ON SOME OF MY PROBLEMS IN NUMBER THEORY I WOULD MOST LIKE TO SEE SOLVED

P. Erdös*

Three years ago I wrote a paper entitled "On the problems in combinatorics I would most like to see solved." In view of the vast quantity of unsolved problems in number theory many of which are in subjects in which I am not particularly competent it would be presumptuous not to change the title a bit. I will mostly discuss problems due to my collaborators and myself. Only in the end will I add a problem or two which I recently heard. I hope the reader will forgive a very old man if he adds some historical reminescences. I must do it while my memory and mind are still more or less intact since I can not hope to escape the two greatest evils, old age and stupidity, very much longer except by the "trivial method", which perhaps mistakenly I do not feel I have to use as yet. Recently at least two books appeared on unsolved problems in number theory. I will refer to these as I and II

I. R. K. Guy, Unsolved problems in intuitive mathematics, Vol. 1 (Number Theory), Springer Problem books, (1981), Berlin, New York.

II. P. Erdos and R. L. Graham, Old and new problems and results in combinatorial number theory, Monographie No. 28 L 'Enseignement Math., 1980.

This paper contains more than 700 references. Many references which I do not give here can be found in this booklet.

My last three papers on problems in number theory were :

Paul Erdos, Some new problems and results in number theory, Number Theory, Proceedings, Mysore, 1981, Lecture Notes in Mathematics 938, 50-71. A forthcoming paper of mine will soon appear in the Proceedings of the meeting in Edmonton, April 1983 and at Marseilles in May 1983.

See also P. Erdos, Many old and on some new problems of mine in number theory, Southeastern Conference, Boca-Raton, 1982.

1. I start with my favourite problem :

Is it true that to every c there is a finite set of congruences

$$a_i \pmod{n_i}, \quad c = n_1 < n_2 < \dots < n_k \tag{1}$$

*Srinivasa Ramanujan Visiting Professor at MATSCIENCE Madras, during January 1984.

so that every integer satisfies as least one of the congruences (1). I offer \$1,000 for a proof or disproof. The current record is $n_1 = 24$. Let me tell you how I conjenctured (1). In 1934 Romanoff proved that the lower density of the integers which can be written in the form $2^k + p$ is positive. Turan and I found a simple proof for his minimal lemma and this started our correspondence. A few months later he asked me whether I can prove that there are infinitely many odd integers not of the form (1). I realised very soon that if I can find a system of congruences in which no n_i equals 6 then I easily get an arithmetic progression no term of which is of the form $2^k + p$. This was very easy to find and by the way Van den Corput also answered Romanoff's question. I also realised that (1) is an interesting problem for its own sake, though I did not realise how difficult it will be. I also realised that a positive answer to (1) will show that for every r there is an arithmetic progression no term of which is of the form $2^k + \theta_r$, where $\omega(\theta_r) \leq r$ ($\omega(n)$ denotes the number of distinct prime factors of n). By the way is it true that to every $\varepsilon > 0$ there is an r so that the lower density of the integers of the form $2^k + \theta_r$ is greater than $1 - \varepsilon$? There are many further problems and even a few results in this area for which I have to refer to I and II, but I want to mention two more problems. Linnik proved that there is an s so that every integer can be written in the form

$$p + q + 2^{k_1} + \dots + 2^{k_{s_\varepsilon}} .$$

Recently Gallagher proved the following stronger result : To every $\varepsilon > 0$ there is an s_ε so that the lower density of the integers of the form

$$p + 2^{k_1} + \dots \quad + 2^{k}$$

is greater than $1 - \varepsilon$. Is it true that for every r there are infinitely many integers not of the form

$$p + 2^{k_1} + \dots \quad + 2^{k_r}$$

and is it true that there is an arithmetic progression no term of which is of the form (2) ? I believe that the answer to the second question is negative for every $r > 2$. In other words I do not think the second question can be answered with the aid of covering congruences, i.e., I think that if $p_1, p_2, \dots p_\ell$ is any set of odd primes and $n > n_o(t)$ then there are always integers k_1 and k_2 so that $2^{k_1} + 2^{k_2} < n$ but $n - 2^{k_1} - 2^{k_2}$ is not a multiple of any of the p's. Crocker answered the first question positively for $r = 2$.

Is it true that if $n \not\equiv 0 \pmod 4$ then

$$n = 2^k + \theta, \qquad \theta \text{ - square free}$$

is always solvable ? I do not believe that covering congruences will help to give a negative answer.

In I and II there are many further problems on covering congruences and also references to the papers mentioned here.

2. This is perhaps my oldest problem. It is more than 50 years old. It was asked independently also by L. Moser. Denote by $h(n)$ the largest integer for which there is a set

$$1 \le a_1 < a_2 < \dots < a_k = n, \qquad k = h(n)$$

so that all the 2^k sums $\sum_{i=1}^{k} \varepsilon_i a_i$ are distinct. Is it true that there is an absolute constant C for which

$$h(n) < \frac{\log n}{\log 2} + C ? \tag{3}$$

L. Moser and I showed that (our proof is given in : P. Erdos, Problems and results in additive number theory, Coll. Theorie des Nombres, Bruxelles (1955), 127-137)

$$h(n) < \frac{\log n}{\log 2} + \frac{\log \log n}{2 \log 2} + C \tag{4}$$

Inequality (4) has not been improved in 40 years. Answering a question of L. Moser and myself Conway and Guy showed that

$$h(n) = \frac{\log n}{\log 2} + 2$$

is possible; say for instance $h(2^{22}) = 24$. But it has been conjectured that

$$h(n) < \frac{\log n}{\log 2} + 3 . \tag{5}$$

I offer 500 dollars for a proof or disproof of (3) and 100 dollars for replacing $1/(2 \log 2)$ in (3) by any smaller constant.

3. Let $a_1 < a_2 < \dots$ be any infinite sequence of integers. Denote by $f(n)$ the number of solutions of $n = a_i + a_j$. Answering a question of Sidon I showed that there is a sequence A for which

$$c_1 \log n < f(n) < c_2 \log n$$

holds for all n. I conjectured that

$$\lim_{n\to\infty} \frac{f(n)}{\log n} = c \ , \qquad 0 < c < \infty \tag{6}$$

is not impossible. Sarkozy and I proved during this meeting that

$$|f(n) - c \log n| = o(\sqrt{\log n}) \tag{7}$$

is impossible. A more general version of (7) will soon appear in a joint paper of Sarkozy and myself dedicated to the memory of Ernst Straus.

Another question posed by Sidon is the following :

An infinite sequence $a_1 \ a_2 \ \dots$ is a $B_k^{(r)}$ sequence of the number of solutions of $a_{i_1} + \dots + a_{i_k} = n$ is at most r. Sidon asked in particular how slowly can a $B_k^{(r)}$ sequence increase ? In particular denote a $B_2^{(1)}$ sequence by a B_2 sequence. Sidon showed that there is a B_2 sequence for which $a_k < k^4$. The greedy algorithm gives $a_k < k^3$ and we both suggested that perhaps there is a B_2 sequence for which $a_k < k^{2+\varepsilon}$ holds. Recently Ajtai, Komlos and Szemeredi proved by an ingenious combinatorial reasoning that there is a B_2 sequence with $a_k < ck^3/\log k$ and so far this is the best upper bound for the growth of a Sidon sequence. It is not difficult to see that for every B_2 sequence we must have

$$\overline{\lim} \ a_k/k^2 = \infty$$

On the other hand as far as I know nobody proved that for a B_3 sequence we must have

$$\overline{\lim} \ a_k/k^3 = \infty$$

An old problem of Turan and myself states that if $f(n) > 0$ for all $n > n_0$ then $\overline{\lim} \ f(n) = \infty$. I offer 500 dollars for proof or disproof of the conjecture. Perhaps $f(n) > 0$ for all $n > n_0$ implies $\overline{\lim} \ f(n)/\log n > 0$ and perhaps if we only assume that $a_k < ck^2$ holds for every k then perhaps $\lim f(n)/\log n > 0$ follows.

Is it true that there is a B_2 sequence for which

$$\underline{\lim} \ \frac{a_k}{k^2} = 1 \ ?$$

If true this is clearly best possible. It would follow from the following combinatorial conjecture which is of interest by itself. Let $a_1 < a_2 < \dots < a_k$ be a sequence of integers satisfying $a_i + a_j \neq a_r + a_s$ i.e., it is B_2 sequence. Then it can be embedded into a perfect difference set, i.e., there is a prime p and $p + 1$ residues $b_1, b_2, \dots, b_{p+1}$ $1 < b_1 < b_2 < \dots < b_{p+1} \leq p^2 + p$ for which $a_i = b_i$ and all the $p^2 + p$ residues

$b_u - b_v$, $u \neq v$ are incongruent (mod p^2+p+1).

For the additive problems discussed in this and the next chapter see the excellent book of Halberstam - Roth.

M.Ajtai, J. Komlos and E. Szemeredi, On a dense Sidon sequence, European Journal of Combinatorics.

4. Let $f(n)=\pm 1$ be an arbitrary number theoretic function. Is it true that to every c there is a d and an m for which

$$\left| \sum_{k=1}^{m} f(kd) \right| > c \text{ ?} \tag{8}$$

Inequality (8) is one of my oldest conjectures. It is more than 50 years old and I offer 500 dollars for a proof or disproof. It is clearly connected with van der Waerden's theorem but I discussed this in many of my papers so I only state one more problem in this connection. Let $\sum \frac{1}{a_i} = \infty$. Is it then true that the a's contain arbitrarily long arithmetic progressions. I offer 3000 dollars for a proof or disproof of this conjecture, which ofcourse would imply that there are arbitrarily long arithmetic progressions among the primes. The longest such progression now on record has 18 terms and is due to Pritchard who presented it at the Denver meeting of the American Mathematical Society in January 1983. (Paul A. Pritchard, Eighteen primes in arithmetic progression, Math. Comp. 41 (1983), 697.) This conjecture is ofcourse connected with the celebrated theorem of Szemeredi.

To conclude this chapter, I state a multiplicative form of (8). Let $f(n) = \pm 1$ be a multiplicative function, i.e., $f(ab)=f(a)f(b)$, when $(a,b)=1$. Is it true then that

$$\overline{\lim} \left| \sum_{k=1}^{n} f(k) \right| = \infty \text{ ?} \tag{9}$$

clearly (9) would follow from (8) but as far as I know (9) has never been proved. Incidentally (9) was conjectured also by Tchudakoff. A possible strengthening of (9) would be that the density of integers n for which

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

is 0 for every c .

5. Now I state some conjectures on prime numbers. Put $d_n = p_{n+1} - p_n$. Turan and I conjectured that $d_n > d_{n+1} > d_{n+2}$ holds for infinitely many n and more generally $d_n > d_{n+1} > \dots > d_{n+k}$ is solvable for every k. To our great annoyance we could

get nowhere with this conjecture. What was even more surprising we could not prove that at least one of the inequalities

$$d_n > d_{n+1} > d_{n+2} \quad \text{or} \quad d_{n+2} > d_{n+1} > d_n \tag{10}$$

has infinitely many solutions. If our conjecture fails then for a certain n_o, $(d_{n_o+i} - d_{n_o+i+1})$ $(-1)^i$ would alway have the same sign, i.e., for some n_o

$$d_{n_o} > d_{n_o+1} , d_{n_o+1} < d_{n_o+2} , d_{n_o+2} > d_{n_o+3} , \dots \tag{11}$$

It is ofcourse quite inconceivable that such an n_o should exist but we could never prove it though perhaps we overlooked a simple idea. *I offer 100 dollars for a proof that an* n_o satisfying (11) does not exist and all the money I can earn, beg, borrow or steal for the proof of the existence of such an n_o ! Consider the k numbers $d_{n+1}, \dots, d_{n+k}$ and order them by size. We ignore the cases if two of them are equal. (Brun's method easily gives that the density of these n's is 0). This gives a permutation of the integers $1, 2, \dots, k$. No doubt all the $k!$ permutations occur, but this ofcourse has never been proved. Denote by $f(k)$ the number of permutations which must occur. Trivially $f(k) \geq k$ (for $k = 2$ this simply means that $d_n > d_{n+1}$ and $d_n < d_{n+1}$ both have infinitely many solutions which is an old result of Turan and myself. $f(3) \geq 4$ is our conjecture and I hope that for large k the inequality $f(k)$ can be considerably improved but I have never carried out the proof.

To end this chapter let me state a few problems about d_n . It seems certain that for every k

$$d_n = d_{n+1} = \dots = d_{n+k}$$

is solvable but even the conjecture that $d_n = d_{n+1}$ has infinitely many solutions seems beyond our reach.

Now I state a few problems about d_n some of which I never considered before, perhaps not all of them are unattackable. Denote by $h(x)$ the largest integer so that for some $n < x$ all the $h(x)$ numbers $d_n, d_{n+1}, \dots, d_{n+h(x)-1}$ are distinct. Estimate $h(x)$ from above and below as well as possible.That $h(x) \to \infty$ follows easily from Brun's method and I expect $h(x) > (\log x)^\alpha$ will also follow for some $\alpha > 0$. I would not be surprised if

$$h(x)/\log x \to 0 \tag{12}$$

but I am not too optimistic about being able to prove (12).

Denote by $r(x)$ the smallest integer t for which $d_n = t, n \leq x$ is not solvable. I would expect that $r(x)/\log x \to \infty$ but this is quite hopeless since even $r(x) \to \infty$ cannot be attacked by methods at our disposal.

Ricci and I proved by using Brun's method that the set of limit points of $d_n/\log n$ has positive measure. There is no doubt that $d_n/\log n$ is in fact everywhere dense in $(0, \infty)$ but this conjecture also is beyond our reach and in fact not a single finite limit point of $d_n/\log n$ is known.

Denote by n_k the product of the first k primes. These problems become much simpler if instead of the primes we consider the integers.

$$1 = a_1 < a_2 < \dots < a_{\varphi(n_k)} = n_k - 1$$

which are relatively prime to n_k, i.e., all whose prime factors are $> p_k$. Perhaps one can determine or estimate the smallest integer $f(k)$ not of the form $a_{i+1} - a_i$. I certainly have not done so but I hope to do so in the future - if there is a future for me !

To end let me state one of my favourite conjectures which is about 45 years old. There is an absolute constant c so that for every n

$$\sum_{i=1}^{\varphi(n)-1} (a_{i+1} - a_i)^2 < \frac{cn^2}{\varphi(n)} \tag{13}$$

Hooley did significant work on (13) but (13) is still open and I offer 500 dollars for a proof or disproof. The prime number analogue of (13) is

$$\sum_{p_k < x} (p_{k+1} - p_k)^2 < c\, x \log x \tag{14}$$

which ofcourse is completely out of reach.

P. Erdos and P. Turan, On some new questions on the distribution of prime numbers, Bull. Amer. Math. Soc. 59 (1948), 685-692.

6. An old conjecture of mine which was probably stated already by Hardy and Littlewood states

$$\pi(x+y) \leq \pi(x) + \pi(y) \tag{15}$$

Without loss of generality we can assume $y \leq x$. To my great surprise Hensley and Richards proved that (15) is incompatible with the prime k-tuple conjecture of Hardy and Littlewood which states : Let $a_1, \ldots, a_k$ be any set of integers which does not form a complete set of residues mod p for any p . Then there are infinitely many integers n so that for every i the integers $n+a_i$, $1 \leq i \leq k$ are all primes. It is ofcourse clear to every "right thinking person" that this conjecture must be true and in fact in the following stronger form : The number of these integers not exceeding x asymptotically

$$\frac{cx}{(\log x)^k}$$

where c is an absolute constant depending only on $a_1, a_2, \ldots, a_k$ and in fact we require that the primes should be consecutive primes. Hardy once remarked that every fool can ask questions about primes which no wise man can answer. In fact a new question about primes has value if it shows up some new aspect. By the way all these questions on k - tuples of primes become easy if primes are replaced by square free numbers. This was certainly known to L. Mirsky decades ago and the only question here would be to determine or estimate the smallest square free number satisfying our conditions.

Now let us return to (15) . E. Straus once remarked that the correct way of stating (15) would have been

$$\pi(x) + 2\pi(\tfrac{y}{2}) \geq \pi(x+y) \tag{16}$$

and indeed (16) is still open.

Hensley and Richards in fact prove that there is an absolute constant c so that

$$\pi(x+y) > \pi(x) + \pi(y) + \frac{cy}{\log^2 y} \tag{17}$$

for every $y > y_o$ and infinitely many values of x. Richards conjectures that for every c there are values of x and y for which (17) holds and I conjecture that c in (17) must be bounded. We ofcourse could not decide who is right. Observe that if Straus' conjecture (16) holds then I am right. Richards and I further have the following plausible and attractive conjecture. Call an integer x 'good' if for all integers $y < x$

$$\pi(x+y) \leq \pi(x) + \pi(y) .$$

Then the density of the good integers is 1. That is, conjecture (15) is only rarely violated. We could only prove that the lower density of good integers is positive. There is no doubt that if x is not good, i.e., if there is a $y < x$ which violates (15), then the number of these y's will be $o(x^{\varepsilon})$ and perhaps $< (\log x)^{c}$; and in fact we conjectured that the largest such y is $o(x^{\varepsilon})$ or even $< (\log x)^{c}$. We could only prove a much weaker result.

D. Hensley and I. Richards, Primes in intervals Acta Arithmetica, 25 (1979), 375-391.

P. Erdos and J. Selfridge, Complete prime subsets of consecutive integers, Proc., Manitoba Conference of Numerical Math., Univ. of Manitoba, Winnipeg 1971, 1-14.

P.Erdos and I. Richards, Density functions for prime and relatively prime numbers, Monatshefte Math., 83 (1977), 99-112.

7. Now I state a few miscellaneous conjectures. Denote by $P(n)$ the greatest prime factor of n . Is it true that the density of integers n for which $P(n) > P(n+1)$ is 1/2 ? Pomerance and I proved a much weaker result but we also proved that for infinitely many n, $P(n) < P(n+1) < P(n+2)$ and conjectured that there are infinitely many integers for which $P(n) > P(n+1) > P(n+2)$ holds. More generally just as in §5 about d_n if we order the number $P(n), P(n+1), \dots, P(n+k)$ by size we get a permutation of the numbers $1, 2, \dots, k$, and we conjectured that each permutation occurs with positive density.

Selfridge and I once considered the following problem : Assume $P(n) = P(m)$, $m>n$. How small can the differences $m-n$ be ? It is not difficult to see that it can be less than $\exp\{(\log x)^{1/2+\varepsilon}\}$ but we did not succeed in getting satisfactory lower bounds.

8. Is it true that for every $\varepsilon > 0$ there are infinitely many primes p for which $P(p-1) < n^{\varepsilon}$? In fact the number of these primes is surely greater than $c_{\varepsilon}\, x/\log x$. This conjecture would imply that the number of solutions of $\varphi(m)=n$ is greater than $n^{1-\varepsilon}$ for suitable n.

9. I conjectured that for every $\varepsilon > 0$ there is an $n_o(\varepsilon)$ so that for every $n > n_o(\varepsilon)$

$$n = a + b, \quad P(ab) < n^{\varepsilon} \tag{18}$$

is solvable. This harmless looking conjecture is probably very difficult. Balog and Sarkozy will publish soon a series of papers on related problems.

One can modify (18) as follows : Put

$$f(n) = \min_{a+b=n} P(ab)$$

Estimate $f(n)$ as well as possible. I expect that

$$\lim_{n\to\infty} \frac{\log f(n)}{\log \log n} = 1/2 . \qquad (19)$$

I think (19) is hopelessly out of reach but a simple computation shows that the limit in (19) if it exists cannot be less than 1/2.

I now mention a rather pretty conjecture of Balog. I stated that very likely for every $\varepsilon > 0$ there are consecutive integers n and $n+1$ for which

$$\min \{P(n), P(n+1)\} > n^{1-\varepsilon} . \qquad (20)$$

This problem led Balog to make the following conjecture which would imply (20) but in my opinion is of more intrinsic interest. Balog's conjecture B_k states as follows : Let $A_k = \{1 \leq a_1 < a_2 < \dots\}$ be an infinite sequence of integers satisfying $a_{i+1} - a_i > 1$ and further if $a \in A_k$ and $a \equiv 0 (\bmod t)$, $P(t) = p_k$ then $\frac{a}{t} \in A_k$ and if $a \in A_k$, $P(t) \leq p_k$ then $t\, a \in A_k$. That is A_k is such that the prime factors $\leq p_k$ can be added or removed freely and we remain in A_k. Now B_k states that the density of A_k is $\leq 1/p_{k+1}$ and a slightly stronger form of B_k states that

$$A_k(x) = \sum_{\substack{a_i \leq x \\ a_k \in A_k}} 1 \leq \left[\frac{x}{p_{k+1}}\right]$$

If the conjecture is true then it is best possible as is shown by the multiples of p_{k+1}. Balog proved his conjecture in a simple and ingenious way for B_1 and he proved for B_2 that the density in this case is $\leq 1/4$. For my problem it would suffice that the density of A_k tends to 0 as $k \to \infty$.

10. To end this paper I state a curious problem which perhaps is not difficult but which caused me lots of trouble. Is it true that there is an absolute constant C so that the number of integers in $\{x, x+n\}$ which have a prime factor p satisfying $\frac{n}{3} < p < \frac{n}{2}$ is at least $c n/\log n$? The result would clearly fail for the primes $\frac{n}{2} < p < n$. To see this put $A = \prod_{\frac{n}{2}<p<n} p$ and consider the interval $(A - \frac{n}{2}, A + \frac{n}{2})$. Clearly A is the only integer in our interval which has prime factor in $(\frac{n}{2}, n)$.

Selfridge and I proved the following curious theorem : For every $\varepsilon > 0$ there is a set of k^2 primes $p_1 < p_2 < \dots < p_{k^2}$ an an interval of length $(3-\varepsilon)p_{k^2}$ which contains only $2k$ distinct multiples of our p's . It is easy to see that every interval of length $> 2p_{k^2}$, contains at least $2k$ distinct multiples of our p_i's. Also I proved that every interval of length greater than $3p_{k^2}$ contains at least $C^{1/2}k$ multiples of the p's. Thus our result with Selfridge is best possible. Nevertheless it is possible that $C^{1/2}k$ can very much be improved and perhaps can be replaced by εk^2.

Our proof with Selfridge is given in : P. Erdos, Problems and results in combinatorial analysis and combinatorial number theory, Southeastern Conference on Combinatorics, Congressum num. 21, (1973), 36 - 38.

11. During my visit in Madras I heard of a forthcoming paper of C. Bantle and F. Grupp. Let $a_1 < a_2 < \dots$ be an *infinite sequence of integers satisfying*

$$(a_i, a_j) = 1, \quad \sum \frac{1}{a_i} < \infty . \tag{21}$$

Let now $b_1 < b_2 < \dots$ be an infinite sequence of integers no one of which has a divisor amongst the a's. Then they prove that

$$b_{i+1} - b_i < b_i^{\frac{9}{20} + \varepsilon} \tag{22}$$

which sharpens previous results of Szemeredi and myself. We ofcourse all expect that (22) is far from being best possible and perhaps in fact $b_{i+1} - b_i < b_i^{\varepsilon}$. If true this must be very deep since it is not even known in the case when $a_1 = p_1^2$ and the b's are the square free numbers. If the b's are the square free numbers then I proved

$$\sum_{b_i < x} (b_{i+1} - b_i)^2 < c.x. \tag{23}$$

and Hooley proved (23) for the exponent 3. It would be nice to try to extend these to the general case.

In this connection the following problems might be of some interest. Determine or estimate the largest integer $f_1(n)$ so that there are integers $\{a_t\}$ satisfying

$$\sum \frac{1}{a_t} < 1, \quad (a_{t_1}, a_{t_2}) = 1$$

and every integer $n < m < n + f_1(n)$ is a multiple of one of the a's. $f_2(n)$ has the same property but here we insist that the a's are primes and for $f_3(n)$ we choose the primes $P(n+i)$ $1 \le i \le f_3(n)$. Finally in case of $f_4(n)$ we drop $(a_i, a_j) = 1$. We only insist that there is an 'a' which is a proper divisor of $n+i$, $1 \le i \le f_4(n)$. I expect that $f_4(n)$ is much larger than $f_i(n)$, $1 \le i \le 3$.

* Mathematics Institute, Hungarian Academy of Sciences, Budapest, HUNGARY.