arithmetic mean of other a's. Straus asked for an estimation of $\max k = A(n)$. The best bounds are at present

$$c_1 n^{\frac{1}{10}} < A(n) < n^{\frac{2}{3}} + \varepsilon. \tag{2}$$

The lower bound is due to Abbott and the upper to Straus and myself. It would be of interest to prove that $\lim \log A(n)(\log n)^{-1} = \alpha$ exists and to determine α.

Graham conjectured: Let $1 \leqslant a_1 < \cdots < a_n$ be n integers. Then

$$\max_{i,j} a_i/(a_i, a_j) \geqslant n.$$

Szemerédi proved this recently. The proof is not yet published.

I conjectured that if $1 \leqslant a_1 < \cdots < a_k \leqslant n$, $\sum 1/a_i < C$, then the number of integers $\leqslant n$ not divisible by any of the a's is $> n^{1-\varepsilon}$ for every $\varepsilon > 0$ and C if $n > n_0(\varepsilon, C)$. It turned out here that my intuition completely misled me since Ruzsa proved that to every $\varepsilon > 0$ there is a $C = C(\varepsilon)$ so that there is a sequence A with $\sum_{i=1}^{k} 1/a_i < C$ so that the number of $m \leqslant n$, with $m \not\equiv 0 \pmod{a_i}$ for $i = 1, \ldots, k$, is $< n^{\varepsilon}$, the exact dependence of ε from C is far from being known. Ruzsa proved that to every $\varepsilon > 0$ there is a $C = C(\varepsilon)$ so that there is a sequence A more than $x^{1-\varepsilon}$ numbers $m < x$, $m \not\equiv 0 \pmod{a_i}$, $i = 1, \ldots, k$. This is not yet cleared up.

the following extremal problem seems to be interesting. Let

$$\sum_i 1/a_i \leqslant 1$$

be any sequence of integers. Denote by $A'(x)$ the number of integers not exceeding x and not divisible by any of the a's. Determine or estimate min $A'(x)$ where the minimum is to be taken over all sequences satisfying (3). I expect that is of the order of magnitude $x/(\log x)^{\alpha}$. It follows from a result of Schinzel and Szekeres that $A'(x) < cx/(\log x)^{\alpha}$ for some $\alpha > 0$.

Here also my intuition was wrong. In 1940 I conjectured that if $1 \leqq a_1 < \cdots < a_k \leqq x$ is a sequence of integers so that the least common multiple of any two a's is greater than x, then $A'(x) > cx$. Szekeres soon proved me wrong and in fact here we have

$$c_2 x/(\log x)^{\beta_1} < A'(x) < c_1 x/(\log x)^{\beta_2}$$

It would be of interest if one could give explicitely the sequence which solves the extremal problem (3), but I very much doubt if this is possible. Ruzsa remarks that trivially $A'(x) > cx/\log x$.

The prime k-tuple conjecture of Hardy and Littlewood states that if $a_1 < \cdots < a_k$ such that the a's do not form a complete set of residues mod p for every p, then there are infinitely many integers n so that all the numbers $n + a_i$, $i = 1, \ldots, k$ are primes. This is of course quite unattackable at present.

On the other hand it is a simple exercise to prove that if $a_1 < \cdots < a_k$ does not form a complete set of residues mod p^2 for every p, then there are infinitely many integers n for which $n + a_i$, $i = 1, 2, \ldots, k$ are all squarefree.

There are difficulties in formulating a reasonable conjecture for infinite sequences. Can the following conjecture be true: Let a_k tend to infinity sufficiently fast, and assume that there is an n so that all the $n + a_k$ are primes. Are there infinitely many such values of n? Is there any hope of proving this for squarefree numbers instead of primes? Or are there values of n for which say $n + 2^{2^k}$ is always a prime? Always squarefree, or infinitely often a prime, or infinitely often squarefree? Unless I overlook a trivial way of getting a counterexample these questions are quite hopeless.

Perhaps the following generalisation for infinite sequences is possible: Let $1 \leqq a_1 < \cdots$ be a sequence which does not contain a complete set of residues mod p for every p, then there are infinitely many values of n for which all the integers $n + a_k$, $a_k < n$ are primes. Perhaps the following modification is slightly less hopeless: Let $1 \leqq a_1 < \cdots$ be a sequence which does not contain a complete set of residues mod p^2 for every p. Then there are infinitely many values of n for which all the integers $n + a_k$, $a_k < n$ are squarefree.

These conjectures just occurred to me, and perhaps some further condition on the thinness of the sequence $A = \{a_k\}$ may be needed. If true they are of course hopeless for the primes and it seems to be hopeless for the square free numbers too. On the other hand if the squares of primes are replaced by a sequence which increases fast enough the conjecture becomes easy. Let $n_1 < n_2 < \cdots$, $(n_i, n_j) = 1$ tend to infinity sufficiently fast and let $a_1 < a_2 < \cdots$ be a sequence of integers which does not contain a complete set of residues (mod n_i) for every i. Then there

are infinitely many integers x so that

$$x + a_i \not\equiv 0 \pmod{n_j}$$

for every n_j and $a_i < x$.

Let me call attention to the following beautiful conjecture of Ostman which is now about 25 years old: Prove that there do not exist two infinite sequences A and B so that the sequence $A + B$ differs from the set of primes in only a finite number of elements. Hornfeck proved that the sets A and B must both be infinite.

Here is a problem which I did not state quite correctly in III: Let $1 \leqq a_1 < \cdots < a_k \leqq n \cdots$ be a sequence of integers satisfying $(a_i, a_j) = 1$. Is it true that there exists an absolute constant C for which (p runs through the primes)

$$\sum \frac{1}{n - u_k} < C + \sum_{p<n} \frac{1}{p}.$$

I state some old problems which are perhaps not hopeless but which have been neglected: Denote by $f(k)$ the minimum number of terms of the square of a polynomial of k terms. Sharpening a result of Rédei and Rényi, I proved that $f(k) < k^{1-c}$. Rényi and I conjectured that $f(k) \to \infty$ as k tends to infinity. The value of $f(k)$ may depend what kind of coefficients we permit (integers, rationals, real or complex numbers). We believe that $f(k) \to \infty$ holds in all cases.

Let $a_1 < \cdots < a_k \leqq n$ be any sequence of integers, $b_1 < b_2 < \cdots$ is the sequence of integers no one of which is the multiple of any of the a's. Put $B(x) = \sum_{b_i < x} 1$. Is it true that for every $m \geqq n$

$$\frac{B(m)}{m} < 2\frac{B(n)}{n}? \tag{1}$$

It is easy to see that if (1) is true it is best possible. The a's consist only of a_1, $n = 2a_1 - 1$, $m = 2a_1$.

Roth conjectured that there is an absolute constant c so that to every k there exists an $n_0 = n_0(k)$ which has the following property: Let $n > n_0$, split the integers not exceeding n into k classes $\{a_i^{(j)}\}$, $j = 1, \ldots, k$. Then the number of distinct integers $m \leqq n$ which for some j, $j = 1, 2, \ldots, k$ can be written in the form $a_{i_1}^{(j)} + a_{i_2}^{(j)}$ is greater than cn.

Let $a_1 < \cdots < a_n$ be a set of integers. Consider the sums $a_i + a_j$ and products $a_i a_j$. I conjectured that there are more than $n^{2-\varepsilon}$ distinct numbers amongst them. This if true seems very difficult. It is possible that Szemerédi and I will be able to prove that the number of distinct terms in this sequence is $> n^{1+c}$ for some $c > 0$.

Graham conjectured: Let $1 \leqq a_1 < \cdots < a_p$ be p not necessarily distinct residues mod p. Assume that $\sum_{i=1}^{p} \varepsilon_i a_i \equiv 0 \pmod{p}$, $\varepsilon_i = 0$ or 1, and not all $\varepsilon_i = 0$ implies $\sum_{i=1}^{p} \varepsilon_i = r$. Does it then follow that there are at most two distinct residues mod p?

Szemerédi and I proved this if $p > C$ i.e. for sufficiently large p.

The following old problem of mine still seems to be open: Let $1 \leqq a_1 < \cdots < a_t \leqq n$ be such that there are at most r a's which are pairwise relatively prime.

Then we obtain max t by taking the multiples of the first r primes not exceeding n. Graham and I have the following related problem: Let $1 \leqq a_1 < a_2 < \cdots < a_k = n$, $(a_i, a_j) \neq 1$. What is the maximum value of k? A reasonable guess seems to be that either max $k = n/p(n)$, where $p(n)$ is the least prime factor of n, or it is the number of integers of the form $2t$, $t \leqq \frac{1}{2}n$, $(t, n) \neq 1$. This problem is stated with many confusing misprints in I (p. 123).

Straus and I conjectured that for $k > k_0$ there always is an $i = i(k)$ so that $p_k^2 < p_{k+i} p_{k-i}$. Pomerance recently proved that this conjecture is completely wrongheaded and that if $a_n^{1/n} \to 1$, $a_n/n \to \infty$, then there are infinitely many values of k so that for every $i < k$, we have

$$a_k^2 > a_{k+1} a_{k-i}. \tag{2}$$

This was in fact conjectured by Selfridge who always disbelieved our conjecture.

Pomerance and I tried unsuccessfully to prove that if the a's are the primes then the density of the integers k for which (2) is satisfied for every $i < k$ is 0.

Sárközi and I proved that if $a_1 < a_2 < \cdots$ is an infinite sequence of integers where no a divides sum of two greater a's, then this sequence has density 0. We could not prove that $\sum 1/a_i < \infty$. Also the following finite problem seems interesting: Let $1 \leqq a_1 < \cdots < a_k \leqq x$ be such that no a_i divides the sum of two greater a's. Then $k \leqq [\frac{1}{3}x] + 1$. Equality holds if $X = 3n$ and the a's are: $2n, 2n+1, \ldots, 3n$. The conjecture $k \leqq [\frac{1}{3}x] + 1$ is still open.

To finish let me state four more problems: Is it true that almost all numbers n have two divisors d_1 and d_2 satisfying $d_1 < d_2 < 2d_1$. This is one of my oldest conjectures and I offer 300 dollars for a proof or disproof. I only proved that the density of these integers exists. I claimed once that I have a proof that the density is 1. I unfortunately have to withdraw this claim.

The second question is only a few days old. Divide the integers into two classes $n_1 < \cdots$ and $m_1 < m_2 < \cdots$. Denote by $N_1 < N_2 < \cdots$ and respectively $M_1 < M_2$ the integers which are distinct sums of the n's respectively the m's. It is easy to see that either the (N_i) or the (M_j) must have upper density 1 (both can of course have lower density 0). It is not clear to me at present how large

$$\limsup \frac{1}{\log x} \max\left(\sum_{N_i < x} \frac{1}{N_i}, \sum_{M_i < x} \frac{1}{M_i} \right)$$

must be. It is easy to see that it can be less than 1, but I expect it to be greater than $\frac{1}{2}$

$$\left(\limsup \left(\frac{1}{\log x} \sum_{a_i < x} 1/a_i \right) \right.$$

is called the upper logarithmic density of the sequence A). Here I mention a theorem and problem of Davenport and myself on the existence of logarithmic density. Let $a_1 < a_2 < \cdots$ be any sequence of integers, $b_1 < b_2 < \cdots$ is the sequence of integers not divisible by any of the a's. Then the b's have logarithmic

density. Now to our problem: To each a_i make correspond an arbitrary set of residues $\{x_j^{(i)}\}$, $j=1, 2, \ldots, t_i$. Let $b_1<b_2<\cdots$ be the sequence of integers for which

$$b_u \not\equiv x_j^{(i)} \pmod{a_i}, \quad j=1, 2, \ldots, t_i$$

always holds if $b_u > a_i$. Is it true that the b's have a logarithmic density? Perhaps this question is not very difficult as far as I know it has not been attacked really seriously.

A sequence of integers $A=\{a_1<a_2<\cdots\}$ is said to have property P if

$$\prod_i a_i = \prod_j a_j$$

(where each product runs over a subset of the a's) is possible only if the number of factors on both sides of the equation is the same. Clearly the integers $\equiv 2 \pmod 4$ have property P. I asked: is there a sequence with property P having density $>1-\varepsilon$? Is there a finite sequence $1<a_1<\cdots<a_k\leqslant n$ having property P and satisfying $k>(1-\varepsilon)n$? Ruzsa showed that the answer to both questions is negative. Ruzsa in fact shows that if A is an infinite sequence with property P then its upper density is less than 1/e. This is best possible. Also if $1\leqslant a_1<\cdots< a_k\leqslant x$ has property P, then $A(x)<(1-c)x$ for an absolute constant $c>0$. The best value of c is not known.

Finally let me tell of a problem where I somewhat made a fool of myself. Herzog and Stewart studied visible lattice points i.e. lattice points $\{u, v\}$, satisfying $(u, v)=1$. One joins two visible lattice points if they are neighbours i.e. if they differ in only one coordinate and there by ± 1. Herzog and Stewart prove that there is only one infinite component and they conjecture that (a, p) $a\not\equiv 0 \pmod p$, p prime, always belongs to the infinite component. Last year when I gave a talk at Michigan State University I asked: Is there an infinite path through visible points no coordinate of which is 1. I though that the answer will be not too hard, and affirmative, and follishly offered 25 dollars for a proof. In the evening Stewart gave the simple proof: $\{p_k, p_{k+1}\}$ can be joined to $\{p_{k+1}, p_{k+2}\}$ by Tchebicheff's theorem which gives the path in question. Perhaps I should have asked: Is there a path going to infinity which avoids points both coordinates of which are primes and also points one coordinate of which is 1. We could further demand that the path is monotone i.e. every step increases the distance from the origin. Is there a monotone path where we change direction after a bounded number of steps?

The following result surely holds, but there are some technical difficulties in giving a rigorous proof: To every ε there is a k so that for all but εn^2 lattice points $\{a, b\}$, $0\leqslant a,\ b\leqslant n$ there is a sequence of visible lattice points $\{u_k, v_k\}$, $k=1, 2, \ldots, (u_0=v_0=1)$, where $u_0<u_1<\cdots$, $v_0<v_1<\cdots$ and $0\leqslant u_{k+1}-u_k< C$, $0\leqslant v_{k+1}-v_k<C$. In other words infinity can be reached with jumps of bounded length.

The following beautiful conjecture is due to Gordon and Motzkin: Is there a

sequence of distinct Gaussian primes $P_1, P_2, \ldots$, for which $|P_{k+1} - P_k| < C$ for some absolute constant C. The answer is almost certainly negative.

J. Komlós, M. Sulyok and E. Szemerédi, Linear problems in combinatorial number theory, Acta Math. Acad. Sci. Hungar. 26 (1975) 113–121.

H.L. Abbott, On a conjecture of Erdös and Strauss on non averaging sets of integers, Proc. 5th British Combinatorial Conference, Aberdeen (1975) 1–4.

P. Erdös and E. Szemerédi, On a problem of Graham, Publicationes Math. 23 (1976) 123–127.

P. Erdös and A. Sárközy, On the divisibility properties of sequences of integers, Proc. London Math. Soc. 21 (1970) 97–101.

P. Erdös, On the number of terms of the square of a polynomial, Nieun Arch. Wiskunde (1949) 63–65.

W. Verdenius, On the number of terms of the square and cube of polynomials Indig. Math. II (1949) 459–465.

Finally I would like to call attention to a forthcoming monograph of R.L. Graham and myself entitled "Old and new Problems and Results in Combinatorial Number Theory" which will soon appear in L'Enseignement Math. See further the forthcoming book of R.L. Graham, B. Rothschild and J. Spencer, Ramsey Theory (Wiley, New York, 1980).