On representation of r-th powers by subset sums

by

E. Lipkin* (Tel-Aviv)

Let $A$ be a set of $x$ natural numbers

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

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 $a_x \leq 3(x-1)$. 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 $a_x \leq 4(x-1)$. 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 $m \geq 1$ 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 $l^{-1} < m \leq l^{1.9}$, then

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

$m$ denotes the smallest integer that does not divide $m$. ([AI])

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 $r$th powers?

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

* Research supported in part by the Fund for Basic Research administered by the Israel Academy of Sciences.
that
\[ f(l, M) = \left(\frac{1}{2} + o(1)\right) l \] if \( M \) is the set of all powers of two and
\[ f(l, M) = \left(\frac{1}{2} + o(1)\right) l \] if \( M \) 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 \( f(l, M_2) \),
\[ f(l, M_2) \geq \left(1 + o(1)\right) \cdot 2^{l/3} \cdot l^{1/3}; \]
N. Alon ([Al]) proved that
\[ f(l, M_2) = O(l / \log l). \]
G. Freiman conjectured a general asymptotic formula
\[ f(l, M_r) = 2^{l/(r+1)} l^{r-1} (l^{r+1}) \left(1 + o(1)\right) \]
for \( r \geq 2 \) and suggested that it can be derived by methods of [EF]. The lower bound \( f(l, M_r) \geq 2^{l/(r+1)} l^{r-1} (l^{r+1}) \left(1 + o(1)\right) \) 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 \( c = c(\varepsilon) > 0 \) such that for every \( l > 0 \) and every \( m \), which satisfies \( l^{1+\varepsilon} \leq m \leq l^2 / \log l \), the inequality
\[ \frac{l}{m} \leq f(l, m) < \frac{c}{m} \]
holds.

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

**Theorem 1.** Let \( \varepsilon \) be an arbitrarily small positive number. Then
\[ f(l, M_2) = O\left(l^{4/5 + \varepsilon}\right). \]

**Theorem 2.** For \( r \geq 10 \)
\[ f(l, M_r) = 2^{l/(r+1)} l^{r-1} (l^{r+1}) \left(1 + O\left(\frac{1}{l}\right)\right) \]
where \( q \) is an arbitrary positive number less than \( l/(6(r+1)) \).

**Theorem 3.** If
\[ C(l \log l)^6 < m < l^{1/3} / (\log l)^3 \]
then
\[ f(l, m) = l/m + h_1 \]
where \( h_1 = c \left(\frac{l \log m}{m \log^2 l}\right) C \) 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 \( A \subset \mathbb{N} \) and \( s, q \in \mathbb{N} \), \( q \geq 2 \) let \( A(s, q) = \{a | a \in A, a = s (\mod q)\} \).

Let \( [a] \) denote the smallest integer \( \geq a \).

\( \alpha_1, \alpha_2, \ldots \) denote positive constants.

\( I = I(N) \) denotes the number of solutions of the equation
\[ x_1 + x_2 + \ldots + x_n = N, \]
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
\[ M = a_1 + \ldots + a_n, \]
\[ D = \frac{1}{x} \sum_{i=1}^{x} a_i^2 - M^2. \]

**Theorem 4.** Let \( A \subset \{1, 2, \ldots, l\} \) be a set (1), \( |A| = x \). Suppose \( x > l^{4/5 + \varepsilon} \), where \( \varepsilon \) is an arbitrarily small positive number and \( l > l_0(\varepsilon), \) and suppose that
\[ |A(s, q)| < x - h \]
for all \( s, q \in \mathbb{N}, q \geq 2, \) where
\[ h = x/\log^2 l. \]

Let \( n \) and \( N \) in (6) satisfy
\[ C_1 \left(\frac{1}{x}\right) (\log l)^6 < n < C_2 \left(\frac{1}{x}\right) \log x \]
(it is possible because of the assumption \( x > l^{4/5 + \varepsilon} \)) and
\[ Mn - C_3 \sqrt{nD} < N < Mn + C_4 \sqrt{nD} \]
where \( C_1, C_2, C_3, C_4 \) are any fixed numbers. Then
\[ I = x^N \left(\frac{1}{2\pi nD} e^{-\left((Mn-N)^2/2nD\right)} + O\left(\frac{x^N}{\sqrt{nD}}\right)\right). \]

**Proof.** It is known that the number of solutions of equation (6)
\[ x_1 + \ldots + x_n = N, x_i \in A \]
is
\[ I = I(N) = x^N \int_0^1 \phi^n(x) e^{-2\pi i n x} \, dx \]

4 - Acta Arithmetica XII, n. 4.
where
\[ \varphi(x) = \frac{1}{x} \sum_{n \neq 0} e^{2i\pi n x}. \]

Define the number
\[ L = C_3 l, \]
where \( C_3 \) is sufficiently large. Since the subintegral function has period 1,
\[ I(N) = x^{-\frac{1}{l}} \int_{-1/L}^{1/L} \varphi^*(x) e^{-2i\pi x N} dx. \]

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
\[ I_1 = \int_{-1/L}^{1/L} \varphi^*(x) e^{-2i\pi \frac{N x}{x}} dx = \frac{1}{\sqrt{2\pi N D}} e^{-\frac{(2\pi N - N)^2}{2\pi N D}} (1 + o(1)) \]
and that
\[ I_2 = \int_{1/L}^{1} \varphi^*(x) e^{-2i\pi \frac{N x}{x}} dx = o(1/\sqrt{N D}) \]
for all \( N, n \) which satisfy (11) and (12).

We first show (15). Let us estimate \( \varphi(x) \) for \( x \in [1/L, 1-1/L] \). Each number \( x \in [0, 1] \) has a representation \( x = p/q + z \), \( (p, q) = 1, 1 \leq q \leq L, |z| < 1/q(L) \); for \( x \in [1/L, 1-1/L] \) we have \( q \geq 2 \). Then we can represent \( \varphi(x) \) in the form
\[ \varphi(x) = \frac{1}{x} \sum_{n \neq 0} e^{2i\pi (p/q + z n)} = \frac{1}{x} \sum_{k = 0}^{q-1} e^{2i\pi \frac{k}{q} x}. \]

where
\[ |z| < \frac{1}{L} \sqrt{1 - \frac{1}{4q^2}}. \]

Denote by \( m_k \) the number of solutions of a congruence \( px_j \equiv k \mod q \) for \( 0 \leq k < q \) and \( 1 \leq j \leq x \). Consider three different cases according to the value of \( q \), for a sufficiently large \( l \). We will use the inequality
\[ \frac{1}{x} \sin x u < \frac{1}{x} \frac{y}{u} - \frac{1}{2} \frac{y^2}{u^2} < 1 - \frac{1}{x} \frac{y^2}{4} \]
which holds for \( 0 < y < \pi/2 \) with \( y \geq 2 \).

1. Case \( q \geq l \). In this case \( m_k \leq 1 \). Then we estimate
\[ |\varphi(x)| \leq \frac{1}{x} \sum_{k = 0}^{q-1} e^{2i\pi \frac{kx}{2q}} = \frac{1}{x} \frac{\sin \frac{\pi x}{2q}}{\sin \frac{\pi}{2q}} \leq 1 - \frac{1}{4.6} \left( \frac{\pi x}{2q} \right)^2 \]
and by (19), using \( q < L \) and \( 1/q > 1/(C_3 l) \), we have
\[ |\varphi(x)| < 1 - \frac{1}{4.6} \frac{\pi^2}{C_3^2} \left( \frac{1}{x} \right)^2. \]

2. Case \( 1 < q < \frac{L}{2} \). 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 \( e^{2i\pi/h} \) and estimate using (17) and (10)
\[ |\varphi(x)| \leq \frac{1}{x} |x - h + h x e^{2i\pi/4q}| \]
\[ = \left| 1 - 2h \frac{x}{x} + \frac{h}{x} \frac{1 + e^{2i\pi/4q}}{2} \right| \leq 1 - \frac{2h}{x} + \frac{1}{x} \frac{1 + e^{2i\pi/4q}}{2} \]
\[ = 1 - \frac{2h}{x} \frac{x}{4q} \sin^2 \frac{\pi}{4q} = 1 - \frac{1}{4.6} \frac{\pi^2}{4q} \frac{1}{x} \left( \frac{1}{x} \right)^2. \]
by \( \sin u < \frac{2}{\pi} u \) and \( \sin^2 \frac{\pi}{4q} > \frac{1}{4q} > \frac{1}{4.6} \left( \frac{1}{x} \right)^2 \).

3. Case \( \frac{L}{2} \leq q < l \). In this case \( m_k \leq \lceil L/q \rceil \leq 2l/q \) for all \( k \). Define \( m = \lceil 2l/q \rceil \) and \( r = \lceil x/(4l/q) \rceil = \lceil xq/(4l) \rceil \). Then \( m \geq 2l/q, r \geq xq/(4l) \) and \( mr \geq x/2 \). Denote \( t = x - mr \), then \( t \leq x/2 \). Replace in the sum (16) \( t \) terms by \( 1 \), \( m \) terms by \( e^{2i\pi/kq} \) for each \( k = 0, 1, \ldots, r - 1 \) and estimate using (17), and (18) since \( r \geq 2 \)
\[ |\varphi(x)| \leq \frac{1}{x} \frac{1}{x} \sum_{k = 0}^{r-1} e^{2i\pi/kq} \]
\[ = t + \frac{m}{x} \sin \frac{\pi t}{2q} = t + \frac{m}{x} \frac{1}{x} \frac{1}{x} \frac{1}{x} \ sin \frac{\pi t}{2q} \]
\[ < x - mr + \frac{m}{x} \frac{1}{x} \left( 1 - \frac{1}{4.6} \left( \frac{\pi t}{2q} \right)^2 \right) \]
\[ = 1 - \frac{m}{x} \frac{\pi^2}{4.6} \frac{r}{x} < 1 - \frac{\pi^2}{2.4.6} \left( \frac{r}{x} \right)^2 \]
in view of \( mr/x \geq 1/2 \) and \( r/q \geq x/4l \).

From these three cases we conclude by (19'), (20), (21) that for all \( x, 1/L \)
\[ \alpha < 1 - \frac{1}{1/L} \]

\[ |\varphi(x)| < 1 - c_0 \frac{1}{\log^2 l} \left( \frac{x}{l} \right)^2 \phi(x) \]

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

\[ (22) \quad |\varphi(x)| < \left( 1 - c_0 \frac{1}{\log^2 l} \left( \frac{x}{l} \right)^2 \phi(x) \right) \leq \left( 1 - c_0 \frac{1}{\log^2 l} \left( \frac{x}{l} \right)^2 \right) \leq \frac{1}{l^2} \]

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

\[ \int_{-1/L}^{1/L} \varphi(x) e^{-2 \sin x} dx = O(1/l^2) = o(1/\sqrt{nD}) \]

and (15) follows.

Next we estimate integral \(I_1 = \int_{-1/L}^{1/L} \varphi(x) e^{-2 \sin x} dx\) to prove (14). By (7), (8) \(D > C x^2\) with some constant \(C\) and by (11) \(nD > C l^2 (\log l)^4\), 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 \(I_1 = \int_{-1/L}^{1/L} \varphi(x) e^{-2 \sin x} dx\) equals the sum of the three integrals. For all \(x \in [-1/L, 1/L]\),

\[ |\varphi(x)| < \frac{1}{C s l^{-1}} = \frac{1}{C s} \]

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

\[ (23) \quad \varphi(x) = \sum_{n \geq 0} \varphi(n) = 2 \sin x M - 2 \sin^2 x (D + M^2) + o(x^2 (D + M^2)) \]

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

\[ |\varphi(x)| e^{-2 \sin x} dx < 1 - e^{-2 \sin x} dx = 1/|x^2| < 1/l^2 \]

holds and we conclude that \(\int_{-1/L}^{1/L} \varphi(x) e^{-2 \sin x} dx = o(1/\sqrt{nD})\). 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 \(N \in \mathbb{N}\) in interval (12) can be represented as a subset sum of \(A, N = \sum_{a_i \in A} a_i\) 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

\[ (25) \quad Q = 1 + o(x^2/\sqrt{nD}) \]

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 \(O(x^{-2}/\sqrt{nD})\) according to Theorem 4. Thus \(Q_i = O\left(n^{-2}/\sqrt{nD}\right)\).

Notice that \(N - 2a_i\) belongs to the interval (12) if we take the number \(C_3\) to be sufficiently large. By (11), \(\sum_{i=1}^{x} Q_i = O(x^2/\sqrt{nD})\) 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), \(x > P\), for some \(\alpha > 0\) and \(l\) be sufficiently large; \(h = x/\log^2 l\). Then there exists \(B \subseteq A\) such that

(i) \(|B| > |A| - (\log_2 (l/x) + 1)h\),
(ii) \(B\) is contained in an arithmetic progression, i.e., for some \(\bar{s}\) and \(\bar{q} \in \mathbb{N}\), \(b_j = \bar{s} + \bar{q} \in B\),
(iii) \(|B| < h\) 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 \(A_{s_0} = A(s_0, q_0)\) we have \(|A_{s_0}| > A - h\). If condition (iii) for \(A_{s_0}\) and \(\bar{q} = q_0\) holds, we put \(B = A_{s_0}\), and if not, we can find \(q_1 > 2q_0\) and \(s_1\) such that for \(A_{s_1} = A(s_1, q_1)\), it is \(|A_{s_1}| - h > |A_{s_0}| - h\). Suppose that we arrived at \(A_k = A_{s_k - 1}(s_k - 1, q_k - 1)\) where

\[ (26) \quad k = [\log_2 (l/x) + 1] \]

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 \(|A_{s_k} - 1| = |A_{s_k}(s_k, q_k)| > |A_{s_k}| - h\). By (26) we have \(2^k > 2\) hence

\[ (27) \quad |A_{s_k} > |A| - (k+1)h > x/2 \geq l/2^k \]
On the other hand, \( A_{k+1} = A_k(s_k, q_k) \) is contained in an arithmetic progression, so we have \( |A_{k+1}| \leq l/q_k \leq l/2^{k+1} \) which contradicts (27).

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

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

**Theorem 6.** Assume that set \( A \) in (1) satisfies the condition \( x > l^{\alpha + t} \) with arbitrary small positive \( \alpha \). 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 \( l > l_0(\alpha) \) each natural number \( N, N \equiv 0 \pmod{d} \) satisfying

\[
C_6 M' \left( \frac{1}{x} \right)^2 \left( \log \frac{l}{x^4} \right) < N < C_7 M' \frac{\sqrt{x}}{\log x}
\]

with some constants \( C_6, C_7 \) can be represented as a subset sum of \( B, \quad N = \sum_{a \equiv 0 \pmod{d}} a_i \) where \( G \leq B \).

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

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 \( t_1 < h/\bar{q} < t_j \), \( j = 1, \ldots, y \), \( y > l^{\alpha + t} > \left( \frac{l}{x} \right)^{\alpha + t} \) where \( 0 < \varepsilon_1 < \varepsilon \). 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 \( n \equiv n_0 \pmod{\bar{q}} \) and

\[
C_1 \left( \frac{1}{\bar{q}^2} \right)^2 \left( \log \frac{1}{\bar{q}} \right)^4 < n < C_2 \frac{\sqrt{y}}{\log y};
\]

then each natural \( \tilde{N} \) in the interval

\[
M'' \tilde{N} - C_3 \sqrt{n D''} < \tilde{N} < M'' \tilde{N} + C_4 \sqrt{n D''}
\]

can be represented as a subset sum of \( T \), i.e. \( \tilde{N} = \tilde{N}_1 + \ldots + \tilde{N}_y, t_j \in T \).

Let us come back to \( B \). From \( b_1 = \bar{s}/\bar{q} + \ldots + (b_j - \bar{s})/\bar{q} = \tilde{N} \) follows

\[
b_1 + \ldots + b_j = \bar{q} \tilde{N} + n \bar{s}.
\]

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

\[
M'' \bar{q} \tilde{N} - C_3 \bar{q} \sqrt{n D''} + \bar{s} n < N < M'' \bar{q} \tilde{N} + C_4 \bar{q} \sqrt{n D''} + \bar{s} n
\]

where \( n \equiv n_0 \pmod{\bar{q}}, n \) 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 \( n \equiv n_0 \pmod{\bar{q}} \). First we take two consecutive \( n \) from interval (11): \( n \) and \( n + \bar{q} \). Interval (29) for \( n + \bar{q} \) looks like

\[
M'' \bar{q} (n + \bar{q}) - C_3 \bar{q} \sqrt{(n + \bar{q}) D''} + \bar{s} (n + \bar{q}) < N
\]

\[
< M'' \bar{q} (n + \bar{q}) - C_4 \bar{q} \sqrt{(n + \bar{q}) D''} + \bar{s} (n + \bar{q})
\]

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

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

or

\[
M'' \bar{q}^2 < C_{10} n D''
\]

for every positive constant \( C_{10} \). Since \( M'' \bar{q}^2 \leq l^2, D'' \geq x^2 \) and \( n \gg (l/x)^2 \log l \), (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 \( \bar{q} M'' < M'' \bar{q} < C_1 \bar{q} M'' \) where \( C_1 \) is a constant. We showed that all \( N \) from the interval (28), satisfying the condition \( N \equiv n_0 \bar{s} \pmod{\bar{q}} \), 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 \( \{ s \} \), \( |A| = x \), satisfying \( x > l^{\alpha + t} \) where \( \alpha \) is an arbitrarily small positive number. Then for \( l > l_0(\alpha) \), there exists a square equal to a subset sum of \( A \). In other words \( f(l, M_2) = O(l^{\alpha + t}) \).

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

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

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

\[
C_7 \frac{\sqrt{x}}{\log x} \left( \frac{C_6 M' \left( \frac{1}{x} \right)^2 \left( \log \frac{l}{x^4} \right)}{\log x \left( \log l \right)^2} \right)^{\beta \log 2}
\]

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 set of all $r$-th powers. For $r \geq 10$ and $q$ being an arbitrary positive number less than $1/6(r+1)$ we have the following asymptotic formula:

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

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

$$a = 2^{-1/(r+1)} r^{2/(r+1)} + 1.$$

Since for any two consecutive primes $p_n$ and $p_{n+1}$ there is $p_{n+1} - p_n < p_n^2$ for any $\theta > 1/20$ ([Hi]),

$$p < 2^{-1/(r+1)} r^{2/(r+1)} + C_{12} 2^{-6/(r+1)} r^{2/(r+1)} + 1.$$

Let $A = \{a_i = p^i | 1 \leq i \leq l/p\}$. We have

$$\sum_{a_i \in A} a_i \leq l \frac{l}{2p} \left(\frac{l}{p} + 1\right) = \frac{l(l+p)}{2p}.$$

Let us show that $\frac{l(l+p)}{2p} > 2^{2/(r+1)}$, or $2p^{r+1} > l(l+p)$. Let $l$ be sufficiently large. All subset sums of our $A$ are divisible by $p$ and none by $p'$, hence subset sum of this $A$ is never an $r$-th power. In this example $|A| = \left\lfloor \frac{l}{p} \right\rfloor$, hence we conclude that

$$f(l, M_r) \geq 2^{-1/(r+1)} r^{2/(r+1)} + C_{12} 2^{-6/(r+1)} r^{2/(r+1)} + \left(1 + O\left(1/l^r\right)\right).$$

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

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

Hence we can apply Theorem 6. We obtain that $A_y$ contains a subset $A_y(\delta, \varepsilon)$ defined by the Lemma; denote $d_0 = (\delta, \varepsilon)$; $M_r$ is an average of elements of $A_y$; then every natural $N, N \equiv 0 \pmod{d_0},$ satisfies

$$C_{12} M_r \left(\frac{l}{p}\right)^2 \log \frac{l}{p} < N < C_{12} M_r \left(\frac{l}{p}\right)^{\sqrt{y}} \log y$$

is a subset sum of $A_y(\delta, \varepsilon)$. Denote by $A$ a set of such integers $N$, denote by $L_0$ and $R_0$ the left and right bounds of $A$. We can calculate using (28) that

$$R_0 L_0 > 2^r$$

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

$$|A'| > 2^{1/(r+1)} r^{r-1/(r+1)} + \frac{1}{2} \left(1 - 1/(r+1) - \varepsilon\right).$$

Construct the set $G = \{A, A + a_1, \ldots, A + a_1 + \ldots + a_{|A|}\}$, 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 $d_0 \in G$ or $(md_0)^e \in G$ with some integer $m > 1$;

First, we check that $d_0 \leq R_0$.

$$R_0 > \sum_{a_j \in A} a_j \geq d_0 \sum_{j=1}^{4} a_j + \frac{d_0}{2} |A|^2 > d_0 2^{1/(r+1)} r^{r-1/(r+1)} + \frac{1}{2} (1 - 1/(r+1) - \varepsilon).$$

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

Secondly, if $d_0 \geq L_0$ then $d_0 \in G$ and we have the $r$-th power represented by a subset sum of $A$. If $d_0 < L_0$ then we take the smallest integer $m (m > 1)$ such that $md_0 \geq L_0$, so that $(m-1)d_0 < L_0$. We use two inequalities:

$$\frac{m^r}{(m-1)^r} \leq 2^r$$

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

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

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

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 \(d_1\) and \(G_1\). To construct \(d_1\), we choose \(d_0 - 1\) elements \(a_1^{(1)}, a_2^{(1)}, \ldots, a_{d_0}^{(1)}\) with the same remainder \(\delta\) modulo \(d_0\) among \(d_0^2\) elements of \(A \setminus A_2\) not divisible by \(d_0\). Denote \(d_1 = (d_0, \delta)\). Consider the set \(\{d, A + d^{(1)}, \ldots, d + d_0^{(1)} + \ldots + a_{d_0}^{(1-1)}\}\). All elements of this set are divisible by \(d_1\) and they are subset sums of \(A\); the elements between \(L_0 + l d_0\) and \(R_0\) form an arithmetic progression with difference \(d_1\). Denote this progression by \(A_1\). Its bounds \(L_1 = L_0 + l d_0\) and \(R_1 = R_0\) satisfy the condition

\[
R_1 / L_1 > 2^{(34, 34')}
\]

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

1. Case 1. All elements of set \(S = (A \setminus A_2) \setminus \{a_1^{(1)}, \ldots, a_{d_0}^{(1)}\}\) except at most \(d_0^2 - 1\) are divisible by \(d_1\). Then we construct, using \(d_1\), an arithmetic progression \(G_1\) like \(G\) before and show that \(G_1\) contains an \(r\)th power.

2. Case 2. If \(S\) there are at least \(d_0^2\) elements not divisible by \(d_1\). Then we proceed to the next step. The process will stop after \(\log_2 l\) steps at most. \(\blacksquare\)

**Theorem 3.** If

\[
C_{14} l (\log l)^6 < m < l^{1/2}(\log l)^3
\]

then

\[
f(l, m) = \frac{l}{m} + h_1
\]

where

\[
h_1 = \frac{l \log m}{C_{14} (\log l)^2}
\]

**Proof.** The lower bound \(f(l, m) \geq \frac{l}{m}\) was obtained by N. Alon ([AI]).

The upper bound is again a corollary of Theorem 6. Let \(m\) be an integer from interval (4). To prove that \(f(l, m) < l/m + h_1\), suppose that \(A\) is an arbitrary set (1) with cardinality

\[
|A| = \frac{l}{m} + h_1
\]

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

\[
l/\log l < x
\]

From (38) we observe that \(x > l^{1/5 + \varepsilon}\), thus we can apply Theorem 6 to \(A\):

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

\[
C_6 M \left(\frac{1}{x}\right)^2 \log l + N < C_7 M \sqrt{x / \log x}
\]

(where \(M\) is the arithmetic mean of the elements of \(A\)) are subset sums of \(A\). Using \(x \in M \leq l\) 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 \(B \subset A\), \(B = \{\bar{a}, \bar{q}\}\) such that each \(N, N \equiv 0 (\mod d)\) lying in the interval (28) is a subset sum of \(B\). Here \(d = (\bar{a}, \bar{q})\), \(M\) is the arithmetic mean of the elements of \(B\). By Lemma \(|B| > x - h(\log (x/a) + 1)\) holds where \(h = x/(\log_2 l)^2\), so using (37) and (36) we estimate

\[
|B| > \frac{1}{m} + h_1 - 1 = \frac{l/m + h_1}{(\log_2 l)^2} \left(\log_2 \frac{1}{x} + 1\right) > \frac{1}{m} + c_1.
\]

On the other hand \(|B| < l/\bar{q}\) and we conclude from \(l/\bar{m} < B < l/\bar{q}\) that \(\bar{m} > \bar{q}\).

Therefore \(\bar{q}\) is a divisor of \(\bar{m}\) as well as \(d\), i.e., \(\bar{m} \equiv 0 (\mod \bar{d})\), hence \(m\) is a subset sum of \(B \subset A\).

**References**


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

Received on 31.1.1987
and in revised form on 12.2.1988

(1746)