Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

P. ERDÖS

Extremal Problems in Number Theory, Combinatorics And Geometry

During my long life I wrote many papers on these subjects [1]. There are many fascinating and difficult unsolved problems in all three of these topics. I have to organize the problems in some order. This is not an easy task and anyway not one of my strong points.

In number theory I will mainly discuss questions related to van der Waerden's theorem on long arithmetic progressions and problems in additive number theory.

In geometry the questions I want to discuss are either metrical problems, e.g. number of distinct distances which must occur between points in a metric space. The metric space usually will be our familiar E_2. I will also discuss incidence problems of points in E_2. These problems have a purely combinatorial interpretation too, but the results in E_2 are completely different than in the finite geometries.

In combinatorics I will discuss Sperner, Ramsey and Turán type problems and will try to emphasize their applications to number theory and geometry.

Since I must, after all, remain myself, I can not entirely refrain from stating some old and new problems which, in my opinion, perhaps have been undeservedly neglected.

I hope the reader will forgive a very old man for some personal and historical reminiscences but to save space I will try to write only facts which I did not mention elsewhere.

1. Number theory

First I discuss problems in number theory. Here some of the most striking and significant questions are those connected with the results of van der Waerden and Szemerédi.

Van der Waerden proved more than 50 years ago that if we partition the integers into two classes, at least one of them contains an arbitrarily large arithmetic progression. Many beautiful and important extensions and modifications are known, e.g. the Hales–Jewett theorem and Hindman's theorem but we have no space to discuss these here. A very nice book on this subject has been published recently by Graham, Rothschild and Spencer [2]. (This book contains a very extensive list of references and the references which I suppress here can be found there.) The finite form of van der Waerden's theorem is:

Let $f(n)$ be the smallest integer for which if we divide the integers not exceeding $f(n)$ into two classes, then at least one of them contains an arithmetic progression of n terms. Van der Waerden's original proof gives an explicit upper bound for $f(n)$ but his bound increases very fast: in fact as fast as the well known Ackerman function (which increases so fast that it is not primitive recursive). The best lower bound due to Berlekamp, Lovász and myself increases only exponentially, like a power of 2. The first task would be to prove (or disprove) that $f(n)^{1/n}$ tends to infinity but that $f(n)$ tends to infinity more slowly than Ackerman's function.

Another equally important task would be to find a sharpening of Szemerédi's theorem: Denote by $r_k(n)$ the smallest integer ϱ for which every sequence $a_1 < \ldots < a_\varrho \leqslant n$ contains an arithmetic progression of k terms. Turán and I conjectured 50 years ago that for every k, $r_k(n) = o(n)$. This conjecture was proved for $k = 3$ by K. F. Roth, then later by Szemerédi for $k = 4$ and finally by Szemerédi for every k. A few years ago Fürstenberg proved Szemerédi's theorem by methods of ergodic theory. This proof does not give an explicit upper bound for $r_k(n)$. Fürstenberg and Katznelson proved the n-dimensional generalization of Szemerédi's theorem for which there is so far no other proof. It is not yet possible to tell the potentialities and possible limitations of this new method [3].

K. F. Roth and F. Behrend proved that

$$\frac{n}{e^{c_1\sqrt{\log n}}} < r_3(n) < \frac{c_2 n}{\log\log n}. \tag{1}$$

No useful upper estimation for $r_k(n)$ is known for $k > 3$. Szemerédi and I observed that it is not even known whether $r_k(n)/r_{k+1}(n) \to 0$. It would be very desirable to improve the upper and lower bounds in (1) and to obtain some useful upper bounds for $r_k(n)$. In particular, is it true that

$$r_k(n) < \frac{n}{(\log n)^l} \tag{2}$$

for every k and l if $n > n_0(k, l)$? I offered a reward of \$ 3,000 for a proof or disproof of (2). (2), if true, would of course imply my old conjecture: If $\sum \frac{1}{a_i} = \infty$, then for every k there are k a's forming an arithmetic progression. This in turn would imply that there are arbitrarily long arithmetic progressions in the primes. Recently 18 primes in an arithmetic progression were found by Pritchard [33]. It seems certain that a much stronger result holds for primes: For every k there are k consecutive primes forming arithmetic progression. But this problem certainly cannot be attacked by any of our present-day methods, and is in fact beyond any methods likely to be at our disposal in the near or distant future. Schinzel's well-known hypothesis H would imply it. Van der Corput, Estermann and Tchudakoff independently deduced by Vinogradov's method that the number of even numbers $2n < x$ which are not the sums of two primes in many ways is less than $x/(\log x)^d$ for every d if $x > x_0(d)$. This was later improved by Montgomery and Vaughan to x^{1-c}. (In fact by Goldbach's conjecture all even numbers but 2 are sums of two primes.) These results immediately imply that there are infinitely many triples of primes in an arithmetic progression. It is not yet known whether there are infinitely many quadruples of primes in an arithmetic progression.

An old conjecture of mine (in fact one of my first conjectures, which perhaps did not receive as much attention as it deserved), can be stated as follows: Let $\{f(n)\}$ be an arbitrary sequence with $f(n) = \pm 1$. Then for every $c > 0$ there are an m and a d such that

$$\left| \sum_{k=1}^{m} f(kd) \right| > c. \tag{3}$$

Note that I permit fewer arithmetic progressions here than in van der Waerden's theorem but I also ask for much less. A weaker variant of (3) states that if $f(n) = \pm 1$ and $f(n)$ is multiplicative, then

$$\overline{\lim_{n\to\infty}} \left| \sum_{k=1}^{n} f(k) \right| = \infty. \tag{4}$$

Tchudakoff [4] stated independently in another context a more general conjecture. Here I just would like to call attention to a large class of problems called problems on irregularities of distributions or discrepancy problems. The first results in this subject were found by van der Corput and Aardenne-Ehrenfest and later by K. F. Roth and W. Schmidt, also by Spencer and myself. Recently very striking results were obtained by J. Beck [32]. Very recently these problems were discussed by V. T. Sós in a more general setting. Her paper [30] will appear soon. Here I restrict myself to problems related to van der Waerden's theorem. Denote by $f(n;l)$ the smallest integer for which if we divide the integers not exceeding $f(n;l)$ into two classes, then there is an arithmetic progression of n terms which contains at least $n/2+l$ terms in one of the classes. $f(n;n/2)$ is our old $f(n)$. I easily proved by the probability method that for $l>\varepsilon n$

$$f(n;l)>(1+c_,)^n. \tag{5}$$

For small values of ε, (5) is perhaps not very far from being best possibler It would be very interesting and useful to obtain good upper and lowe. bounds for $f(n;l)$. It would be especially interesting if one could determine the dependence of $f(n;l)$ on l. The trouble is that there are no nontrivial upper bounds known for $f(n;l)$, not even if l is bounded. As far as I know the only result of this kind is due to J. Spencer who determined $f(2n;1)$, i.e. he determined the smallest integer $f(2n;1)$ for which one cannot divide the integers $1,2,\ldots,f(2n;1)$ into two classes so that every arithmetic progression of $2n$ terms contains precisely n terms from each class [5].

2. Combinatorics and additive number theory

Now I discuss problems in combinatorial additive number theory. For a fuller history and discussion of such problems I refer the reader to the excellent book [6] by Halberstam and Roth. Perhaps my oldest conjecture (more than 50 years old) is the following: Denote by A a sequence $1\leqslant a_1<\ldots<a_k\leqslant n$ of integers. Assume that all the sums $\sum_{l=1}^{k}\varepsilon_i a_i$ ($\varepsilon_1 = 0$ or 1) are distinct. Is it true that $k=\log n/\log 2+C$ for some absolute constant C? The powers of 2 show that $C\geqslant 1$. L. Moser raised this problem independently. Moser and I proved that

$$k<\frac{\log n}{\log 2}+\frac{\log\log n}{2\log 2}+O(1).$$

Conway and Guy [7] showed that $C \geqslant 2$. They found 24 integers not exceeding 2^{22} for which all the subset sums are different. It has been conjectured that their construction is perhaps optimal and that $C = 3$.

Now I discuss some problems of Sidon. Sidon called a finite or infinite sequence A a $B_k^{(r)}$ sequence if the number of representations of every integer n as the sum of k or fewer a's is at most r. Let us first assume $k = 2$, $r = 1$, i.e. we assume that the sums $a_i + a_j$ are all distinct. First we consider infinite B_2 sequences (for $r = 1$ a $B_k^{(1)}$ sequence will be denoted by B_k). Sidon asked: Determine the slowest possible growth of a B_2 sequence. The greedy algorithm immediately gives that there is a B_2 sequence for which $a_k < ck^3$. On the other hand I proved that for every B_2 sequence

$$\limsup_{k\to\infty} \frac{a_k}{k^2(\log k)} \geqslant 1. \tag{6}$$

I have been able to improve (6). Further I proved that there is a B_2 sequence for which

$$\liminf \frac{a_k}{k^2} \geqslant \frac{1}{2}. \tag{7}$$

Krückeberg replaced $\frac{1}{2}$ in (6) by $1/\sqrt{2}$ and I conjectured that $1/\sqrt{2}$ can further be improved to 1 which, if true, would be best possible. I could prove (7) if I could prove that if $a_1 < a_2 < \ldots < a_l$ is a B_2 sequence, then it can be embedded into a B_2 sequence $a_1 < \ldots < a_l < a_{l+1} < \ldots < a_k$ with $a_k = (1+o(1))k^2$. Perhaps the following stronger result holds: Every B_2 sequence can be embedded into a perfect difference set. Rényi and I proved by probabilistic methods that to every $\varepsilon > 0$ there is an $r = r(\varepsilon)$ for which there is a $B_2^{(r)}$ sequence satisfying $a_k < k^{2+\varepsilon}$ for every k. I would expect that in fact there is a B_2 sequence satisfying $a_k < k^{2+\varepsilon}$, but the proof of this is nowhere in sight. In fact $a_k < ck^3$ remained unimproved for nearly 50 years. Recently Ajtai, Komlós and Szemerédi [8] proved by a novel and very ingenious combinatorial method that there is a B_2 sequence for which

$$a_k < \frac{ck^3}{\log k}.$$

This new method was recently applied by Komlós, Pintz and Szemerédi to Heilbronn's problem [9].

Denote by $f_k(x)$ the largest integer for which there is a B_k sequence having $f_k(x)$ terms not exceeding x. Turán and I proved that

$$\big(1+o(1)\big)x^{1/2} < f_2(x) < x^{1/2}+cx^{1/4}. \tag{8}$$

Lindstrom proved $f_2(x) < x^{1/2}+x^{1/4}+1$ which at present is the best upper bound for $f_2(x)$.

The lower bound of (8) was also proved by Chowla. Turán and I conjectured that

$$f_2(x) = x^{1/2}+O(1). \tag{9}$$

(9), if true, is probably very difficult. Bose and Chowla proved that for every k

$$f_k(x) \geqslant \big(1+o(1)\big)x^{1/k}.$$

Bose observed that our method with Turán fails to give $f_k(x) \leqslant \big(1+o(1)\big)x^{1/k}$ and in fact this problem is still open. In fact I could never prove that if A is infinite B_k sequence ($k>2$), then

$$\limsup a_l/l^k = \infty.$$

Sidon asked me more than 50 years ago: Denote by $f(n)$ the number of solutions of $n = a_i+a_j$. Is there a basis of order 2 (i.e. every integer is the sum of two a's) for which $f(n) = o(n^\varepsilon)$, for every $\varepsilon > 0$? By probabilistic methods I proved that there is a basis of order 2 for which

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

which is very much stronger than Sidon's conjecture. Turán and I further conjectured that for every basis of order 2 we have

$$\overline{\lim_{n\to\infty}} f(n) = \infty. \tag{10}$$

Perhaps (10) already follows if we assume only that $a_k < ck^2$ holds for some c and every k.

Is there a basis of order 2 for which

$$f(n)/\log n \to 1? \tag{11}$$

Probably (11) will not be quite easy, since (unless I overlook an obvious idea) the probability method does not seem to help with (11).

D. Newman and I conjectured that there is a $B_2^{(2)}$ sequence which is not the union of a finite number of B_2 sequences. Three years ago I found a very simple proof of this conjecture [10]. Nesetril and Rödl proved the related conjecture for $B_2^{(r)}$ sequences. In fact I proved that there is a $B_2^{(2)}$ sequence having n^3 terms no subsequence of which having more than $2n^2$ terms is a B_2 sequence. To see this consider the $B_2^{(2)}$ sequence

$$4^u+4^v, \quad 1 \leqslant u \leqslant n,\ n < v \leqslant n+n^2.$$

This is a $B_2^{(2)}$ sequence of n^3 terms and no subsequence having more than $2n^2$ terms can be a B_2 sequence. To see this consider a complete bipartite graph of n black and n^2 white vertices. The black vertices correspond to the numbers 4^u and the white vertices to 4^v. 4^u+4^v corresponds to the edge joining 4^u and 4^v. A simple graph theoretic argument shows that every subgraph of $2n^2$ edges contains a C_4, i.e. a rectangle. This shows that the subsequence is not a B_2 sequence. Observe that $n^2 = (n^3)^{2/3}$. I cannot decide if the exponent $\frac{2}{3}$ is best possible. Perhaps it could be improved to $\frac{1}{2}$ but I doubt it [11]. V. T. Sós and I considered $B_2'^{(r)}$ sequences, i.e. sequences $a_1, \ldots, a_l, \ldots$, where the number of solutions of $m = a_i - a_j$ is at most r. Of course, for $r=1$ we obtain our familiar B_2 sequences. We could not decide whether there is a $B_2'^{(r)}$ sequence which is not the union of a finite number of B_2 sequences. It is easy to construct a B_2 sequence for which every integer has a unique representation $a_i - a_j$. On the other hand it is easy to see that if $a_k < \lambda k^2$, then the number of solutions of $a_j - a_i = t$ cannot be bounded. We plan to write a paper at a later date on these and other problems on $B_2'^{(r)}$ sequences.

To complete this chapter I state two unsolved problems: Let $\varphi(n) \to 0$ and $1 \leqslant a_1 < a_2 < \ldots < a_x \leqslant n$ be the largest set of integers for which the number of distinct integers of the form $a_i + a_j$ is $\leqslant (1+\varphi(n))\binom{n}{2}$. I can prove that $x \geqslant (1+o(1))\dfrac{2}{3^{1/2}}\, n^{1/2}$, and hope that for some constant $\varepsilon > 0$, $x < (1-\varepsilon)(2n)^{1/2}$. This, if true, would imply that a harmonious graph of n vertices must have fewer than $(1-\eta)\binom{n}{2}$ edges [12] for some constant $\eta > 0$.

Silverman and I asked: Let $h(n)$ be the largest integer for which there is a sequence $0 < a_1 < \ldots < a_{h(n)} \leqslant n$ so that none of the sums $a_i + a_j$ is a square. How large is $h(n)$? This harmless looking question leads to surprising complications [13].

3. Geometrical problems, global results

Next I discuss geometrical problems. Let $x_1, \ldots, x_n$ be distinct points in E_k. Denote by $d(x_i, x_j)$ the distance between x_i and x_j. Let t denote the number of distinct distances and let $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_t$ denote the multiplicities of the distances $\left(\sum_{i=1}^{t} a_i = \binom{n}{2}\right)$. We will mainly study the maximum possible value of a_1 and the minimal value of t. We will study the problems both globally and locally, where "locally" means that we study the distribution of "distances from one point x_i". We will study these problems both if the points are in general position and also if they are restricted in various ways, e.g. they form a convex set or no three are on a line. It will turn out that all these problems lead to various interesting and difficult questions and we are far from their final solutions. Many of them lead to interesting combinatorial extremal problems. G. Purdy and I hope to finish a book on these geometrical problems before this decade is over. We have written several joint papers on this subject.

V. T. Sós and I tried to obtain conditions $\left(\text{other than the trivial } \sum_{i=1}^{t} a_i = \binom{n}{2}\right)$ which would permit us to embed the points in E_k. This question is nontrivial even for $k = 1$ and we have obtained only preliminary results (in many cases with the help of various colleagues). We hope to return to these problems elsewhere. I was told of the following attractive conjecture of Specker. It is easy to see that for every choice of the multiplicities $a_1, \ldots, a_t$ the points can be embedded into E_k for some $k \leqslant n-1$. Specker conjectured that $k = n-1$ is needed only for the regular simplex, i.e., if $a_1 = \binom{n}{2}$. I have never looked seriously at this nice conjecture but I am told that it does not seem to be trivial. V. T. Sós and I raised the following question: Is there an $f(k)$ so that if the n points are in E_k and the minimum multiplicity $a_t > n$ then $n < f(k)$.

Now let us return to our subject. First of all I wish to remark that both on the metrical and on the incidence problems important progress has been made in the last 2 years by J. Beck, J. Spencer, F. Chung, E. Szemerédi and W. Trotter. I have to apologize that I mainly restrict myself to my own problems (not because I consider them more important but because I know more about them). I just want to remark that very recently P. Ungár [31] completely solved a problem of Scott by proving that n points in E_2 determine at least $n-1$ distinct directions (sharpening an earlier result of Burnet and Purdy).

Denote by $f_k(n)$ the largest possible value of a_1 and $g_k(n)$ the smallest possible value of t. Probably $k = 2$ is the most interesting and difficult case. For $k = 1$ everything is trivial: $f_1(n) = g_1(n) = n-1$. For $k > 2$ interesting problems remain, but to save space I will only refer to the literature.

I observed in 1945 that

$$n^{1+c/\log\log n} < f_2(n) < c_2 n^{3/2} \tag{12}$$

and conjectured that the lower bound in (12) is best possible or at least not far from being best possible. The lower bound is given by the triangular or square lattice and perhaps some sort of lattice gives the true lower bound. V. T. Sós and I conjectured that the n points which give $f_2(n)$ must contain an equilateral triangle or a square or at least a set of 4 points which determine at most 2 (or perhaps 3) distinct distances. Further we asked: Is it true that $f_2(n) - a_2 \to \infty$ as $n \to \infty$? Is it true that the configurations which maximize a_1 are the same which minimize t? The answer is almost certainly no. Join two points if their distance is 1. Assume that the distance 1 occurs $f_2(n)$ times. We could get no useful properties of this graph. Of course, it must be connected. It is easy to see that this graph cannot contain a $K(2, 3)$ and this was the way I originally proved $f_2(n) < cn^{3/2}$. With a little trouble we could enumerate all the forbidden subgraphs having fewer than k vertices, as long as k is not too large. Once I hoped that the exclusion of finitely many of these forbidden graphs will give $f_2(n) < n^{1+\varepsilon}$. But now I rather believe that for $n > n_0(k)$ there is a graph of n vertices and $cn^{3/2}$ edges not containing any of the forbidden graphs having $\leqslant k$ vertices.

Szemerédi proved 10 years ago that $f_2(n) = o(n^{3/2})$. Two years ago Beck and Spencer proved $f_2(n) < n^{3/2-\varepsilon}$ for some $\varepsilon > 0$. This was improved by Fan Chung, Szemerédi, Trotter and Spencer to $f_2(n) < n^{4/3}$. Unfortunately their method does not seem to give $f_2(n) < n^{1+\varepsilon}$. I also observed in 1946 that $g_2(n) > \sqrt{n-1} - 1$. This was improved by L. Moser to $g_2(n) > cn^{2/3}$ and last year Fan Chung proved $g_2(n) > cn^{5/7}$. This has also been improved to $g_2(n) > cn^{3/4}$.

4. Distance distribution, local results

I conjectured that if $x_1, x_2, \ldots, x_n \in E_2$ and one denotes by $t_i(n)$ the number of different distances from x_i then

$$\max_i t_i(n) > cn/(\log n)^{1/2}.$$

Beck proved $\max_i t_i(n) > cn^{5/7}$ and this was also improved to $\max_i t_i(n) > n^{3/4}$. In fact I conjectured that

$$\sum_{i=1}^{n} t_i(n) > cn^2/(\log n)^{1/2}. \tag{13}$$

Perhaps (13) is a bit too optimistic but as far as I know no counterexample is known.

I conjectured that for any n points $x_1, \ldots, x_n$ in the plane there is an x_i so that the number of points equidistant from x_i is $\sigma(n^\varepsilon)$ and perhaps it is less than $n^{c/\log\log n}$. The lattice points again show that (if true) this conjecture is best possible. It is trivial that this result holds with $cn^{1/2}$ instead of n^ε and recently Beck proved it with $o(n^{1/2})$. Denote by $a_i(n)$ the largest number of points equidistant from x_i. The most optimistic conjecture is that

$$\sum_{i=1}^{n} a_i(n) < n^{1+c/\log\log n}. \tag{14}$$

Again, perhaps (14) is a bit too optimistic.

5. Distance distributions with conditions

Below we shall assume some additional properties $\mathscr{P}$ of the points $x_1, \ldots, x_n$, and denote the corresponding functions by $f_k(\mathscr{P}, n)$, $g_k(\mathscr{P}, n)$, $t_i(\mathscr{P}, n)$. Let $\mathscr{C}$ denote that $x_1, \ldots, x_n$ form a convex set. I conjectured and Altman proved that

$$g_2(C, n) = \left[\frac{n}{2}\right]. \tag{15}$$

Szemerédi conjectured that (15) remains true if we assume only that no three of the points are on a line, but his proof gives only $g_2(\mathscr{L}, n) \geqslant \left[\frac{n}{3}\right]$ (where $\mathscr{L}$ denotes the above property). L. Moser and I conjectured that

$$f_2(C, 3n+1) = 5n$$

but we could not even prove $f_2(C, n) < cn$. I also conjectured that for a convex set

$$\max_i t_i(C, n) \geqslant \left[\frac{n}{2}\right] \tag{16}$$

but (16) is still open. Perhaps (16) remains true if we assume only that no three of the x_i's are on a line: $\max_i t_i(\mathscr{L}, n) \geqslant [n/2]$. I further conjectured that in the convex case there is always an x_i so that no three of the other vertices are equidistant from it. This was disproved by Danzer but perhaps remains true if 3 is replaced by 4. Here again convexity could be replaced by the condition $\mathscr{L}$.

Let $\mathscr{L}^*$ denote that no three points are on a line, no four on a circle. Is it then true that

$$g_2(\mathscr{L}^*, n)/n \to \infty ?$$

I could not even exclude $g_2(\mathscr{L}^*, n) > cn^2$, but perhaps here I overlook an obvious argument. I could not exclude the possibility that $g_2(\mathscr{L}^*, n) = n-1$ and $a_i = i$, $1 \leqslant i \leqslant n-1$. I thought that this is impossible for $n \geqslant 5$, but colleagues found not quite trivial examples for $n = 5$ and $n = 6$.

Let $\mathscr{F}$ denote that every set of 4 points determines at least 5 distinct distances. Is it then true that $g_2(\mathscr{F}, n) > cn^2$? Is the chromatic number of the hypergraph formed by the quadruples determining $\geqslant 5$ distances bounded? If the points are on a line, this chromatic number is 2.

I could not prove $g_2(\mathscr{F}, n) > cn^2$ even if we assume that every set of 5 points determine at least 9 distinct distances.

Below we delete $\mathscr{P}$ from our notation. Very likely, if our set contains no isosceles triple (i.e., if every set of three points determines three distinct distances) then $g_2(n)/n \to \infty$.

Assume finally that the points are on a line and that every set of 4 points determines at least 4 distinct distances. Then $g_1(n)/n \to \infty$ but $g_1(n)$ can be less than $n^{1+\varepsilon}$. The number of these problems could clearly be continued but it is high time to stop.

6. Incidence problems

Now I discuss incidence problems. Sylvester conjectured and Gallai proved that if n points are given in E_2, not all on a line, then there is always a line which goes through exactly two of the points. The finite geometries show that special properties of the plane (or E_k) must be used here. Motzkin conjectured that for $n > n_0$ there are at least $[n/2]$ such lines. He further observed that, if true, this conjecture is best possible for infinitely many n. Hansen recently proved this conjecture, sharpening a previous result of W. Moser and L. M. Kelly. Hansen's proof has not yet been published.

Croft, Purdy and I conjectured that if $x_1, \ldots, x_n$ is any set of n points in E_2, then the number of lines which contain more than l points is less than cn^2/l^3. The lattice points in the plane show that, if true, this conjecture is best possible. Szemerédi and Trotter recently proved this conjecture. Thus, in particular, there are fewer than $cn^{1/2}$ lines which contain more than $n^{1/2}$ points. The finite geometries again show that special properties of the plane have to be used.

A few weeks later Beck independently proved our conjecture by a different method but in a slightly weaker form. The strong form of our conjecture was needed to prove another conjecture of mine. Denote by $L_1, \ldots, L_m$ the lines determined by our points. By a well-known result of de Bruijn and myself $m \geqslant n$. Denote by $|L_i|$ the number of points on L_i, $|L_1| \geqslant |L_2| \geqslant \ldots \geqslant |L_m|$. I conjectured that the number of distinct sets of cardinalities $\{|L_1|, \ldots, |L_m|\}$ is between $e^{c_1 n^{1/2}}$ and $e^{c_2 n^{1/2}}$. The lower bound is easy and Szemerédi and Trotter proved the upper bound. I have a purely combinatorial conjecture. Let $|S| = n$ and let $A_i \subset S$, $1 \leqslant i \leqslant m$ be a partially balanced block design of S, i.e. every pair $\{x_i, x_j\}$ be contained in one and only one of the A's. I conjectured that the number of distinct sets of cardinalities in $\{|A_1|, \ldots, |A_m|\}$ is between $n^{c_1 n^{1/2}}$ and $n^{c_2 n^{1/2}}$. The upper bound is easy but the lower bound is still open. Rödl recently informed me that the lower bound is also easy.

About 100 years ago Sylvester asked himself the following problem. Assume that no four of the x_i's are on a line. Determine or estimate the largest number of triples of points which are on a line. $\frac{1}{3}\binom{n}{2}$ is a trivial upper bound and Sylvester proved that $n^2/6 - cn$ is possible. The best possible value of c is not yet known. Thus here the difference between the plane and block designs is not so pronounced. A few years ago Burr, Grünbaum and Sloane [14] wrote a comprehensive paper on this subject. They gave a plausible conjecture for the exact maximum. Their paper contains extensive references. Some of their proofs are simplified in a forthcoming peper of Füredi and Palásthy.

Surprisingly, an old conjecture of mine has so far been intractable. Assume $k \geqslant 4$ and that no $k+1$ of our points are on a line. Let $l_k(n)$ be the maximum number of k-tuples which are on a line. Then $l_k(n) = o(n^2)$.

B. Grünbaum proved that $l_k(n) > cn^{1+1/k-2}$ is possible and perhaps this result is best possible.

Dirac conjectured that if $x_1, \ldots, x_n$ are n points not all on a line and we join every two of them, then there is always an x_i so that at least $n/2 - c$ distinct lines go through x_i. If true, then apart from the value of c this is easily seen to be best possible. Szemerédi and Trotter and a few weeks later Beck proved this conjecture with $c_1 n$ instead of $n/2$. Finally Beck proved the following old conjecture of mine. Let there be given n points, at most $n-k$ on a line. Then these points determine at least ckn distinct lines. Perhaps the correct value of c is $\frac{1}{6}$ in any case. Beck gets a very small value of c. Many very interesting questions have completely been omitted, e.g. Borsuk's conjecture [15]. Some more geometric problems will be mentioned in the last chapter on combinatorial problems, where combinatorial theorems directly imply geometric or number-theoretic results.

7. Combinatorial problems

In this final chapter I discuss combinatorial problems. Many mathematicians, including myself, wrote several survey papers on this subject [16] and therefore I will try to keep this chapter short. Also recently appeared an excellent book of Bollobás [17] and several very interesting papers of Simonovits will soon appear. Thus, apart from a few favourite problems, I will mention only results having applications in number theory or geometry. Perhaps the first significant result in this subject is the following theorem of Sperner [18]: Let $|S| = n$, $A_i \subset S$, $1 \leqslant i \leqslant T_n$ be a family of subsets of S no one of which contains the other, then

$$\max T_n = \binom{n}{[n/2]}. \tag{17}$$

Sperner's theorem was forgotten for a long time, perhaps even by its discoverer. When I first met Sperner in Hamburg more than 20 years ago, I asked him about his result. He first thought that I asked him about his much better known lemma in dimension theory, and it turned out that he all but forgot about (17). (17) in fact was used a great deal in the theory of additive arithmetical functions. As far as I know, the first use of (17) was due to Behrend and myself. Behrend and I proved (Behrend [19] a few months earlier) that if $1 \leqslant a_1 < \ldots < a_h \leqslant x$ is a sequence of integers in which no one divides the other, then

$$\sum_{c=1}^{h} \frac{1}{a_i} < \frac{c \log x}{(\log\log x)^{1/2}}. \tag{18}$$

Pillai and both of us observed that (18) is best possible, apart from the value of c. Later Sárközy, Szemerédi and I [20] determined the best value of c in (18). It seems likely that there is no simple characterization of the extremal sequences.

Now I discuss some extremal problems on graphs and hypergraphs. As stated, many papers and the book of Bollobás have appeared recently on this subject, thus I will be very sketchy. Let $G^{(r)}$ be an r-uniform hypergraph (i.e. the basic elements of $G^{(r)}$ are its vertices and r-tuples). For $r = 2$ we obtain the ordinary graphs [21]. $G(n, e)$ will denote a graph (r-uniform hypergraph) of n vertices and e edges (r-tuples). Let $f(n; G^{(r)})$ be the smallest integer so that every $G\big(n; f(n; G^{(r)})\big)$ (i.e. every r-graph of n vertices and $f\big(n; G^{(r)}(n)\big)$ edges (i.e. r-tuples)) contains $G^{(r)}$ as a subgraph. If $r = 2$ and $G^{(2)}$ is a complete graph of l vertices $K(l)$, Turán determined more than 40 years ago $f\big(n; K(l)\big)$ for every l. He also asked for the determination of $f(n; G)$ for more general graphs. Thus started an interesting and fruitful new chapter in graph theory. In particular he asked for the determination of $f\big(n; K^{(r)}(l)\big)$ where $K^{(r)}(l)$ is the complete r-graph of l vertices. This problem is probably very difficult. It is easy to see that

$$\lim_{n\to\infty} f\big(n; K^{(r)}(l)\big) \Big/ \binom{n}{r} = c_{r,l}$$

always exists. $c_{2,l} = 1-1/(l-1)$, but for no $l > r > 2$ is the value of $c_{r,l}$ known. Turán had some plausible conjectures. One possible reason for the difficulty of this problem is that (while Turán proved the uniqueness of his extremal graphs for $r = 2$ and every l) W. G. Brown, and in more general form Kostochka, proved that for $r > 2$ there are many different extremal graphs [22]. Now, for $r = 2$ I state some of our favourite conjectures with Simonovits. It is well known that

$$f(n; C_4) = \big(\tfrac{1}{2} + o(1)\big)\, n^{3/2}. \tag{19}$$

Try to characterize the (bipartite) graphs for which

$$f(n; G) < cn^{3/2}. \tag{20}$$

(One can easily see that if $f(n, G) = o(n^2)$, then G is bipartite.) Our conjecture (perhaps more modestly it should be called a guess) is that (20) holds if any only if G is bipartite and has no subgraph each vertex of which has degree (or valency) greater than 2. Unfortunately we could neither prove the necessity nor the sufficiency of this attractive, illuminating (but perhaps misleading) conjecture. A weaker conjecture, having a better chance of being true, states: Let G satisfy (20). Define G' by adjoining a new vertex to G and join it to two vertices of G of different colour. Then G' also satisfies (20). Further, we conjectured that if G is bipartite, then there is some rational a, $1 \leqslant a < 2$, for which

$$\lim f(n; G)/n^a = c, \quad 0 < c < \infty. \tag{21}$$

Further, for every rational $a \in [1, 2)$ there is a G for which (21) holds [23]. It is well known that (21) is false for $r > 2$, but perhaps for every $G^{(r)}$

$$\lim_{n\to\infty} f(2n; G^{(r)})/f(n; G^{(r)}) = c(G^r) \tag{22}$$

exists and differs from 0 and ∞.

Nearly 50 years ago I investigated the following extremal problem in number theory: Let $1 \leqslant a_1 < a_2 < \ldots < a_h \leqslant x$. Assume that all the products $a_i a_j$ are distinct. Put $f(x) = \max a_h$: estimate $f(x)$ as accurately as possible. I proved that there are positive constants $c_2 \geqslant c_1 > 0$ such that [23]

$$\pi(x) + c_1 x^{3/4}/(\log x)^{3/2} < f(x) < \pi(x) + c_2 x^{3/4}(\log x)^{3/2} \tag{23}$$

(where $\pi(x)$ is the number of primes $\leqslant x$). The proof of (23) was based on the inequality

$$c_3 n^{3/2} < f(n; c_4) < c_4 n^{3/2}. \tag{24}$$

(24) was proved at that time by E. Klein and me.

Another number theoretic application of graph theory is as follows: Denote by $K^{(r)}(t, \ldots, t)$ the r-graph of rt vertices $x_1^{(j)} \ldots x_r^{(j)}$, $1 \leqslant j \leqslant t$ having t^r edges $\{x_{i_1}^{(1)}, x_{i_2}^{(2)}, \ldots, x_{i_t}^{(t)}\}$. I proved (for $t = 2$ this was proved earlier in a sharper form by Kövári and T. Sós and Turán) that

$$f\big(n, K^r(t, \ldots, t)\big) = O(n^{r-1/t^{r-1}}). \tag{25}$$

I deduced from (25) the following result: Let $1 \leqslant b_1 < b_2 < \ldots$ be an infinite sequence of integers. Denote by $g(n)$ the number of solutions of $n = b_i b_j$. Assume that $g(n) > 0$ for all n. Then

$$\limsup_{n=\infty} g(n) = \infty. \tag{26}$$

The additive analog of (26) is an old problem of Turán and mine, and, as stated in Chapter 1, is still open [24].

I just state one more theorem of Simonovits and mine which has direct consequences to some of the problems discussed in Chapter 2. We proved that (for $n > n_0$)

$$f(n; K(r, r, 1)) = \left[\frac{n^2}{4}\right] + \frac{n}{2} + 1 \tag{27}$$

and (27) implies that for $n \equiv 0 \pmod 8$

$$f_4(n) = \frac{n^2}{4} + \frac{n}{2} + 1. \tag{28}$$

The simplicity of (28) is in curious contrast to the difficulty of (10) [25].

Now I have to say a few words on Ramsey's theorem. Very much work has been done on this subject and to save space I only state one or two of my favourite problems and refer to the literature. (The list of references is, of course, far from complete [26].) Let $G_1, \ldots, G_k$ be k graphs, and let $r(G_1, \ldots, G_k)$ be the smallest integer n for which if one colours the edges of the complete graph $K(n)$ by k colours arbitrarily, then for some i, $1 \leqslant i \leqslant k$ the i-th colour contains G_i as a subgraph.

It is surprisingly difficult to get *good* upper or lower bounds for these functions, e.g. it is not yet known whether the limit of $r(K(m); K(m))^{1/m}$ exists. It is known that it is between $2^{1/2}$ and 4. The sharpest known inequality for $r(K(3), K(m))$ states

$$\frac{c_2 m^2}{(\log m)^2} < r((K(3), K(m)) < \frac{c_1 m^2}{\log m}. \tag{29}$$

The proof of (29) uses probabilistic methods. Presumably

$$r(K(r), K(m)) > \frac{cm^{r-1}}{(\log m)^{a_r}} \tag{30}$$

for some constant a_r, but (30) resisted so far all attempts for $r > 3$. It seems very likely that

$$r(K(m), C_4) < m^{2-\varepsilon} \tag{31}$$

holds, but it is not even known that (C_3 is a triangle, $C_3 = K(3)$)

$$r\big(K(m), C_4\big)/r\big(K(m), C_3\big) \to 0. \tag{32}$$

Szemerédi recently observed that

$$r\big(K(m), C_4\big) < \frac{cm^2}{(\log m)^2}. \tag{33}$$

(33), in view of (31), only just fails to prove (32). Ajtai, Komlós and Szemerédi [8] in fact proved the following lemma, which immediately gives (33), and was crucial to the proof of (7):

Trivially, every $G(n; kn)$ has an independent set of size $> n/2k$. Now, if one assumes that our $G(n; kn)$ has no triangles, then the largest independent set has size $> (cn \log k)/k$ (which, apart from the value of the constant, is best possible). In fact, the result remains true even if we assume only that the number of triangles is abnormally small. Several unsolved problems remain, e.g. if we assume only that our $G(n; kn)$ contains no $K(r)$, can we ensure an independent set of size much larger than $n/2k$. The results in this case are not yet in their final form [27].

I just mention one application of Ramsey's theorem. 50 years ago E. Klein asked: Is there an $f(n)$ such that if $x_1, \ldots, x_{f(n)}$ is a set of $f(n)$ points in the plane, no three on a line, then one can always find a subset of n points forming the vertices of a convex n-gon. Szekeres deduced this from Ramsey's theorem. He also conjectured that $f(n) = 2^{n-1}+1$. Later we proved

$$2^{n-1}+1 \leqslant f(n) \leqslant \binom{2n-4}{n-2}.$$

The first unsettled case is whether $f(6) = 17$ or not.

To finish the paper I want to state a conjecture of mine which would have some geometric applications: Is it true that there is an $n = n(\varepsilon)$ so that if $\varepsilon n < k < (\frac{1}{2} - \varepsilon)n$ and $|S| = n$, $A_i \subset S$, $1 \leqslant i \leqslant T_n$ is a family of subsets of S so that for every $1 \leqslant i_1 < i_2 \leqslant T_n$

$$|A_{i_1} \cap A_{i_2}| \neq k,$$

then $T_n < (2-\varepsilon)^n$. Peter Frankl just proved this conjecture.

References

[1] Here is a partial list of my papers on these subjects:

I. Extremal Problems in Number Theory. In: A. L. Whiteman (ed.), *Proc. Symp. Amer. Math.* VIII, *Theory of Numbers*, Amer. Math. Soc., 1963, pp. 181–189.

II. Combinatorial Problems in Geometry and Number Theory. In: D. K. Ray-Chaudhuri (ed.), *Ibid.* XXXIV, *Relations between Combinatorics and Other Parts of Mathematics*, Amer. Math. Soc., 1979, pp. 149–162.

III. On the Applications of Combinatorial Analysis to Number Theory, Geometry and Analysis, *Actes du Congres Int. des Math. Nice 3* (1970), pp. 201–210.

IV. Remarks on Number Theory IV and V, Extremal Problems in Number Theory (in Hungarian), *Mat. Lapok* **13** (1962), pp. 228–253, and *ibid.* **17** (1966), pp. 135–155.

V. On Some Problems in Elementary and Combinatorial Geometry, *Annali di Mat. Pura et Applicata* **53** (1975), pp. 99–108. Most of the results in the chapter on geometry for which I do not give references are found in this paper.

VI. On the Combinatorial Problems which I would Most Like to See Solved, *Combinatorica* **1** (1981), pp. 25–42.

VII. Some Old and New Problems in Various Branches of Combinatorics, *Proc. Tenth Southeastern Conference in Combinatorics Congressus Num. XXIII*, 1979, Vol. 1, pp. 19–37. This paper lists many of my relevant papers. For a large list of problems in combinatorial and elementary number theory see P. Erdös and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, *Monographie de l'Enseignement Math.* **28**, Geneve, 1981, and R. K. Guy, Usolved Problems in Number Theory, *Problem Books in Math.* 1, edited by P. R. Halmos, Springer Verlag, 1982. For problems in geometry see the booklet of W. Moser, *Research Problems in Discrete Geometry*, Dept. of Math., McGill Univ., 1981.

[2] Graham R. L., Rothschild B. L., and Spencer J. H., *Ramsey Theory*, A. Wiley, Interscience Publication, 1980.

[3] Fürstenberg H., *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[4] Tchudakoff N. G., Theory of the Characters of Number Semigroups, *Journal Indian Math. Soc.* **20** (1951), pp. 11–15.

[5] Spencer J., Problem 185, *Canad. Math. Bull.* **16** (1973), p. 464.

[6] Halberstam H. and Roth K. F., *Sequences*, Oxford Univ. Press, 1960. Recently this book was reprinted by Springer-Verlag.

[7] Guy R. K., Sets of Integers, whose Subsets Have Distinct Sums. In: A. Rosa, E. Salidumi and J. Turgeon (eds.), *Theory and Practice of Combinatorics*, Annals of Discrete Math., dedicated to A. Kotzig, North-Holland, 1982, pp. 141–154.

[8] Ajtai M., Komlós J., and Szemerédi E., A Dense Infinite Sidon Sequence, *European J. of Combinatorics* **2** (1981), pp. 1–11.

[9] Komlós J., Pintz J., and Szemerédi E., On Heilbronn's Triangle Problem, *J. London Math. Soc.* **25** (1982), pp. 13–29.

[10] Erdös P., Some Applications of Ramsey's Theorem to Additive Number Theory, *Europ. J. of Combinatorics* **1** (1980), pp. 43–46.

[11] Komlós J., Sulyok M., and Szemerédi E., Linear Problems in Combinatorial Number Theory, *Acta Math. Acad. Sci. Hungar.* **26** (1975), pp. 113–121.

[12] Graham R. L. and Sloane M. T. A., S. I. A. M. J., *Algebraic and Discrete Methods,* On Additive Bases and Harmonious Graphs, Vol. 1 (1980), pp. 382–404.

[13] Lagarias J., Odlyzko A. M., and Shearer J., On the Density of Sequences of Integers the Sum of No Two of Which Is a Square, I: Arithmetic Progressions, II: General Sequences, *J. Comb. Theory, Ser. A* **33** (1982), pp. 167–184 and **34** (1983), pp. 123–139.

[14] Burr S. A., Grünbaum B., and Sloane N. J. A., The Orchard Problem, *Geometriae Dedicata* **2** (1974), pp. 397–424. This paper has extensive references and many open problems and interesting historical remarks.

[15] *Proc. Symp. in Pure Math. VII, Convexity,* Amer. Math. Soc., 1963. In particular: Grünbaum B., Borsuk's Problem and Related Questions, pp. 271–284, for further problems see also Danzer L., Grünbaum B., and Klee V., Helly's Theorem and Its Relatives, pp. 101–180.

For further problems and results on combinatorial geometry see e. g. Grünbaum B., Arrangements and Spreads, *Regional conference series in Math.* **10,** Amer. Math. Soc. Providence (1972); Motzkin Th., The Lines and Planes Connecting the Points of a Finite Set, *Trans. Amer. Math. Soc.* **70** (1951), pp. 451–464; Hadwiger, Debrunner and Klee, *Combinatorial Geometry in the Plane,* Holt, Rinehart and Winston.

Erdös P., Graham R. L., Montgomery P., Rothschild R. L., Spencer T., and Straus E. G., Euclidean Ramsey Theorems, I, II and III, *Journal Comb. Theory, Ser. A* **14** (1973), pp. 341–363; *Infinite and Finite Sets,* Coll. Math. J. Bolyai, North-Holland, 1973, pp. 529–557 and 559–583.

Grünbaum B., *Arrangements and Spreads,* Amer. Math. Soc. (1972).

[16] Burr S. A., Generalized Ramsey Theory for Graphs – A Survey. In: *Graphs and Combinatorics,* Springer-Verlag, Berlin, 1974, pp. 53–75; also: A Survey of Noncomplete Ramsey Theory for Graphs, *Annals of the New York Academy of Sciences, Topics in Graph Theory* **328** (1979), pp. 58–75.

Chung F. R. K., and Grinstead L. M., A Survey of Bounds for Classical Ramsey Numbers, *Journal of Graph Theory* **7** (1983), pp. 25–38.

Erdös P., and Kleitman D. J., Extremal Problems among Subsets of a Set, *Proc. Second Chapel Hill Conf. on Comb. Math.,* Univ. of North Carolina, Chapel Hill, 1970, pp. 146–170; see also *Discrete Math.* **8** (1974), pp. 281–294, and Erdös P., *Art of Counting,* M.I.T. Press, 1973.

Erdös P. and Simonovits M., Some Extremal Problems in Graph Theory, Combinatorial Theory and Its Application, *Proc. Coll. Balatonfüred* (1969), pp. 377–390.

Katona Gy., Extremal Problems for Hypergraphs, *Combinatorics, Proc. MATO Advanced Study, Breukelen 1979,* Part 2, pp. 13–42.

Kleitman D. J., On an Extremal Property of Antichains in Partial Orders, *ibid.* pp. 77–90.

[17] Bollobás B., *Extremal Graph Theory,* Academic Press, London and New York, 1978.

Simonovits M., Extremal Graph Theory. In: Beineke and Wilson (eds.), *Selected Topics in Graph Theory II,* Academic Press.

Simonovits M., Extremal Graph Problems, Degenerate Extremal Problems and Supersaturated Graphs, *Proc. Conference on Silver Jubilee of University of Waterloo* (1982).

[18] Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Zeitschrift* **27** (1928), pp. 544–548.
[19] Behrend F., On Sequences of Numbers Not Divisible One by Another, *J. London Math. Soc.* **10** (1935), pp. 42–49; see laso: Elliott T. D. A., *Probabilistic Number Theory*, Vol. 1, *Grundlehren der Math. Wissenschaften* **239**, Springer-Verlag, pp. 214–216.
[20] Erdös P., Sàrközy A., and Szemerédi E., On Divisibility Properties of Sequences of Integers, *Number Theory, Coll. Math. Soc. J. Bolyai* **2**, North-Holland, 1970, pp. 35–47.
[21] Berge C., *Graphs et hypergraphs*, Monographies Universitaires de Mathematiques, Dunod, 1970.
[22] Kostochka A. V., A Class of Constructions for Turán's 3–4 Problem, *Combinatorica* **2** (1982), pp. 192–198.
[23] Erdös P., On some Applications of Graph Theory to Number Theoretic Problems, *Publ. Ramanujan Inst.* **1** (1969), pp. 131–136.
[24] Erdös P., On Extremal Problems of Graphs and Generalized Graphs, *Israel T. Math.* **2** (1964), pp. 183–190; and On the Multiplicative Representation of Integers, *ibid.*, pp. 251–261. For further problems on hypergraphs see Erdös P., On Some Extremal Problems on r-Graphs, *Discrete Math.* **1** (1971), pp. 1–6.
[25] Erdös P., On Some Applications of Graph Theory to Geometry, *Canad, J. Math.* **19** (1967), pp. 968–971; Simonovits M., A Method for Solving Extremal Problems in Graph Theory Stability Problems, *Proc. Coll. on Graph Theory Held in Tihany (Hungary) 1966*; see also: Erdös P., Extremal Problems in Graph Theory, *Proc. Symp. Graph Theory, Smolenice 1963*, pp. 29–36. For further extremal results see Erdös P. and Simonovits M., An Extremal Graph Problem, *Acta Math. Acad. Sci. Hung.* **22** (1971), pp. 275–282.
[26] See e.g. the papers of Burr quoted in [16]; also: Erdös P. and Spencer J., *Probabilistic Methods in Combinatorics*, Acad. Press, New York, 1974.
[27] Ajtai M., Erdös P., Komlós J., and Szemerèdi E., On Turàn's Theorem for Sparse Graphs, *Combinatorica* **1** (1981), pp. 313–318.
[28] Erdös P. and Szekeres E., A Combinatorial Problem in Geometry, *Compositio Math.* **2** (1935), pp. 963–970; and: On Some Extremal Problems in Elementary Geometry, *Ann. Univ. Sci. Budapest. Sect. Math.* **3–4** (1961), pp. 53–62; see also: *The Art of Counting, Selected Writings of P. Erdös*, MIT Press, 1973.
[29] Frankl P. and Wilson R. M., Intersection Theorems with Geometric Consequences, *Combinatorica* **1** (1981), pp. 357–368.
[30] Sós V. T., Irregularities of Partitions, Ramsey Theory, Uniform Distribution, *Surveys in Combinatorics*, Invited papers for the 9th British Combinatorial Conference 1983, pp. 201–245.
[31] Ungàr P., $2n$ Non Collinear Points Determine at least $2n$ Directions, *J. Comb. Theory* **33** (1982), pp. 343–347.
[32] Beck J., On a Problem of K. F. Roth Concerning Irregularities of Point Distributions.
[33] Pritchard P.A., Eighteen Primes in an Arithmetic Progression, *Math. Comp.* **41** (1983), p. 697.