

LINEAR PROBLEMS IN COMBINATORIAL NUMBER THEORY

By

J. KOMLÓS, M. SULYOK and E. SZEMERÉDI (Budapest)

Introduction

In many problems of combinatorial number theory one has to estimate the density of a sequence of positive integers, where the only restriction on the sequence is that some linear relations do not hold for its subsets. We mention two such problems which have been investigated by many authors. The first one is the Sidon-problem. A sequence a_i (finite or infinite) is called a B_2 -sequence if every integer can be written in at most one way in the form $a_i + a_j$. In general, a sequence is a B_h -sequence if the sums $a_{i_1} + \dots + a_{i_h}$ are all different. Let $F_h(n)$ denote the largest number k for which there exists a B_h sequence $a_1 < a_2 < \dots < a_k \leq n$. Sidon applied the estimations on $F_2(n)$ for Fourier-analysis. Erdős, Turán, Rényi, Singer, Chwila and many others investigated the asymptotic behaviour of the functions $F_h(n)$ and proved that $F_2(n) \sim \sqrt{n}$ and $c_1 n^{1/h} < F_h(n) < c_2 n^{1/h}$. (Further base-problems can be found in [1].)

The following problem is similar in character, but is much harder: How many numbers can be chosen from the first n integers not containing an arithmetic progression of k terms. Denote this number by $r_k(n)$. ROTH and BEHREND [2], [3] proved that

$$n^{1 - \frac{\varepsilon}{\sqrt{\log n}}} < r_3(n) < \frac{n}{\log \log n},$$

and SZEMERÉDI [4], [5] showed that for all $k \geq 3$, $r_k(n) = o(n)$.

One can ask the same questions for arbitrary n integers (not necessarily the first n).

The results in this direction are much weaker. In the case of the Sidon problem the corresponding lower bound is only $n^{1/3}$, for the problem of arithmetic progression — to our knowledge — no lower bound is known. However, one can expect that the case of the first n integers is “the worst case”. That is the statement we are going to formulate and prove in this paper under very general circumstances.

§ 1. The theorem

We are given a system of linear equations

$$\sum_{i=1}^r \alpha_i^{(l)} x_i = 0, \quad l = 1, 2, \dots, L.$$

We say that the numbers x_1, \dots, x_r satisfy relation ϱ (and denote by $\varrho(x_1, \dots, x_r)$) if all the above equations hold true. Define relation φ for sets A of integers as follows:

$\varphi(A)$ if A contains no r -tuples (x_1, \dots, x_r) , $\{x_1, \dots, x_r\} \subset A$ satisfying relation ϱ . We obtain different relations if in the definition of φ we restrict ourselves to r -tuples of different numbers x_1, \dots, x_r , or allow arbitrary r -tuples. Our theorem will apply to both types of relations.

$|A|$ denotes the number of elements in A , and we put

$$\|A\| = \max_{\substack{B \subset A \\ \varphi(B)}} |B|.$$

Define the functions $f(n)$ and $g(n)$:

$$f(n) = \|\{1, 2, \dots, n\}\|$$

and

$$g(n) = \min_{|A|=n} \|A\|.$$

Obviously $g(n) \leq f(n)$.

Now the statement mentioned in the introduction can be formulated as follows:

THEOREM. *For every linear relation ϱ there exists a positive number $c=c(\varrho)$ for which $g(n) > cf(n)$ for all n large enough.*

Obviously, we can restrict ourselves for rational and thus for integer coefficients $\alpha_i^{(l)}$, as we do in the sequel.

A relation ϱ is called invariant under translation if

$$\sum_{i=1}^r \alpha_i^{(l)} = 0, \quad l = 1, 2, \dots, L.$$

Put

$$\alpha = \max_{1 \leq l \leq L} \sum_{i=1}^r |\alpha_i^{(l)}|.$$

REMARK 1. In the case of relations ϱ invariant under translation the constant $c(\varrho)$ does not depend on the relation ϱ , only on the number α . Probably it is also true for relations not invariant under translation, but we could not prove it. (It is also possible that in both cases c is an *absolute* constant.)

REMARK 2. It is easy to see that in the case of relations not invariant under translation $f(n) > cn$ and we will prove that $g(n) > c'n$ (Lemma 7). In the case of the above mentioned translation-invariant relations $F_k(n)$, $r_k(n)$ one has $f(n) = o(n)$, and SZEMERÉDI's result [5] shows that this is so for every translation-invariant relation.

§ 2. The proof of the theorem

All over the proof the number n is assumed to be large enough so that all approximations hold true.

We will often use the following simple but basic remark:

REMARK 3. Given $a_1 < \dots < a_n$ and a positive integer q . Assume the numbers a_i can be written as $a_i = h_i q + r_i$ with $|r_i| < \frac{q}{\alpha}$, $i = 1, 2, \dots, n$. Then the mapping