[12] W. Staś, *Über einige Abschätzungen in Idealklassen*, ibid. 6 (1960), 1–10.
[13] P. Turán, *Über eine neue Methode in der Analysis und deren Anwendungen*, Akadémiai Kiadó, Budapest 1953.

INSTITUTE OF MATHEMATICS OF THE ADAM MICKIEWICZ UNIVERSITY
Poznań, Poland

Received on 17.3.1987
and in revised form on 25.1.1988 (1714)

ACTA ARITHMETICA
LII (1989)

On representation of r-th powers by subset sums

by

E. LIPKIN* (Tel-Aviv)

Let A be a set of x natural numbers

(1) $$A = \{a_1, \ldots, a_x\}, \quad 1 \leqslant a_1 < a_2 < \ldots < a_x \leqslant l, \quad |A| = x.$$

Let $\mathcal{M}$ be a given set of integers. Denote by $f(l, \mathcal{M})$ the maximum cardinality of a set A which contains no subset $B \subseteq A$ such that $\sum_{a_i \in B} a_i \in \mathcal{M}$.

Recently Erdős and Freud, and N. Alon proposed the following four similar problems:

1. Let $a_x \leqslant 3(x-1)$. Does there exist a subset $B \subseteq A$ such that $\sum_{a_i \in B} a_i$ is a power of two? ([Er].)

2. Let $a_x \leqslant 4(x-1)$. Does there exist a subset $B \subseteq A$ such that $\sum_{a_i \in B} a_i$ is a square-free number? ([Er].)

3. What is a maximal cardinality of set A which contains no subset $B \subseteq A$ such that $\sum_{a_i \in B} a_i$ is a square? In other words what is $f(l, \mathcal{M})$ if $\mathcal{M} = M_2$ is the set of all squares? ([Er].)

4. Let $f(l, m)$ denote for $m \geqslant 1$ the maximum cardinality of a set $A \subseteq \{1, \ldots, l\}$ which contains no subset $B \subseteq A$ such that $\sum_{a_i \in B} a_i = m$.

Conjecture of N. Alon is that if $l^{1.1} \leqslant m \leqslant l^{1.9}$, then

$$f(l, m) = (1 + o(1)) \frac{l}{\bar{m}} \quad \text{as } l \to \infty;$$

$\bar{m}$ denotes the smallest integer that does not divide m. ([Al])

G. Freiman stated a natural generalization of problem 3 of P. Erdős:

3′. What is $f(l, \mathcal{M})$ in the case when $\mathcal{M} = M_r$ is the set of all rth powers?

Problems 1 and 2 are considered in [Al] and [EF]. In [Al] it is shown

* Research supported in part by the Fund for Basic Research administered by the Israel Academy of Sciences.

that $f(l, \mathcal{M}) = (\frac{1}{3}+o(1))l$ if $\mathcal{M}$ is the set of all powers of two and $f(l, \mathcal{M}) = (\frac{1}{4}+o(1))l$ if $\mathcal{M}$ is the set of all square-free numbers. [EF] gives a positive answer for both questions 1 and 2 by analytical method.

In this paper we use the methods of [EF] to study problems 3, 3′ and 4.

Concerning these problems the following is known:

P. Erdős ([Er]) found a lower bound for $f(l, M_2)$,

$$f(l, M_2) \geqslant (1+o(1)) \cdot 2^{1/3} \cdot l^{1/3};$$

N. Alon ([Al]) proved that

$$f(l, M_2) = O(l/\log l).$$

G. Freiman conjectured a general asymptotic formula

$$f(l, M_r) = 2^{1/(r+1)} l^{(r-1)/(r+1)} (1+o(1))$$

for $r \geqslant 2$ and suggested that it can be derived by methods of [EF]. The lower bound $f(l, M_r) \geqslant 2^{1/(r+1)} l^{(r-1)/(r+1)} (1+o(1))$ follows using arguments from [Er]. For large r, A is more dense, hence it is simpler to use analytical method.

N. Alon in [Al] proved that for every fixed $\varepsilon > 0$, there exists a constant $c = c(\varepsilon) > 1$ such that for every $l > 0$ and every m, which satisfies $l^{1+\varepsilon} \leqslant m \leqslant l^2/\log l$, the inequality

$$\left\lceil \frac{l}{\bar{m}} \right\rceil \leqslant f(l, m) < c\frac{l}{\bar{m}}$$

holds.

In our paper we prove the following three theorems concerning problems 3, 3′ and 4.

THEOREM 1. *Let ε be an arbitrarily small positive number. Then*

$$f(l, M_2) = O(l^{4/5+\varepsilon}). \tag{2}$$

THEOREM 2. *For $r \geqslant 10$*

$$f(l, M_r) = 2^{1/(r+1)} l^{(r-1)/(r+1)} \left(1+O\left(\frac{1}{l^{\varrho}}\right)\right) \tag{3}$$

where ϱ is an arbitrary positive number less than $1/(6(r+1))$.

THEOREM 3. *If*

$$Cl(\log l)^6 < m < l^{3/2}/(\log l)^3 \tag{4}$$

then

$$f(l, m) = l/\bar{m} + h_1 \tag{5}$$

where $h_1 = c\frac{l \log \bar{m}}{\bar{m} \log^2 l}$, C and c are some constants.

In order to prove Theorems 1, 2 and 3 we first will establish several results about additive properties of set A (Theorems 4, 5, 6) using analytical method of [EF]; see also [F1], [F2], [FJM].

We use the following notation.

For each set $A \subset N$ and $s, q \in N$, $q \geqslant 2$ let $A(s, q) = \{a \mid a \in A, a \equiv s \pmod q\}$.

Let $\lceil a \rceil$ denote the smallest integer $\geqslant a$.

$C_1, C_2, \ldots$ denote positive constants.

$I = I(N)$ denotes the number of solutions of the equation

$$x_1 + x_2 + \ldots + x_n = N, \tag{6}$$

where $x_i \in A$. $Q = Q(N)$ denotes the number of solutions of equation (6), such that all x_i are different, i.e. $x_i \neq x_j$ for $i \neq j$. Denote

$$M = \frac{a_1 + \ldots + a_x}{x}, \tag{7}$$

$$D = \frac{1}{x} \sum_{i=1}^{x} a_i^2 - M^2. \tag{8}$$

THEOREM 4. *Let $A \subset \{1, 2, \ldots, l\}$ be a set (1), $|A| = x$. Suppose $x > l^{4/5+\varepsilon}$, where ε is an arbitrarily small positive number and $l > l_0(\varepsilon)$, and suppose that*

$$|A(s, q)| < x - h \tag{9}$$

for all $s, q \in N$, $q \geqslant 2$, where

$$h = x/\log^2 l. \tag{10}$$

Let n and N in (6) satisfy

$$C_1 \left(\frac{l}{x}\right)^2 (\log l)^4 < n < C_2 \frac{\sqrt{x}}{\log x} \tag{11}$$

(*it is possible because of the assumption $x > l^{4/5+\varepsilon}$*) *and*

$$Mn - C_3\sqrt{nD} < N < Mn + C_4\sqrt{nD} \tag{12}$$

where C_1, C_2, C_3, C_4 are any fixed numbers. Then

$$I = \frac{x^n}{\sqrt{2\pi n D}} e^{-(Mn-N)^2/2nD} + o\left(\frac{x^n}{\sqrt{nD}}\right).$$

Proof. It is known that the number of solutions of equation (6) $x_1 + \ldots + x_n = N$, $x_i \in A$ is

$$I = I(N) = x^n \int_0^1 \varphi^n(\alpha) e^{-2\pi i \alpha N} d\alpha$$

where

$$\varphi(\alpha) = \frac{1}{x}\sum_{a\in A} e^{2\pi i\alpha a}.$$

Define the number

(13) $$L = C_5 l$$

where C_5 is sufficiently large. Since the subintegral function has period 1,

$$I(N) = x^n \int_{-1/L}^{1-1/L} \varphi^n(\alpha)\, e^{-2\pi i\alpha N}\, d\alpha.$$

Divide the interval $[-1/L, 1-1/L]$ into two parts $[-1/L, 1/L]$ and $[1/L, 1-1/L]$. Correspondingly, $I(N)$ equals the sum of the two integrals I_1 and I_2. To prove the assertion of Theorem 4 it is sufficient to prove that

(14) $$I_1 = \int_{-1/L}^{1/L} \varphi^n(\alpha)\, e^{-2\pi i\alpha N}\, d\alpha = \frac{1}{\sqrt{2\pi nD}} e^{-(Mn-N)^2/2nD}(1+o(1))$$

and that

(15) $$I_2 = \int_{1/L}^{1-1/L} \varphi^n(\alpha)\, e^{-2\pi i\alpha N}\, d\alpha = o(1/\sqrt{nD})$$

for all N, n which satisfy (11) and (12).

We first show (15). Let us estimate $\varphi(\alpha)$ for $\alpha\in[1/L, 1-1/L]$. Each number $\alpha\in[0, 1]$ has a representation $\alpha = p/q+z$, $(p, q) = 1$, $1\leqslant q\leqslant L$, $|z| < 1/(qL)$; for $\alpha\in[1/L, 1-1/L]$ we have $q\geqslant 2$. Then we can represent $\varphi(\alpha)$ in the form

(16) $$\varphi(\alpha) = \frac{1}{x}\sum_{a\in A} e^{2\pi i(pa/q+za)} = \frac{1}{x}\sum_{k=0}^{q-1} \sum_{\substack{a\in A\\ pa\equiv k \ (\mathrm{mod}\ q)}} e^{2\pi i(k/q+za)}$$

where

(17) $$|za| < \frac{1}{qL}\cdot l < \frac{1}{4q}.$$

Denote by m_k the number of solutions of a congruence $pa_j\equiv k \pmod q$ for $0\leqslant k<q$ and $1\leqslant j\leqslant x$. Consider three different cases according to the value of q, for a sufficiently large l. We will use the inequality

(18) $$\frac{1}{y}\frac{\sin yu}{\sin u} < \frac{1}{y}\,\frac{yu-\frac{1}{2}\frac{(yu)^3}{6}}{u-u^3/6} < 1-\frac{1}{4}\frac{(yu)^2}{6}$$

which holds for $0<yu<\pi/2$ with $y\geqslant 2$.

1. Case $q\geqslant l$. In this case $m_k\leqslant 1$. Then we estimate

(19) $$|\varphi(\alpha)| \leqslant \frac{1}{x}\Big|\sum_{k=0}^{x-1} e^{2\pi ik/2q}\Big| = \frac{1}{x}\frac{\sin(\pi x/2q)}{\sin(\pi/2q)} < 1-\frac{1}{4\cdot 6}\left(\frac{\pi x}{2q}\right)^2$$

and by (19), using $q<L$ and $1/q > 1/(C_5 l)$, we have

(19′) $$|\varphi(\alpha)| < 1-\frac{1}{4\cdot 6}\frac{\pi^2}{4}\frac{1}{C_5^2}\left(\frac{x}{l}\right)^2.$$

2. Case $1<q<8\frac{l}{x}$. By (9) $m_k<x-h$ holds for every k, therefore in the sum (16) we can replace $(x-h)$ terms by 1, h terms by $e^{2\pi i/2q}$ and estimate using (17) and (10)

(20) $$\begin{aligned}|\varphi(\alpha)| &\leqslant \frac{1}{x}|x-h+he^{2\pi i/2q}|\\ &= \left|1-2\frac{h}{x}+\frac{h}{x}(1+e^{2\pi i/2q})\right| \leqslant 1-2\frac{h}{x}+\frac{h}{x}|1+e^{2\pi i/2q}|\\ &= 1-4\frac{h}{x}\sin^2\frac{\pi}{4q} = 1-\frac{4}{\log^2 l}\sin^2\frac{\pi}{4q} < 1-\frac{1}{\log^2 l}\frac{1}{64}\left(\frac{x}{l}\right)^2\end{aligned}$$

by $\sin u > \frac{2}{\pi}u$ and $\sin^2\frac{\pi}{4q} > \frac{1}{4q^2} > \frac{1}{4\cdot 64}\left(\frac{x}{l}\right)^2$.

3. Case $8\frac{l}{x}\leqslant q<l$. In this case $m_k\leqslant\lceil l/q\rceil < 2l/q$ for all k. Define $m = \lceil 2l/q\rceil$ and $r = \lceil x/(4l/q)\rceil = \lceil xq/(4l)\rceil$. Then $m\geqslant 2l/q$, $r\geqslant xq/(4l)$ and $mr\geqslant x/2$. Denote $t = x-mr$, then $t\leqslant x/2$. Replace in the sum (16) t terms by 1, m terms by $e^{2\pi ik/2q}$ for each $k = 0, 1, \ldots, r-1$ and estimate using (17), and (18) since $r\geqslant 2$

(21) $$\begin{aligned}|\varphi(\alpha)| &\leqslant \frac{t}{x}+\frac{1}{x}\Big|m\sum_{k=0}^{r-1} e^{2\pi ik/2q}\Big|\\ &= \frac{t}{x}+\frac{m\sin(\pi r/2q)}{x\sin(\pi/2q)} = \frac{t}{x}+\frac{mr}{x}\cdot\frac{1}{r}\frac{\sin(\pi r/2q)}{\sin(\pi/2q)}\\ &< \frac{x-mr}{x}+\frac{mr}{x}\left(1-\frac{1}{4\cdot 6}\left(\frac{\pi r}{2q}\right)^2\right)\\ &= 1-\frac{mr}{x}\cdot\frac{\pi^2}{4\cdot 6\cdot 4}\left(\frac{r}{q}\right)^2 < 1-\frac{\pi^2}{2\cdot 4\cdot 6\cdot 4\cdot 4^2}\left(\frac{x}{l}\right)^2\end{aligned}$$

in view of $mr/x\geqslant 1/2$ and $r/q\geqslant x/4l$.

From these three cases we conclude by (19′), (20), (21) that for all α, $1/L$

$< \alpha < 1 - 1/L$

$$|\varphi(\alpha)| < 1 - c_0 \frac{1}{\log^2 l}\left(\frac{x}{l}\right)^2$$

holds with an appropriate constant c_0 for a sufficiently large l. Then by the left side of (11) the estimation

$$(22)\quad |\varphi(\alpha)|^n < \left(1 - c_0 \frac{1}{\log^2 l}\left(\frac{x}{l}\right)^2\right)^n \ll \left(1 - c_0 \frac{1}{\log^2 l}\left(\frac{x}{l}\right)^2\right)^{C_1 (l/x)^2 (\log l)^4} \ll \frac{1}{l^2}$$

follows. By (7) and (8) we observe that $D < cl^2$ where c is some constant, so by (11), $nD < cl^2 \sqrt{l}$. Thus, (22) implies in (15) that

$$\int_{1/L}^{1-1/L} \varphi^n(\alpha) e^{-2\pi i \alpha N} d\alpha = O(1/l^2) = o(1/\sqrt{nD})$$

and (15) follows.

Next we estimate integral $I_1 = \int_{-1/L}^{1/L} \varphi^n(\alpha) e^{-2\pi i \alpha N} d\alpha$ to prove (14). By (7), (8) $D > Cx^2$ with some constant C and by (11) $nD > Cl^2 (\log l)^4$, hence for $b = \sqrt{(\log l)/nD}$, $b < 1/L$ holds. Divide the interval $[-1/L, 1/L]$ into three parts $[-1/L, -b]$, $[-b, b]$, $[b, 1/L]$. Correspondingly $I_1 = \int_{-1/L}^{1/L}$ equals the sum of the three integrals. For all $\alpha \in [-1/L, 1/L]$,

$$|\alpha a| < \frac{1}{C_5 l} \cdot l = \frac{1}{C_5}$$

holds in view of (13). By the Taylor expansion formula $e^{2\pi i \alpha a} = 1 + 2\pi i \alpha a - 2\pi^2 \alpha^2 a^2 + o(\alpha^2 a^2)$, then we have

$$(23)\quad \varphi(\alpha) = \frac{1}{x}\sum_{a \in A} e^{2\pi i \alpha a} = 1 + 2\pi i \alpha M - 2\pi^2 \alpha^2 (D + M^2) + o(\alpha^2 (D + M^2))$$

$$= e^{2\pi i \alpha M - 2\pi^2 \alpha^2 D + o(\alpha^2 D)}.$$

Because of (23) for $1/L > |\alpha| \geq b = \sqrt{(\log l)/nD}$ and for sufficiently large l

$$(24)\quad |\varphi^n(\alpha) e^{-2\pi i \alpha N}| < e^{-\pi^2 \alpha^2 nD} < e^{-\pi^2 \log l} = 1/(l^{\pi^2}) < 1/l^2$$

holds and we conclude that $\int_{-1/L}^{-b} + \int_{b}^{1/L} = o(1/\sqrt{nD})$. For the principal part of I_1 one can obtain the estimation (14) in the usual way.

This completes the proof of Theorem 4. ∎

THEOREM 5. *Let us assume that all the conditions of Theorem 4 are satisfied. Then each number $N \in \mathbf{N}$ in interval (12) can be represented as a subset sum of A, $N = \sum_{a_i \in B} a_i$ where $B \subseteq A$.*

Proof. Recall that $Q = Q(N)$ denotes the number of solutions of equation (6) such that all x_i are different, i.e. $x_i \neq x_j$ for $i \neq j$. Let us show that

$$(25)\quad Q = I + o(x^n/\sqrt{nD}).$$

If at least two unknowns in the solution of equation (6) are equal to a_i, denote the number of such solutions by Q_i. There are $n(n-1)/2$ ways to choose a pair of unknowns.

The number of solutions of the equation $y_1 + \ldots + y_{n-2} = N - 2a_i$ where $y_i \in A$, is $O(x^{n-2}/\sqrt{nD})$ according to Theorem 4. Thus $Q_i = O\left(n^2 \frac{x^{n-2}}{\sqrt{nD}}\right)$. Notice that $N - 2a_i$ belongs to the interval (12) if we take the number C_3 to be sufficiently large. By (11), $\sum_{i=1}^{x} Q_i = O(x^n/((\log x)^2 \sqrt{nD}))$ which produces (25). This implies the assertion of the theorem. ∎

The set A in (1) does not necessarily satisfy condition (9). Let us show that for a large subset B of A the condition of type (9) holds.

LEMMA. *Let A be the set (1), $x > l^\alpha$ for some $\alpha > 0$ and l be sufficiently large; $h = x/\log^2 l$. Then there exists $B \subseteq A$ such that*

(i) $|B| \geq |A| - (\log_2 (l/x) + 1) h$,

(ii) *B is contained in an arithmetic progression, i.e. for some $\bar{s}$ and $\bar{q} \in \mathbf{N}$, $b_j \equiv \bar{s} \pmod{\bar{q}}$ holds for each $b_j \in B$,*

(iii) *$|B(s, q)| < |B| - h$ for all s and $q > \bar{q}$, $\bar{q} | q$.*

Proof. If condition (iii) for $B = A$ and $\bar{q} = 1$ holds the proof is over. Otherwise there exist some $q_0 \geq 2$ and some integer s_0 such that for $A_1 = A(s_0, q_0)$ we have $|A_1| \geq A - h$. If condition (iii) for A_1 and $\bar{q} = q_0$ holds, we put $B = A_1$, and if not, we can find $q_1 \geq 2q_0$ and s_1 such that for $A_2 = A_1(s_1, q_1)$, it is $|A_2| \geq |A_1| - h \geq |A| - 2h$. Suppose that we arrived at $A_k = A_{k-1}(s_{k-1}, q_{k-1})$ where

$$(26)\quad k = \lceil \log_2 (l/x) + 1 \rceil.$$

Let us show that for A_k condition (iii) holds. Suppose that on the contrary, we can find s_k and $q_k \geq 2q_{k-1} \geq 2^{k+1}$ such that $|A_{k+1}| = |A_k(s_k, q_k)| > |A_k| - h$. By (26) we have $2^k \geq 2\frac{l}{x}$ hence

$$(27)\quad |A_{k+1}| > |A| - (k+1) h > x/2 \geq l/2^k.$$

On the other hand, $A_{k+1} = A_k(s_k, q_k)$ is contained in an arithmetic progression, so we have $|A_{k+1}| \leqslant l/q_k \leqslant l/2^{k+1}$ which contradicts (27).

To complete the proof of the Lemma, we put $B = A_k$ and $\bar{s} = s_{k-1}$, $\bar{q} = q_{k-1}$. ■

As a corollary of Theorem 5 and the Lemma we obtain our central auxiliary result.

THEOREM 6. *Assume that set A in* (1) *satisfies the condition $x > l^{4/5+\varepsilon}$ with arbitrary small positive ε. Let $B = A(\bar{s}, \bar{q})$ be the set which we find applying the Lemma. Denote by M', D' corresponding values* (7) *and* (8) *for set B. Denote $d = (\bar{s}, \bar{q})$. Then for $l > l_0(\varepsilon)$ each natural number N, $N \equiv 0 (\mathrm{mod}\, d)$ satisfying*

$$(28)\qquad C_6 M' \left(\frac{l}{x}\right)^2 (\log l)^4 < N < C_7 M' \frac{\sqrt{x}}{\log x}$$

with some constants C_6, C_7 can be represented as a subset sum of B, $N = \sum_{a_i \in G} a_i$ where $G \subseteq B$.

Proof. We will prove the assertion of the theorem for all N satisfying (28) belonging to some class $m (\mathrm{mod}\, \bar{q})$, $d \mid m$. Since m is arbitrary, this does not restrict generality. Let n_0 be a solution of the congruence $n_0 \bar{s} \equiv m (\mathrm{mod}\, \bar{q})$.

We have $B = \{b_j, b_j = \bar{s} + t_j \bar{q}\}$, $j = 1, \ldots, y$. Define $T = \{t_1, \ldots, t_y\}$ where $t_j = (b_j - \bar{s})/\bar{q}$. The numbers t_j satisfy the inequality $t_j < b_j/\bar{q} \leqslant l/\bar{q}$ and $y > l^{4/5+\varepsilon_1} > \left(\frac{1}{\bar{q}}\right)^{4/5+\varepsilon_1}$ where $0 < \varepsilon_1 \leqslant \varepsilon$. From (iii) which is valid for $B = A(\bar{s}, \bar{q})$ it follows that condition (9) is valid for T. Therefore we can apply Theorem 5 to the set T: denote by M'', D'' the corresponding values (7) and (8) for T; let n satisfy the conditions $n \equiv n_0 (\mathrm{mod}\, \bar{q})$ and

$$(11')\qquad C_1 \left(\frac{l}{\bar{q}y}\right)^2 \left(\log \frac{l}{\bar{q}}\right)^4 < n < C_2 \frac{\sqrt{y}}{\log y};$$

then each natural $\tilde{N}$ in the interval

$$(12')\qquad M'' n - C_3 \sqrt{nD''} < \tilde{N} < M'' n + C_4 \sqrt{nD''}$$

can be represented as a subset sum of T, i.e. $\tilde{N} = t_{j_1} + \ldots + t_{j_n}$, $t_j \in T$.

Let us come back to B. From $(b_{j_1} - \bar{s})/\bar{q} + \ldots + (b_{j_n} - \bar{s})/\bar{q} = \tilde{N}$ follows

$$b_{j_1} + \ldots + b_{j_n} = \bar{q}\tilde{N} + n\bar{s}.$$

We deduce by using (12′) that each element N of the form $N = \bar{q}\tilde{N} + n\bar{s}$ and from the interval

$$(29)\qquad M'' \bar{q}n - C_3 \bar{q} \sqrt{nD''} + \bar{s}n < N < M'' \bar{q}n + C_4 \bar{q} \sqrt{nD''} + \bar{s}n$$

where $n \equiv n_0 (\mathrm{mod}\, \bar{q})$, n belonging to (11′), can be represented as a subset sum of B.

Now we will show that sequence of intervals (29) covers interval (28) when n runs over interval (11′) and $n \equiv n_0 (\mathrm{mod}\, \bar{q})$. First we take two consecutive n from interval (11): n and $n + \bar{q}$. Interval (29) for $n + \bar{q}$ looks like

$$(29')\qquad M'' \bar{q}(n+\bar{q}) - C_3 \bar{q} \sqrt{(n+\bar{q})D''} + \bar{s}(n+\bar{q}) < N < M'' \bar{q}(n+\bar{q}) - C_4 \bar{q} \sqrt{(n+\bar{q})D''} + \bar{s}(n+\bar{q}).$$

Let us show that two neighboring intervals (29) and (29′) intersect. It is sufficient to check that

$$M'' \bar{q}(n+\bar{q}) - C_3 \bar{q} \sqrt{(n+\bar{q})D''} + \bar{s}(n+\bar{q}) < M'' \bar{q}n + C_4 \bar{q} \sqrt{nD''} + \bar{s}n$$

or

$$(30)\qquad M''^2 \bar{q}^2 < C_{10} nD''$$

for every positive constant C_{10}. Since $M''^2 \bar{q}^2 \leqslant l^2$, $D'' \gg x^2$ and $n \gg (l/x)^2 \log^4 l$, (30) is satisfied. Secondly we observe that the union of intervals (29) covers interval (28) when n runs over (11), provided constant C_6 is sufficiently large relative to C_8, and C_7 is sufficiently small relative to C_9. Also we use that $\bar{q}M'' < M' < C_{11} \bar{q} M''$ where C_{11} is a constant. We showed that all N from the interval (28), satisfying the condition $N \equiv n_0 \bar{s} (\mathrm{mod}\, \bar{q})$, can be represented as subset sums of A. This completes the proof. ■

Now we can prove the main Theorems 1, 2, 3.

THEOREM 1. *Let A be a set* (1), *$|A| = x$, satisfying $x > l^{4/5+\varepsilon}$ where ε is an arbitrarily small positive number. Then for $l > l_0(\varepsilon)$, there exists a square equal to a subset sum of A. In other words $f(l, M_2) = O(l^{4/5+\varepsilon})$.*

Proof. By Theorem 6, all numbers N in interval (28) and of the form $N = t \cdot d$, $t \in \mathbf{N}$ are subset sums of A. Consider $t = s \cdot d$, $s \in \mathbf{N}$. Then

$$(31)\qquad \frac{1}{d^2} C_6 M' \left(\frac{l}{x}\right)^2 (\log l)^4 < s < C_7 M' \frac{\sqrt{x}}{\log x} \frac{1}{d^2}.$$

The left end of this interval is greater than 1, since $d \leqslant \bar{q} < l/x$. The ratio of the upper bound to the lower bound in (31)

$$C_7 \frac{\sqrt{x}}{\log x} \Big/ C_6 \left(\frac{l}{x}\right)^2 (\log l)^4 > \frac{C}{\log x (\log l)^4} l^{5\varepsilon/2}$$

is greater than two for a sufficiently large l. The segment $[s, 2s]$ contains a square, as does the interval (31). Multiplying it by d^2 we obtain a square contained in (28), represented by a subset sum of A. ■

THEOREM 2. *Let M_r be the set of all r-th powers. For $r \geqslant 10$ and ϱ being an arbitrary positive number less than $1/6(r+1)$ we have the following asymptotic formula*:

$$f(l, M_r) = 2^{1/(r+1)} l^{(r-1)/(r+1)}(1+O(1/l^{\varrho})). \tag{3}$$

Proof. The lower bound is given for $r = 2$ by Erdös ([Er]). In the same way for $r \geqslant 2$ we construct the A whose subset sum is never an rth power. Let p be the least prime greater than

$$a = 2^{-1/(r+1)} l^{2/(r+1)}+1. \tag{32}$$

Since for any two consecutive primes p_n and p_{n+1} there is $p_{n+1}-p_n \ll p_n^{\theta}$ for any $\theta > 11/20$ ([HI]) then

$$p < 2^{-1/(r+1)} l^{2/(r+1)}+C_{12} 2^{-\theta/(r+1)} l^{2\theta/(r+1)}. \tag{33}$$

Let $A = \{a_i = p \cdot i|\ 1 \leqslant i \leqslant l/p\}$. We have $\sum_{a_i \in A} a_i \leqslant p\frac{l}{2p}\left(\frac{l}{p}+1\right) = \frac{l(l+p)}{2p}$. Let us show that $p^r > \frac{l(l+p)}{2p}$, or $2p^{r+1} > l(l+p)$.

Indeed,

$$\begin{aligned} 2p^{r+1} &> 2a^{r+1} > 2(a-1)^{r+1}+2(r+1)(a-1)^r \\ &\geqslant l^2+2(r+1)\, 2^{-r/(r+1)} l^{2r/(r+1)} > l^2+lp \end{aligned}$$

by (32) and (33) for l sufficiently large. All subset sums of our A are divisible by p and none by p^r, hence subset sum of this A is never an rth power. In this example $|A| = \left[\frac{l}{p}\right]$, hence we conclude that

$$\begin{aligned} f(l, M_r) &\geqslant \frac{l}{2^{-1/(r+1)} l^{2/(r+1)}+C_{12} 2^{-\theta/(r+1)} l^{2\theta/(r+1)}} \\ &> 2^{1/(r+1)} l^{(r-1)/(r+1)}\left(1+O\left(\frac{1}{l^{\varrho}}\right)\right). \end{aligned}$$

The upper bound in the asymptotic formula (3) we obtain as a consequence of Theorem 6. To prove $f(l, M_r) < 2^{1/(r+1)} l^{(r-1)/(r+1)}+l^{(r-1)/(r+1)-\varrho}$, we suppose on the contrary that A is an arbitrary set (1) with cardinality $|A| = 2^{1/(r+1)} l^{(r-1)/(r+1)}+l^{(r-1)/(r+1)-\varrho}$. We will show that some subset sum of A is the rth power of an integer. Take $y = [\frac{1}{3} l^{(r-1)/(r+1)-\varrho}]$ elements of A, denote this subset by A_y; $|A_y| = y$. Because of $r \geqslant 10$ and $0 < \varrho < 1/6(r+1)$, we have

$$\frac{r-1}{r+1}-\varrho > \frac{4}{5}.$$

Hence we can apply Theorem 6. We obtain that A_y contains a subset $A_y(\bar{s}, \bar{q})$ defined by the Lemma; denote $d_0 = (\bar{s}, \bar{q})$; M_y is an average of elements of A_y; then every natural N, $N \equiv 0 (\operatorname{mod} d_0)$, satisfying

$$C_6 M_y \left(\frac{l}{y}\right)^2 (\log l)^4 < N < C_7 M_y \frac{\sqrt{y}}{\log y} \tag{28'}$$

is a subset sum of $A_y(\bar{s}, \bar{q})$. Denote by Δ a set of such integers N, denote by L_0 and R_0 the left and right bounds of Δ. We can calculate using (28') that

$$R_0/L_0 > 2^r \tag{34}$$

for sufficiently large l. Consider 2 cases.

Case 1. All elements of $A \setminus A_y$ are divisible by d_0 except at most d_0^2-1. Delete from $A \setminus A_y$ the elements not divisible by d_0, denote by A' the set of remaining elements. Clearly

$$|A'| > 2^{1/(r+1)} l^{(r-1)/(r+1)}+\tfrac{1}{3} l^{(r-1)/(r+1)-\varrho}. \tag{35}$$

Construct the set $G = \{\Delta, \Delta+a_1, \ldots, \Delta+a_1+\ldots+a_{|A'|}\}$, where a_j runs over A'. G is an arithmetic progression with the difference d_0, all elements of G are divisible by d_0 and they are subset sums of A. Denote the left and right bounds of G by L_0' and R_0', then $L_0' = L_0$, $R_0' > R_0$. We will show that $d_0^r \in G$ or $(md_0)^r \in G$ with some integer $m > 1$:

First, we check that $d_0^r \leqslant R_0'$.

$$R_0' \geqslant \sum_{a_j \in A'} a_j \geqslant d_0 \sum_{j=1}^{|A'|} j > \frac{d_0}{2}|A'|^2 > d_0 2^{-(r-1)/(r+1)} l^{2(r-1)/(r+1)}$$

holds in view of (35). On the other hand, since all elements of A' are divisible by d_0 and $a_j \leqslant l$, we have $d_0|A'| \leqslant l$. Hence $d_0 \leqslant l/|A'| < 2^{-1/(r+1)} l^{2/(r+1)}$ in view of (35) and hence $d_0^r < d_0 2^{-(r-1)/(r+1)} l^{2(r-1)/(r+1)}$. Therefore $d_0^r < R_0'$.

Secondly, if $d_0^r \geqslant L_0'$ then $d_0^r \in G$ and we have the rth power represented by a subset sum of A. If $d_0^r < L_0'$ then we take the smallest integer m $(m > 1)$ such that $m^r d_0^r \geqslant L_0'$, so that $(m-1)^r d_0^r < L_0'$. We use two inequalities:

$$\frac{m^r}{(m-1)^r} \leqslant 2^r \quad (\text{for } m > 1) \quad \text{and} \quad \frac{R_0'}{L_0'} > 2^r$$

which holds by (34) since $L_0' = L_0$ and $R_0' > R_0$. It follows that

$$m^r d_0^r \leqslant 2^r (m-1)^r d_0^r < 2^r L_0' < R_0'.$$

We obtained that $m^r d_0^r < R_0'$ and consequently $m^r d_0^r \in G$.

Case 2. In $A \setminus A_y$ there are at least d_0^2 elements not divisible by d_0. Then we proceed to the second step of the process by constructing two progressions Δ_1 and G_1. To construct Δ_1 we choose d_0-1 elements $a_1^{(1)}, a_1^{(2)}, \ldots, a_1^{(d_0-1)}$ with the same remainder δ modulo d_0 among d_0^2 elements of $A \setminus A_y$ not divisible by d_0. Denote $d_1 = (d_0, \delta)$. Consider the set $\{\Delta, \Delta + a_1^{(1)}, \ldots, \Delta + a_1^{(1)} + \ldots + a_1^{(d_0-1)}\}$. All elements of this set are divisible by d_1 and they are subset sums of A; the elements between $L_0 + ld_0$ and R_0 form an arithmetic progression with difference d_1. Denote this progression by Δ_1. Its bounds $L_1 = L_0 + ld_0$ and $R_1 = R_0$ satisfy the condition

$$(34') \qquad R_1/L_1 > 2^r$$

because of (28′) and (34). Now we again consider 2 cases.

Case 1. All elements of set $S = (A \setminus A_y) \setminus \{a_1^{(1)}, \ldots, a_1^{(d_0-1)}\}$ except at most $d_1^2 - 1$ are divisible by d_1. Then we construct, using Δ_1, an arithmetic progression G_1 like G before and show that G_1 contains an rth power.

Case 2. In S there are at least d_1^2 elements not divisible by d_1. Then we proceed to the next step. The process will stop after $\log_2 l$ steps at most. ■

THEOREM 3. *If*

$$(4) \qquad C_{13}\, l (\log l)^6 < m < l^{3/2}/(\log l)^3$$

then

$$f(l, m) = l/\bar{m} + h_1$$

where

$$(36) \qquad h_1 = C_{14} \frac{l}{\bar{m}} \frac{\log \bar{m}}{\log^2 l}.$$

Proof. The lower bound $\left[\frac{l}{\bar{m}}\right] \leqslant f(l, m)$ was obtained by N. Alon ([Al]). The upper bound is again a corollary of Theorem 6. Let m be an integer from interval (4). To prove that $f(l, m) < l/\bar{m} + h_1$ we suppose that A is an arbitrary set (1) with cardinality

$$(37) \qquad |A| = x = \left[\frac{l}{\bar{m}} + h_1\right]$$

and will show that m has a representation as a subset sum of A. By (37) we have $x > l/\bar{m}$ and in view of $\bar{m} < \log l$

$$(38) \qquad l/\log l < x.$$

From (38) we observe that $x > l^{4/5+\varepsilon}$, thus we can apply Theorem 6 to A:

(a) If A satisfies condition (9) then all N in the interval

$$(28'') \qquad C_6\, M \left(\frac{l}{x}\right)^2 (\log l)^4 < N < C_7\, M \frac{\sqrt{x}}{\log x}$$

(where M is the arithmetic mean of the elements of A) are subset sums of A. Using $x \ll M \ll l$ and (38) we observe that interval (4) is contained in (28″), so each m from interval (4) is a subset sum of A.

(b) If set A does not satisfy condition (9), then by Theorem 6 there exists a subset $B \subset A$, $B = A(\bar{s}, \bar{q})$ such that each N, $N \equiv 0 \pmod d$ lying in the interval (28″) is a subset sum of B. Here $d = (\bar{s}, \bar{q})$, M is the arithmetic mean of the elements of B. By Lemma $|B| \geqslant x - h(\log_2(x/a) + 1)$ holds where $h = x/(\log_2 l)^2$, so using (37) and (36) we estimate

$$|B| > \frac{l}{\bar{m}} + h_1 - 1 - \frac{l/\bar{m} + h_1}{(\log_2 l)^2}\left(\log_2 \frac{l}{x} + 1\right) > \frac{l}{\bar{m}} + ch_1.$$

On the other hand $|B| \leqslant l/\bar{q}$ and we conclude from $l/\bar{m} < B < l/\bar{q}$ that $\bar{m} > \bar{q}$. Therefore $\bar{q}$ is a divisor of m as well as d, i.e. $m \equiv 0 \pmod d$, hence m is a subset sum of $B \subset A$. ■

References

[Al] N. Alon, *Subset sums*, J. Number Theory, to appear.

[EF] P. Erdős and G. Freiman, *On two additive problems*, J. Number Theory, to appear.

[Er] P. Erdős, *Some problems and results on combinatorial number theory*, Proc. 1st China Conference in Combinatorics (1986), to appear.

G. Freiman, *Waring problem with an increasing number of terms* (Russian), Proc. Elabuga Pedagogical Institute 3 (1958), 105–119.

[F2] — *An analytical method of analysis of linear Boolean equations*, Ann. N.Y. Acad. Sci. 337 (1980), 97–102.

[FJM] G. Freiman, A. Judin, D. Moskvin, *Structural theory of set summation and local limit theorems for independent lattice random variables*, Probability Theory and its Applications 19 (1974), 52–62. (Translation).

[HI] D. R. Heath-Brown and H. Iwaniec, *On the difference between consecutive primes*, Invent. Math. 55 (1979), 49–69.

SCHOOL OF MATHEMATICAL SCIENCES
FACULTY OF EXACT SCIENCES
RAYMOND AND BEVERLY SACKLER
TEL-AVIV UNIVERSITY
Ramat-Aviv, 69978 Tel-Aviv, Israel

Received on 31.8.1987
and in revised form on 12.2.1988 (1746)