On a linear diophantine problem of Frobenius

by

P. ERDÖS (Budapest) and R. L. GRAHAM (Murray Hill, N. J.)

Introduction. Given integers $0 < a_1 < ... < a_n$ with $gcd(a_1, ..., a_n) = 1$, it is well-known that the equation $N = \sum_{k=1}^n x_k a_k$ has a solution in non-negative integers x_k provided N is sufficiently large. Following [9], we let $G(a_1, ..., a_n)$ denote the greatest integer N for which the preceding equation has no such solution.

The problem of determining $G(a_1, \ldots, a_n)$, or at least obtaining nontrivial estimates, was first raised by G. Frobenius (cf. [2]) and has been the subject of numerous papers (e.g., cf. [1], [2], [3], [4], [7], [8], [9], [11], [12], [13]). It is known that:

$$G(a_1, a_2) = (a_1 - 1)(a_2 - 1) - 1 \quad ([2], [11]);$$

$$G(a_1, \dots, a_n) \leq (a_1 - 1)(a_n - 1) - 1 \quad ([2], [4]);$$

$$G(a_1, \dots, a_n) \leq \sum_{k=1}^{n-1} a_{k+1} d_k / d_{k+1}$$

where $d_k = \gcd(a_1, \ldots, a_k)$ ([2]). The exact value of G is also known for the case in which the a_k form an arithmetic progression ([1], [13]).

In this paper, we obtain the bound

$$G(a_1,\ldots,a_n)\leqslant 2a_{n-1}\left[\frac{a_n}{n}\right]-a_n,$$

which in many cases is superior to previous bounds and which will be seen to be within a constant factor of the best possible bound. We also consider several related extremal problems and obtain an exact solution in the case that $a_n - 2n$ is small compared to $n^{1/2}$.

A general bound. As before, we consider integers $0 < a_1 < \ldots < a_n$ with $gcd(a_1, \ldots, a_n) = 1$.

THEOREM 1.

(1)
$$G(a_1,\ldots,a_n) \leqslant 2a_{n-1}\left[\frac{a_n}{n}\right] - a_n.$$

Proof. Let g denote a_n , let m denote $\left[\frac{a_n}{n}\right]$ and let A denote the set $\{0, a_1, \ldots, a_{n-1}\}$ of residues modulo g. Consider the sum

$$\mathscr{C} = \underbrace{A + \ldots + A}_{m} = \{b_1 + \ldots + b_m : b_k \epsilon A\} \pmod{g}.$$

By a strong theorem of Kneser ([10]; cf. also [6], p. 57), there exists a (minimal) divisor g' of g such that

$$\mathscr{C} = \underbrace{A^{(g')} + \ldots + A^{(g')}}_{m} \pmod{g}$$

where

$$A^{(g')} = \{a + rg' \colon 0 \leqslant r < g/g', \ a \in A\} \pmod{g}$$

and such that

(2)
$$\frac{|\mathscr{C}|}{g} \ge \frac{mn}{g} - \frac{m-1}{g'}.$$

Assume \mathscr{C} does not contain a complete system of residues modulo g. Since $gcd(a_1, \ldots, a_{n-1}, g) = 1$ then $A^{(g')}$ must consist of more than one congruence class mod g'. By the theorem of Kneser and the minimality of g', it follows that \mathscr{C} must contain at least m+1 distinct residue classes mod g'; thus

$$(3) \qquad \qquad \frac{|\mathscr{C}|}{g} \geqslant \frac{m+1}{g'}.$$

Note that $g \ge n$ and $m = \lfloor g/n \rfloor$ imply

(4)
$$m+1 > \frac{1}{2} \left(\frac{m-1}{\frac{mn}{g} - \frac{1}{2}} \right).$$

Suppose now that $|\mathscr{C}| \leq \frac{1}{2}g$. By (2) and (4) we have

$$rac{mn}{g} - rac{m-1}{g'} \leqslant rac{1}{2}, \quad `g' \leqslant rac{m-1}{rac{mn}{g} - rac{1}{2}} < 2\,(m+1).$$

Hence, by (3),

$$\frac{|\mathscr{C}|}{g} \! \geqslant \! \frac{m\!+\!1}{g'} \! > \! \frac{m\!+\!1}{2\,(m\!+\!1)} = \! \frac{1}{2}$$

which is a contradiction.

We may therefore assume $|\mathscr{C}| > \frac{1}{2}g$. But in this case it is easily seen that $\mathscr{C} + \mathscr{C}$ contains a complete residue system mod g. It follows that the least possible integer not representable in the form

$$x_1b_1+\ldots+x_{2m}b_{2m}+xg$$

with $x_k \ge 0$, $x \ge 0$, $b_k \epsilon A$, is given by

$$2m \cdot \max_{a \in A} (a) - g = 2a_{n-1} \left[\frac{a_n}{n} \right] - a_n.$$

This proves the theorem.

Note that in the case that n = 2 and a_2 is odd we have

$$G(a_1, a_2) \leqslant 2a_1 \left[\frac{a_2}{2} \right] - a_2 = a_1 a_2 - a_1 - a_2$$

which is best possible.

An extremal problem. The question of the estimation of G naturally suggests the following extremal problem. For integers n and t, define g(n, t) by

$$g(n, t) = \max_{a_i} G(a_1, \ldots, a_n)$$

where the max is taken over all a_i satisfying

(5)
$$0 < a_1 < \ldots < a_n \leq t, \quad \gcd(a_1, \ldots, a_n) = 1.$$

By Theorem 1 the following result is immediate.

COROLLARY. $g(n, t) < 2t^2/n$.

On the other hand, it is not hard to see that for the set $\{x, 2x, ..., (n-1)x, x^*\}$ with x = [t/(n-1)] and $x^* = (n-1)[t/(n-1)]-1$,

$$g(n, t) \ge G(x, \ldots, x^*) \ge \frac{t^2}{n-1} - 5t \quad \text{for} \quad n \ge 2.$$

Thus, g(n, t) is bounded below by essentially t^2/n .

Of course, for n = 2, the exact value of g is given by g(2, t) = (t-1)(t-2)-1. It appears that

$$g(3, t) = \left[\frac{(t-2)^2}{2}\right] - 1,$$

with the sets $\{t/2, t-1, t\}$ or $\{t-2, t-1, t\}$ for t even and $\{(t-1)/2, t-1, t\}$ for t odd achieving this bound. However, this has not yet been established. It follows from the Corollary that $g(n, cn) < 2c^2n$ and $g(n, n^2) < 2n^3$; again, the truth probably differs from these estimates by a factor of 1/2 for large n.

26 - Acta Arithmetica XXI.

Determination of g(n, 2n+k). The remainder of the paper will be concerned with the determination of g(n, 2n+k) for n large compared to k. It follows easily from density considerations that g(n, 2n+k)= 2n+2k-1 for $k \leq -1$ (cf. [12]). It was shown in [5] that g(n, 2n)= 2n+1 and g(n, 2n+1) = 2n+3. It was also proved in [5] that for k fixed g(n, 2n+k) = 2n+h(k) for some function h of k provided n is sufficiently large. The exact value of h(k) is given by the next result.

THEOREM 2. For k fixed, if n is sufficiently large then

$$g(n, k) = \begin{cases} 2n+2k-1 & \text{for} \quad k \leq -1, \\ 2n+1 & \text{for} \quad k = 0, \\ 2n+4k-1 & \text{for} \quad k \geq 1 \text{ and } n-k \equiv 1 \pmod{3}, \\ 2n+4k+1 & \text{for} \quad k \geq 1 \text{ and } n-k \not\equiv 1 \pmod{3}. \end{cases}$$

Proof. By previous remarks we may restrict ourselves to $k \ge 2$. Assume for a fixed integer $K \ge 2$ the theorem holds for all k < K. Let $A = \{a_1, \ldots, a_n\}$ be a set satisfying (5) with k = K and n large (to be specified later). We first establish

(6)
$$g(n, k) \leq \begin{cases} 2n + 4K - 1 & \text{if } n - K \equiv 1 \pmod{3}, \\ 2n + 4K + 1 & \text{if } n - K \not\equiv 1 \pmod{3}. \end{cases}$$

Let S(A) denote the set of sums $\{\sum_{i=0}^{n} x_i a_i : x_i \ge 0\}$ we are considering and let G(A) abbreviate $G(a_1, \ldots, a_n)$. Note that if there exists an x, $1 \le x \le 2n+K$, with $x \in S(A)$, $x \notin A$, then the set $A' = A \cup \{x\}$ satisfies

$$0 < a'_1 < \ldots < a'_{n+1} = 2n + K = 2(n+1) + K - 2.$$

By the induction hypothesis

$$G(A) = G(A') \leq 2(n+1) + 4(K-2) + 1 = 2n + 4K - 5 < 2n + 4K - 1$$

so that (6) certainly holds in this case. Hence, we may assume A and S(A) agree below 2n + K.

Next, suppose $2n+K+1 \in S(A)$. Then for $A' = A \cup \{2n+K+1\}$ we have

$$0 < a'_1 < \ldots < a'_{n+1} = 2n + K + 1 = 2(n+1) + K - 1$$

so that by the induction hypothesis

$$G(A) = G(A') \leq 2(n+1) + 4(K-1) + 1 = 2n + 4K - 1$$

and (6) holds in this case. Hence, we may assume

$$2n+K+1\notin S(A)$$
.

Now, suppose $2n + K + 2\epsilon S(A)$, $2n + K + 3\epsilon S(A)$. For $A' = A \cup \cup \{2n + K + 2, 2n + K + 3\}$ we have

$$0 < a'_1 < \ldots < a'_{n+2} = 2n + K + 3 = 2(n+2) + K - 1.$$

By the induction hypothesis

$$egin{aligned} G(A) &= G(A') \leqslant egin{cases} 2\,(n+2) + 4\,(K-1) - 1 & ext{if} \ (n+2) - (K-1) &\equiv 1 (ext{mod} \ 3), \ 2\,(n+2) + 4\,(K-1) + 1 & ext{if} \ (n+2) - (K-1) &
eq 1 (ext{mod} \ 3) \ &= egin{cases} 2n + 4K - 1 & ext{if} \ n-k &\equiv 1 \ (ext{mod} \ 3), \ 2n + 4K + 1 & ext{if} \ n-k &
eq 1 \ (ext{mod} \ 3), \ \end{aligned}$$

so that (6) holds in this case. Hence we may assume that either

 $2n+K+2\notin S(A)$ or $2n+K+3\notin S(A)$.

There are two cases:

(I) Suppose $a_1 \leq 3K$. If at least 3K consecutive integers belong to A then by successively adding a_1 to these integers, we infer that G(A) < 2n+K and (6) holds in this case. Therefore, we may assume that A does not contain 3K consecutive integers.

Since we have assumed $2n + K + 1 \notin S(A)$ then for all $i, 1 \leq i \leq 2n + K$, either $i \notin A$ or $2n + K + 1 - i \notin A$. Thus, for exactly $\left[\frac{K+1}{2}\right]$ values of jwe have $j \notin A$ and $n + K + 1 - j \notin A$. For a given integer f(K), if n is sufficiently large then for some $t \leq \left[\frac{K+1}{2}\right] f(K)$, each of the integers t+i, $1 \leq i \leq f(K)$, satisfies either

$$t+i\epsilon A$$
 or $2n+K+1-(t+i)\epsilon A$.

Consequently, for some t', $t+1 \leq t' \leq t+3K$, we have

$$2n+K-t'+1\epsilon A.$$

There are several possibilities:

(i) Suppose $2n + K - t' \epsilon A$. If $t' + 2 \epsilon A$ then we would have 2n + K - t' + 2, $2n + K - t' + 3 \epsilon S(A)$ which contradicts our assumptions on A. We may therefore assume

$$2n+K-t'-1\epsilon A$$
.

But now consider t'+3. If $t'+3 \epsilon A$ then as before we find 2n+K-t'+2, $2n+K-t'+3 \epsilon S(A)$ which is a contradiction. Hence, we must have

$$2n+K-t'-2\epsilon A$$
.

We can continue this argument to conclude that

$$2n+K-t'-s\epsilon A$$
 for $0 \leq s \leq 3K-1$,

provided $f(K) \ge 6K$ and *n* is sufficiently large. But this is a sequence of 3K consecutive integers in *A* and since this contradicts our assumption on *A*, then case (i) is impossible.

(ii) Suppose $2n + K - t' \notin A$. Then we have

 $t'+1\epsilon A$.

If we now have $t' + 2 \epsilon A$ then as before 2n + K - t' + 2, $2n + K - t' + 3 \epsilon S(A)$ which is a contradiction. Therefore, we may assume $t' + 2 \epsilon A$, i.e.,

$$2n+K-t'-1\epsilon A$$
.

Now, by using the same arguments as in (i) we can argue that t'+3, $2n+K-t'-3, \ldots, t'+2r+1, 2n+K-t'-2r-1 \epsilon A$ for 2r < f(K)-3K if n is sufficiently large. In particular we have

$$t'+2j+1\epsilon A$$
, $0\leqslant j<rac{1}{2}ig(f(K)-3Kig)$

where $t' \leq \left[\frac{K+1}{2}\right] f(K) + 3K$. Since $a_1 \leq 3K$ then by successively adding $2a_1$ to the integers t' + 2j + 1, we see that all integers x of the form x = t' + 2s + 1, $s \geq 0$, belong to S(A) provided

$$6K \leq f(K) - 3K$$
.

Of course if $t' \equiv 0 \pmod{2}$, then by adding $t' + 1 \epsilon A$ to the integers t' + 2s + 1, $s \ge 0$, we see that all integers $\ge 2 \left[\frac{K+1}{2}\right] f(K) + 6K + 2$ belong to S(A). For *n* sufficiently large, this certainly implies (6). We may therefore assume

$$t' \equiv 1 \pmod{2}$$

and consequently all even integers $\geq t'+1$ belong to S(A). In fact, is it clear that if $x \in A$ is an odd integer and $x \leq 2n + K - (t'+1)$ then all odd integers $\geq 2n + K$ (and hence all integers $\geq 2n + K$) belong to S(A). Thus, we may assume that

$$x \in A$$
, $x \text{ odd } \Rightarrow x > 2n - \left[\frac{K+1}{2}\right] f(K) - 2K$.

Further, if K is odd then 2n+K+1 is even and therefore belongs to S(A) for n sufficiently large. This contradicts our assumption on A and we may assume K is even.

404

Now, let u be the largest integer such that $2n + K - 2u + 1 \epsilon A$. Since K is even it follows that

$$u < \frac{1}{2} \left(\left[\frac{K+1}{2} \right] f(K) + 3K + 1 \right) \cdot$$

Consider the K+1 integers 2u+2j, $1 \le j \le K+1$. By the definition of u none of the integers 2n+K-(2u+2j)+1 belongs to A. Since there are at most $\left[\frac{K+1}{2}\right] = \frac{K}{2}$ of these integers for which both $2u+2j \notin A$ and $2n+K-(2u+2j)+1 \notin A$ then we see that at least $K+1-\frac{K}{2}=\frac{K}{2}+1$ of them belong to A, say,

$$2u+2j_1,\ldots,2u+2j_t\epsilon A, \quad t \ge K/2+1.$$

Forming the sums

$$(2n+K-2u+1)+(2u+2j_i), \quad i=1,2,...,t,$$

we obtain at least K/2+1 sums $2n+K+2j_i+1$ which are $\ge 2n+K+3$ and $\le 2n+3K+3$ and which belong to S(A). But all the even integers 2n+K+2r, $1 \le r \le K+1$, also belong to S(A). Hence, S(A) contains at least n+(K/2+1)+K+1 integers which are less than or equal to 2n+3K+3 and we can find a subset $A' \subseteq S(A)$ with

$$0 < a'_1 < \ldots < a'_{n+3K/2+2} = 2n+3K+3-d$$

for some integer $d \ge 0$. Since

$$(2n+3K+3-d) - (2+3K/2+2) \leq -1$$

then by the induction hypothesis we conclude that all integers $\geq 2n + 3K + 3 - d$ belong to S(A). If $d \geq 1$ then in fact all integers $\geq 2n + 3K + 2$ belong to S(A); if d = 0 then since 2n + 3K + 2 is even then we still have all integers $\geq 2n + 3K + 2 \epsilon S(A)$. Thus,

$$G(A) \leq 2n+3K+1$$
.

But for $K \ge 2$, $4K - 1 \ge 3K + 1$ so that

$$G(A) \leq 2n+4K-1$$

and (6) holds in this case. This concludes case (I).

- (II) Suppose $a_1 > 3K$. There are two cases:
- (i) Suppose $a_1 > n + \left[\frac{K+1}{2}\right]$. Thus, exactly $\left[\frac{K+1}{2}\right]$ of the inte-

gers which are $> n + \left[rac{K+1}{2}
ight]$ and < 2n + K are missing from A. This

implies that for some $i, 1 \leq i \leq \left[\frac{K+1}{2}\right] + 1$, both $n+2\left[\frac{K+1}{2}\right] + 1 + i \epsilon A$ and $n+2\left[\frac{K+1}{2}\right] + 2 - i \epsilon A$, i.e., $2n+4\left[\frac{K+1}{2}\right] + 3 \epsilon S(A)$. Of course, the same argument can be repeated for $2n+4\left[\frac{K+1}{2}\right] + 4$, etc., so that for *n* sufficiently large, $2n+4\left[\frac{K+1}{2}\right] + j + 2 \epsilon S(A)$ for $1 \leq j \leq 4\left[\frac{K+1}{2}\right] + 4$. Hence S(A) contains a subset A' with

$$0 < a'_1 < \ldots < a'_{n+4\left[\frac{K+1}{2}\right]+3} = 2n + 8\left\lfloor\frac{K+1}{2}\right\rfloor + 5 - d$$

for some $d \ge 0$. Since

$$2\left(n+4\left\lceil\frac{K+1}{2}\right\rceil+3\right) > 2n+8\left\lceil\frac{K+1}{2}\right\rceil+5-d$$

then by the induction hypothesis all integers $> 2n+8\left[\frac{K+1}{2}\right]+5$ belong to S(A). But since $2n+4\left[\frac{K+1}{2}\right]+j+2\epsilon S(A)$ for $1 \le j \le 4\left[\frac{K+1}{2}\right]+3$ then all integers $> 2n+4\left[\frac{K+1}{2}\right]+2$ belong to S(A). However, $4\left[\frac{K+1}{2}\right]+2 \le 4K-1$ for $K \ge 2$ so that (6) holds in this case.

(ii) Suppose $a_1 \leq n + \left[\frac{K+1}{2}\right]$. Consider the 3K-1 integers $2n + K - a_1 + i + 1$, $1 \leq i \leq 3K - 1$. Since a_1 is the least element of A then at least $3K - 1 - \left[\frac{K+1}{2}\right]$ of these integers must belong to A. Adding a_1 to each of them gives at least $3K - 1 - \left[\frac{K+1}{2}\right]$ integers in S(A) which are > 2n + K and $\leq 2n + 4K$. Thus, S(A) contains a subset A' with

$$0 < a'_1 < \ldots < a'_{n+3K-1-\left[rac{K+1}{2}
ight]} = 2n+4K-a$$

for some $d \ge 0$.

For $K \ge 4$,

$$2 \Big(n + 3K - 1 - \left[\frac{K+1}{2} \right] \Big) > 2n + 4K - d$$

so that by the induction hypothesis

$$G(A) \leqslant G(A') \leqslant 2n + 4K - 1$$

and (6) holds. Hence, we may assume $K \leq 3$. There are two cases.

Suppose K = 2. If $2n - a_1 + j \epsilon A$, $4 \leq j \leq 6$, then $2n + j \epsilon S(A)$, $4 \leq j \leq 6$. Thus S(A) contains a subset A' with

$$0 < a'_1 < \ldots < a'_{n+3} = 2n+6$$

and by the induction hypothesis

$$G(A) \leqslant G(A') \leqslant 2n+7$$

so that (6) holds in this case.

If at least one of $2n - a_1 + j$, $4 \leq j \leq 6$, is missing from A, then in fact, exactly one of $2n - a_1 + j$, $4 \leq j \leq 6$, is missing from A, and all of $2n - a_1 + j \epsilon A$, $1 \leq j \leq 9$. Hence, $2n + j \epsilon S(A)$, $7 \leq j \leq 9$, and S(A) contains a subset A' with

$$0 < a'_1 < \ldots < a'_{n+5} \leqslant 2n+9$$
.

By the induction hypothesis

$$G(A') \leqslant 2n+8$$

and since 2n+7, $2n+8 \in S(A)$ then

$$G(A) \leqslant 2n+6$$

which satisfies (6) in this case.

The case K = 3 is similar and will be omitted. It can be checked that the condition that n be sufficiently large in the preceding arguments is satisfied, for example, by taking $n > 20K^2$.

This concludes case (II) and (6) is proved.

We next exhibit specific sets A which satisfy (6) with equality for n arbitrarily large. There are three cases.

(i) $n-K \equiv 1 \pmod{3}$. Write n = 3m+K+1 and let

$$A \ = \bigcup_{i=1}^{2m+K} \left\{ 3i
ight\} \cup \bigcup_{j=1}^{m+1} \left\{ 3m+3K+5-3j
ight\}.$$

The least element of S(A) which is $\equiv 1 \pmod{3}$ is 2(3m+3K+2) = 6m+6K+4 so that

$$2n+4K-1 = 6m+6K+1 \notin S(A).$$

Therefore $0 < a_1 < \ldots < a_n = 2n + K$ and $G(A) \ge 2n + 4K - 1$. (ii) $n - K \equiv 2 \pmod{3}$. Write n = 3m + K + 2 and let

$$A = \bigcup_{i=1}^{2m+K+1} \{3i\} \cup \bigcup_{j=1}^{m+1} \{3m+3K+7-3j\}.$$

(iii) $n-K \equiv 0 \pmod{3}$. Write n = 3m+K and let

$$A = \bigcup_{i=1}^{2m+K} \{3i\} \cup \bigcup_{j=1}^{m} \{6m+3K+2-3j\}.$$

It is easy to see in (ii) and (iii) that A satisfies (5) and $G(A) \ge 2n + 4K + 1$.

The examples in (i), (ii) and (iii) together with (6) establish the theorem for k = K. This completes the induction step and the theorem is proved.

Acknowledgment. The authors wish to thank E. G. Straus for important suggestions in the proof of Theorem 1.

Added in proof: The conjecture $g(3, t) = \left[\frac{(t-2)^2}{2}\right] - 1$ has recently been settled in the affirmative by M. Lewin (personal communication).

References

- P. T. Bateman, Remark on a recent note on linear forms, Amer. Math. Monthly 65 (1958), pp. 517-518.
- [2] Alfred Brauer, On a problem of partitions, Amer. J. Math. 64 (1942), pp. 299-312.
- [3] and B. M. Seelbinder, On a problem of partitions, II, Amer, J. Math. 76 (1954), pp. 343-346.
- [4] and J. E. Shockley, On a problem of Frobenius, J. Reine Angew. Math. 211 (1962), pp. 215-220.
- [5] P. Erdös, Problem P-84, Can. Math. Bull. 14 (1971), pp. 275-277.
- [6] H. Halberstam and K. F. Roth, Sequences I, London 1966.
- [7] B. R. Heap and M. S. Lynn, A graph theoretic algorithm for the solution of a linear diophantine equation, Numerische Math. 6 (1964), pp. 346-354.
- [8] - On a linear diophantine problem of Frobenius: an improved algorithm, Numerische Math. 7 (1965), pp. 226-231.
- [9] S. M. Johnson, A linear diophantine problem, Can. J. Math. 12 (1960), pp. 390-398.
- [10] M. Kneser, Abschätzungen der asymptotischen Dichte von Summenmengen, Math. Zeitschr. 58 (1953), pp. 459-484.
- [11] N. S. Mendelsohn, A linear diophantine equation with applications to nonnegative matrices, Ann. N. Y. Acad. Sci. 175 art. 1 (1970), pp. 287-294.
- [12] M. Nagata and H. Matsumura, Súgaku 13 (1961-62), p. 161; Math. Rev. 25 no. 3 # 2386 (1963).

HUNGARIAN ACADEMY OF SCIENCES Budapest, Hungary BELL TELEPHONE LABORATORIES, INCORPORATED Murray Hill, New Jersey

Received on 25. 8. 1971

(220)