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 $N - \lceil N/n \rceil$ 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 $N - \lceil N/n \rceil$ 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 $\mathcal{L} = \{ L_i(x_1, \ldots, x_m) = \sum_{j=1}^m a_{ij} x_j : 1 \leq i \leq n \}$ 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 $t_j$, $1 \leq j \leq m$, at least one of the values $L_i(t_1, \ldots, t_m)$, $1 \leq i \leq n$, 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 $\mathcal{L}$. 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 $\mathcal{L} = \{ L_i(x_1, \ldots, x_m) : 1 \leq i \leq n \}$ is a set of linear forms, we say that a set $R \subseteq [1, N]$ is $\mathcal{L}$-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 $\mathcal{L}$-free, we say that $\mathcal{L}$ hits $R$. Define

$$S_\mathcal{L}(N) = \max_{R} |R|$$

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

$$\delta(\mathcal{L}) = \liminf_{N} S_\mathcal{L}(N)/N.$$

As an example, consider the system $\mathcal{L}_n = \{ x_1 + kx_2 : 0 \leq k < n \}$. The condition that $R$ is $\mathcal{L}_n$-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 $\delta(\mathcal{L}_n) = 0$.

3. Augmented Arithmetic Progressions

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

$$\mathcal{L}_n^* = \{ x_1 + kx_2 : 0 \leq k < n \} \cup \{ x_2 \}.$$

In this case, $\mathcal{L}_n^*$ 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 \( \mathcal{L}_n^* \) differs sharply from that of \( \mathcal{L}_n \) as the following examples indicate.

**Example 1** Let \( R_1 \subseteq [1, N] \) be defined by

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

Clearly \( R_1 \) is \( \mathcal{L}_n^* \)-free since

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

Thus

\[
\delta(\mathcal{L}_n^*) \geq 1 - n^{-1}.
\]

(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 \( \mathcal{L}_n^* \) 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 \( 0 \not\in R_2 \). Note that

\[
|R_2| = N - \lfloor N/n \rfloor = |R_1|.
\]

(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 \( |R| > N - \lfloor N/n \rfloor \). Then \( \mathcal{L}_n^* \) hits \( R \).

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

\[
\Delta \leq \lfloor N/n \rfloor
\]

(3)

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

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

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

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

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

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 \( t \), \( 1 \leq t \leq n \), 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 \( |T_i \cap R| \leq n-1 \) for all \( i \). Thus, since \( |A_1 \cap R| = 1 \), then by (5)

\[
n |R| \leq (n-1)(N-(n-1)\Delta) + 2\Delta \sum_{j=1}^{n-1} (n-j) - (n-1)(\Delta - 1)
\]

\[
= (n-1)N + \Delta(-n(n-1)^2 + n(n-1) - (n-1)) + n-1
\]

\[
= (n-1)(N+1), \tag{6}
\]

which implies

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

This proves Theorem 1. \( \square \)

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

\[
S_{\mathcal{L}^*}(N) = N - \left[\frac{N}{n}\right] \tag{8}
\]

and consequently

\[
\delta(\mathcal{L}^*_n) = 1 - n^{-1}.
\]

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 \( \mathcal{L} = \{x, 2x, 3x\} \). 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 \( 2^a3^b \), \( a, b \geq 0 \).

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 \( \mathcal{L} \)-free if and only if \( R(t) = R \cap C(t) \) is \( \mathcal{L} \)-free for all \( t \) with \( (t, 6) = 1 \). For indeed, \( \mathcal{L} \) can hit \( R \) only if for some \( x, \{x, 2x, 3x\} \supseteq R \). However, this implies that \( \mathcal{L} \) hits \( R(t) \) for some \( t \) relatively prime to 6. Thus, a maximal \( \mathcal{L} \)-free set \( R \) is formed by taking the union of maximal \( \mathcal{L} \)-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 \( \mathcal{L} \)-free if and only if \( X_1 = \{d_k: k = 1, \ldots, r\} \subseteq C(1) \) is \( \mathcal{L} \)-free. Thus, if \( f(r) \) denotes the cardinality of the largest \( \mathcal{L} \)-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 \( \mathcal{L} \)-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 \( \mathcal{L} \)-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 $K(\mathcal{L})$ denote the set \{\(k \colon f(k) > f(k - 1)\)\}, the telescoping sum in (11) becomes

\[
\delta(\mathcal{L}) = \frac{1}{3} \sum_{k \in K(\mathcal{L})} \frac{1}{d_k}.
\]  

(12)

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

\[
\begin{array}{cccccc}
 k & f(k) & k & f(k) & k & f(k) \\
 1 & 1 & 13 & 9 & 25 & 17 \\
 2 & 2 & 14 & 10 & 26 & 18 \\
 3 & 2 & 15 & 11 & 27 & 18 \\
 4 & 3 & 16 & 11 & 28 & 19 \\
 5 & 4 & 17 & 12 & 29 & 20 \\
 6 & 5 & 18 & 13 & 30 & 20 \\
 7 & 5 & 19 & 13 & 31 & 21 \\
 8 & 6 & 20 & 14 & 32 & 22 \\
 9 & 7 & 21 & 14 & 33 & 22 \\
 10 & 7 & 22 & 15 & 34 & 23 \\
 11 & 8 & 23 & 16 & 35 & 24 \\
 12 & 8 & 24 & 17 & 36 & 25 \\
\end{array}
\]

Thus,

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

(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 $\mathcal{L}$-free set

\[
R_k = \{2^a3^b : i = 1, \ldots, f(k)\} \subseteq \{d_1, \ldots, d_k\}
\]

in which all \(a_i - b_i\) are congruent modulo 3.

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

5. Forms in One Variable—The General Case

Let $\mathcal{L}$ denote the set of linear forms \{\(a_1x, \ldots, a_nx\)\} 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 $D(\mathcal{L}) = (d_1 < d_2 < \cdots)$ denote the set of all integers of the form
q_i^a \cdots q_i^a, \alpha_i \geq 0. For each \( k \) let \( f(k) \) denote the cardinality of a maximal \( \mathcal{L} \)-free subset of \( \{d_1, \ldots, d_k\} \). Finally, let \( K(\mathcal{L}) \) be defined by

\[
K(\mathcal{L}) = \{k : f(k) > f(k - 1)\}.
\]

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

**Theorem 2**

\[
\delta(\mathcal{L}) = \prod_{j=1}^{r} (1 - q_j^{-1}) \sum_{k \in K(\mathcal{L})} d_k^{-1}
\]  \hspace{1cm} (14)

6. Concluding Remarks

One problem with a representation such as (14) is that it is not clear how to describe \( K(\mathcal{L}) \) so as to be able to evaluate \( \sum_{k \in K(\mathcal{L})} d_k^{-1} \). Several systems \( \mathcal{L} = \mathcal{L}(a_1, \ldots, a_n) = \{a_1 x, \ldots, a_n x\} \) 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 \( K(\mathcal{L}) \) are not difficult and are omitted.

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

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

**REFERENCES**
