REPRINTED FROM
NUMBER THEORY AND ALGEBRA
© 1977
ACADEMIC PRESS, INC.
NEW YORK SAN FRANCISCO LONDON

On Extremal Density Theorems for Linear Forms

R. L. GRAHAM *H. S. WITSENHAUSEN*

BELL LABORATORIES
MURRAY HILL, NEW JERSEY

J. H. SPENCER†

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS‡

A typical question in extremal number theory is one which asks how large a subset R may be selected from a given set of integers so that R possesses some desired property. For example, it is not difficult to see that if R is a subset of the integers $[1, 2, \ldots, 2N]$ *and R has more than N elements then there are integers x and y in R so that* $x + y$ *is also in R. The sets* $\{1, 3, 5, \ldots, 2N - 1\}$ *or* $\{N + 1, N + 2, \ldots, 2N\}$ *show that this bound cannot be improved.*

In this note we prove several general results of this type. In particular, we show that if $R \subseteq \{1, 2, \ldots, N\}$ *and R has more than* $N - [N/n]$ *elements, then for some integers x and y, the integers* $x, x + y, x + 2y, \ldots, x + (n - 1)y$ *and y all belong to R. Furthermore the bound* $N - [N/n]$ *is best possible.*

† The work done by this author was done while he was a consultant at Bell Laboratories.
‡ Present address: SUNY at Stony Brook, Stony Brook, New York.

1. Introduction

Suppose $\mathscr{L} = \{L_i(x_1, \ldots, x_m) \equiv \sum_{j=1}^{m} a_{ij}x_j\colon 1 \le i \le n\}$ is a set of linear forms in the variables x_j with integer coefficients a_{ij}. The question we consider is the following:

How large may a subset R of $\{1, 2, \ldots, N\}$ be so that for every choice of positive integers t_j, $1 \le j \le m$, at least one of the values $L_i(t_1, \ldots, t_m)$, $1 \le i \le n$, is not in R.

Unfortunately, this question appears to be rather difficult and very few general results are currently available. In this paper we study this problem for several important special sets $\mathscr{L}$. It will be seen that even in these simple cases, the problem is not without interest.

2. Preliminaries

Let $[1, N]$ denote the set $\{1, 2, \ldots, N\}$. If $\mathscr{L} = \{L_i(x_1, \ldots, x_m)\colon 1 \le i \le n\}$ is a set of linear forms, we say that a set $R \subseteq [1, N]$ is *$\mathscr{L}$-free* if for any choice of positive integers $t_1, \ldots, t_m$, at least one of the values $L_i(t_1, \ldots, t_m)$ does not belong to R. If R is not $\mathscr{L}$-free, we say that $\mathscr{L}$ *hits* R. Define

$$S_{\mathscr{L}}(N) = \max_R \, |R|$$

where the max is taken over all $R \subseteq [1, N]$ that are $\mathscr{L}$-free and $|R|$ denotes the cardinality of R. Also, define $\delta(\mathscr{L})$, called the *critical density* of $\mathscr{L}$, by

$$\delta(\mathscr{L}) = \liminf_N S_{\mathscr{L}}(N)/N.$$

As an example, consider the system $\mathscr{L}_n = \{x_1 + kx_2\colon 0 \le k < n\}$. The condition that R is $\mathscr{L}_n$-free means exactly that R contains no arithmetic progression of n terms.

For this example, a recent result of Szemerédi [2], however, asserts that any infinite set of integers of positive upper density contains arbitrarily long arithmetic progressions. From this it follows at once that $\delta(\mathscr{L}_n) = 0$.

3. Augmented Arithmetic Progressions

We now consider a system closely related to $\mathscr{L}_n$ which we denote by $\mathscr{L}_n^*$. It is defined by

$$\mathscr{L}_n^* = \{x_1 + kx_2\colon 0 \le k < n\} \cup \{x_2\}.$$

In this case, $\mathscr{L}_n^*$ hits R if and only if R contains an arithmetic progression of

n terms together with the common difference of the progression. However, the critical density of $\mathscr{L}_n^*$ differs sharply from that of $\mathscr{L}_n$ as the following examples indicate.

Example 1 Let $R_1 \subseteq [1, N]$ be defined by

$$R_1 = \{x \in [1, N]: x > [N/n]\}.$$

Clearly R_1 is $\mathscr{L}_n^*$-free since

$$t_1 + (n-1)t_2 \geq n(1 + [N/n]) > N \qquad \text{for} \quad t_1, t_2 \in R_1.$$

Thus

$$\delta(\mathscr{L}_n^*) \geq 1 - n^{-1}. \tag{1}$$

Example 2 Suppose n is prime and let $R_2 \subseteq [1, N]$ be defined by

$$R_2 = \{x \in [1, N]: x \not\equiv 0 \pmod{n}\}.$$

Then $\mathscr{L}_n^*$ cannot hit R_2 since for *any* integers t_1 and t_2, either $t_2 \equiv 0 \pmod{n}$ or $t_1 + kt_2$, $0 \leq k < n$, runs through a complete residue system modulo n and therefore represents $0 \notin R_2$. Note that

$$|R_2| = N - [N/n] = |R_1|. \tag{2}$$

The following result shows that equality holds in (1) and, in fact, (2) is best possible.

Theorem 1 *Suppose* $R \subseteq [1, N]$ *with* $|R| > N - [N/n]$. *Then* $\mathscr{L}_n^*$ *hits* R.

Proof Let R satisfy the hypothesis of the theorem and suppose R is $\mathscr{L}_n^*$-free. Let Δ denote the least element of R. Then we may assume

$$\Delta \leq [N/n] \tag{3}$$

since otherwise $|R| \leq N - [N/n]$. Define the arithmetic progressions $T_i \subseteq [1, N]$ by

$$T_i = \{i + k\Delta: 0 \leq k < n\}, \qquad 1 \leq i \leq N - (n-1)\Delta.$$

Also, define $A_j, A'_j \subseteq [1, N]$ for $1 \leq j \leq n$ as follows:

$$A_j = \begin{cases} [(j-1)\Delta + 1, j\Delta] & \text{for} \quad 1 \leq j < n, \\ [(n-1)\Delta + 1, N] & \text{for} \quad j = n; \end{cases}$$

$$A'_j = \begin{cases} [N - j\Delta + 1, N - (j-1)\Delta] & \text{for} \quad 1 \leq j < n, \\ [1, N - (n-1)\Delta] & \text{for} \quad j = n. \end{cases}$$

By (3), we see that

$$|A_n| = |A'_n| \geq \Delta.$$

Also, it is easily checked that if $x \in A_j \cap A'_{j'}$ then $j + j' = n + t$ for some t, $1 \le t \le n$, and

$$|\{i: x \in T_i\}| = t. \tag{4}$$

We claim the following equation holds:

$$n|R| = \sum_{i=1}^{N-(n-1)\Delta} |T_i \cap R| + \sum_{j=1}^{n-1} (n-j)(|A_j \cap R| + |A'_j \cap R|). \tag{5}$$

To prove (5), let $x \in R$. Then for some k and k', $x \in A_k \cap A'_{k'}$. Since the A_j are disjoint, as are the A'_j, then the contribution x makes to the second sum on the right-hand side of (4) is just $(n-k) + (n-k')$. Let $k + k' = n + t$. Hence, by (4), x contributes exactly t to the first sum in (5). Therefore, each $x \in R$ contributes exactly

$$(n-k) + (n-k') + (k+k'-n) = n$$

to the right-hand side of (5) so that Eq. (5) is indeed valid. But by hypothesis, since $\Delta \in R$, then $|T_i \cap R| \le n - 1$ for all i. Thus, since $|A_1 \cap R| = 1$, then by (5)

$$\begin{aligned} n|R| &\le (n-1)(N-(n-1)\Delta) + 2\Delta \sum_{j=1}^{n-1} (n-j) - (n-1)(\Delta - 1) \\ &= (n-1)N + \Delta(-(n-1)^2 + n(n-1) - (n-1)) + n - 1 \\ &= (n-1)(N+1), \end{aligned} \tag{6}$$

which implies

$$|R| \le \left[\frac{(n-1)(N+1)}{n}\right] = N - \left[\frac{N}{n}\right]. \tag{7}$$

This proves Theorem 1. ∎

Of course, it follows from (1) and (7) that

$$S_{\mathscr{L}_n^*}(N) = N - [N/n] \tag{8}$$

and consequently

$$\delta(\mathscr{L}_n^*) = 1 - n^{-1}.$$

4. Forms in One Variable—A Special Case

As a prelude to a discussion in the next section of the general case of linear forms in one variable (i.e., with $m = 1$), we consider first the special case $\mathscr{L} = \{x, 2x, 3x\}$. This example in fact has all the essential features of the general case.

To begin, we let $D = \{d_1 < d_2 < \cdots\}$ denote the set of all integers of the form $2^a 3^b$, $a, b \geq 0$.

Let N be a fixed positive integer. For $1 \leq t \leq N$ with $(t, 6) = 1$, let $C(t)$ denote the set

$$C(t) = [1, N] \cap \{td_k: k = 1, 2, \ldots\}.$$

Note that a set $R \subseteq [1, N]$ is $\mathscr{L}$-free if and only if $R(t) = R \cap C(t)$ is $\mathscr{L}$-free for all t with $(t, 6) = 1$. For indeed, $\mathscr{L}$ can hit R only if for some x, $\{x, 2x, 3x\} \supseteq R$. However, this implies that $\mathscr{L}$ hits $R(t)$ for some t relatively prime to 6. Thus, a maximal $\mathscr{L}$-free set R is formed by taking the union of maximal $\mathscr{L}$-free subsets from $C(t)$ for each t, $(t, 6) = 1$. However, it is clear that

$$X_t = \{td_k: k = 1, \ldots, r\} \subseteq C(t)$$

is $\mathscr{L}$-free if and only if $X_1 = \{d_k: k = 1, \ldots, r\} \subseteq C(1)$ is $\mathscr{L}$-free. Thus, if $f(r)$ denotes the cardinality of the largest $\mathscr{L}$-free subset of $\{d_1, \ldots, d_r\}$ and $h(r)$ denotes the number of $t \in [1, N]$, $(t, 6) = 1$, with $|C(t)| = r$, then for any $\mathscr{L}$-free set $R \subseteq [1, N]$,

$$|R| \leq \sum_{r=1}^{\infty} f(r)h(r). \tag{9}$$

For fixed r, $|C(t)| = r$ if and only if

$$td_r \leq N < td_{r+1}$$

i.e.,

$$N/d_{r+1} < t \leq N/d_r .$$

Thus,

$$h(r) \to \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)N\left(\frac{1}{d_r} - \frac{1}{d_{r+1}}\right) \quad \text{as} \quad N \to \infty \tag{10}$$

and, therefore, for maximal $\mathscr{L}$-free sets $R_N \subseteq [1, N]$,

$$\lim_{N\to\infty} \frac{|R_N|}{N} = \frac{1}{3} \sum_{r=1}^{\infty} f(r)\left(\frac{1}{d_r} - \frac{1}{d_{r+1}}\right). \tag{11}$$

But

$$f(r + 1) - f(r) \leq 1,$$

so that letting $K(\mathscr{L})$ denote the set $\{k: f(k) > f(k-1)\}$, the telescoping sum in (11) becomes

$$\delta(\mathscr{L}) = \frac{1}{3} \sum_{k \in K(\mathscr{L})} \frac{1}{d_k}. \tag{12}$$

Unfortunately, there does not seem to be any simple way to determine the elements of $K(\mathscr{L})$. The first few values are given in Table 1.

TABLE 1

k	$f(k)$	k	$f(k)$	k	$f(k)$
1	1	13	9	25	17
2	2	14	10	26	18
3	2	15	11	27	18
4	3	16	11	28	19
5	4	17	12	29	20
6	5	18	13	30	20
7	5	19	13	31	21
8	6	20	14	32	22
9	7	21	14	33	22
10	7	22	15	34	23
11	8	23	16	35	24
12	8	24	17	36	25

Thus,

$$K(\mathscr{L}) = \{1, 2, 4, 5, 6, 8, 9, 11, 13, 14, 15, 17, 18, 20,$$
$$22, 23, 24, 26, 28, 29, 31, 32, 34, 35, 36, \ldots\}. \tag{13}$$

It may be that $f(k) = 1 + [2k/3]$ if $k \not\equiv 0 \pmod 3$ and, perhaps, for all k, there is always a maximal $\mathscr{L}$-free set

$$R_k = \{2^{a_i}3^{b_i}: i = 1, \ldots, f(k)\} \subseteq \{d_1, \ldots, d_k\}$$

in which all $a_i - b_i$ are congruent modulo 3.

It would also be interesting to know if $\delta(\mathscr{L})$ is irrational.

5. Forms in One Variable—The General Case

Let $\mathscr{L}$ denote the set of linear forms $\{a_1 x, \ldots, a_n x\}$ where $A = \{a_1 < \cdots < a_n\}$. Let $P(A) = \{q_1, \ldots, q_r\}$ be the set of primes dividing the a_i and let $D^{(\mathscr{L})} = (d_1 < d_2 < \cdots)$ denote the set of all integers of the form

$q_1^{\alpha_1} \cdots q_r^{\alpha_r}$, $\alpha_i \geq 0$. For each k let $f(k)$ denote the cardinality of a maximal $\mathscr{L}$-free subset of $\{d_1, \ldots, d_k\}$. Finally, let $K(\mathscr{L})$ be defined by

$$K(\mathscr{L}) = \{k : f(k) > f(k-1)\}.$$

By using essentially the same arguments as in the previous section, the following theorem can be proved.

Theorem 2

$$\delta(\mathscr{L}) = \prod_{j=1}^{r} (1 - q_j^{-1}) \sum_{k \in K(\mathscr{L})} d_k^{-1} \tag{14}$$

6. Concluding Remarks

One problem with a representation such as (14) is that it is not clear how to describe $K(\mathscr{L})$ so as to be able to evaluate $\sum_{k \in K(\mathscr{L})} d_k^{-1}$. Several systems $\mathscr{L} = \mathscr{L}(a_1, \ldots, a_n) = \{a_1 x, \ldots, a_n x\}$ of forms in one variable are known, however, for which such a description can be given. We list a sample of these below. The arguments needed to determine the sets $K(\mathscr{L})$ are not difficult and are omitted.

1. $\delta(\mathscr{L}(1, p, p^2, \ldots, p^{m-1})) = (p^m - p)/(p^m - 1)$ for p prime. Thus, $\delta(\mathscr{L}(1, 2)) = \frac{2}{3}$ as expected.
2. $\delta(\mathscr{L}(1, n)) = n/(n+1)$.
3. $\delta(\mathscr{L}(2, 3)) = \frac{3}{4}$.
4. $\delta(\mathscr{L}(1, 2, 8)) = \frac{57}{62}$. Some recent results of Harlambis [1] are relevant here.

It seems quite likely that almost all systems $\mathscr{L}$ have $\delta(\mathscr{L})$ irrational although not even *one* such $\mathscr{L}$ is known at present!

REFERENCES

[1] N. M. Harlambis, "Sets with missing differences or missing patterns," PhD Dissertation, Univ. California, Los Angeles, 1973.

[2] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.