New analytical results in subset-sum problem

Gregory A. Freiman*

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA
School of Mathematics, Raymond and Beverly Sackler, Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel

Received 2 March 1989
Revised 17 September 1990

Abstract


An analytical method is developed to prove that, for the integer set \( A \subseteq [1, l] \), with \( l > l_0 \) and \( |A| = m > c_1 l^{1/2} (\log l)^{1/2} \), the set \( A^* \) of subset sums contains a long arithmetic progression of length larger than \( c_2 m^2 \). Here \( l_0 \) and \( c_1 \) are sufficiently large constants and \( c_2 \) is some positive constant.

This result gives a possibility to solve new algorithmic and combinatorial problems connected with subset sums.

1. Introduction

Let \( A = \{a_1, a_2, \ldots, a_m\} \), \( 1 \leq a_1 < a_2 < \cdots < a_m \leq l \), \( a_i \in \mathbb{N} \), \( l \in \mathbb{N} \) and

\[
a_1 x_1 + a_2 x_2 + \cdots + a_m x_m \leq n, \tag{1}
\]

where \( a_i \in A \), \( x_i \in \{0, 1\} \), \( 1 \leq i \leq m \), \( n \in \mathbb{R} \).

The problem of finding the maximum of the linear form on the left-hand side in (1) such that the inequality (1) holds is known as the subset-sum problem (SSP) (see, for example, [12]).

Analytical methods were used in [8,9] in order to solve SSP when \( m \) is sufficiently large compared to \( l \), namely, \( m > l^{2/3} + \varepsilon \), where \( \varepsilon \) is an arbitrary small positive number and \( l > l_0 \), \( l_0 \) being a sufficiently large positive constant.

In this paper we prove the existence of a long arithmetic progression in the set \( A^* = \{S_B = \sum_{b \in B} b \mid B \subseteq A\} \) of subset sums when \( m > c_1 l^{1/2} (\log l)^{1/2} \) and \( l > l_0 \). Here and

Correspondence to: Gregory A. Freiman, School of Mathematics, Tel Aviv University, Tel Aviv, Israel.
* Research supported by the Fund of Basic Research administrated by the Israel Academy of Sciences.
later $l_0, c_1, c_2, \ldots$ denote sufficiently large constants if they are not defined differently. This result gives the possibility of solving SSP in a wider range of problems than before.

We formulate the main result in the following theorem.

**Theorem 1.** Let $A = \{a_1, a_2, \ldots, a_m\}$ be a set of distinct integers from the segment $[1, l]$, let $l > l_0$ and

$$m > c_1 l^{1/2}(\log l)^{1/2}.$$  \hspace{1cm} (2)

Then there exists a natural number $d$,

$$d \leq \frac{3l}{m},$$  \hspace{1cm} (3)

such that each integer $M$, with $d \mid M$ and $M \in \mathbb{N}$, belongs to $A^*$.

We use the notation $f \ll g$ or $g \gg f$ instead of $f = O(g)$, and write $f \asymp g$ when $f = O(g)$ and $g = O(f)$.

Sections 2–4 are devoted to the proof of Theorem 1.1. In Section 2 we consider the case where elements of $A$ are 'well distributed'. Then we obtain the asymptotic formula for the number of representations of integers by subset sums, which gives (4) with $d = 1$. To study the situation where the asymptotic formula is not valid, we define in Section 3 a map which transforms our problem into a two-dimensional one with increased density. Such maps were introduced in [6, 7]. The high density allows us to use the results of [10]. In Section 4 the proof of Theorem 1.1 is completed, and we discuss its possible applications in Section 5.

2. Proof of Theorem 1.1

Let

$$I = 2^m \int_0^1 \phi(x) e^{-2\pi i M x} dx,$$

where $\phi(x) = \prod_{j=1}^m \phi_j(x)$ and $\phi_j(x) = \frac{1}{2}(1 + e^{2\pi i x j})$, $j = 1, 2, \ldots, m$. The number $I$ is equal to the number of solutions of the equation

$$a_1 x_1 + a_2 x_2 + \cdots + a_m x_m = M,$$

where $a_i \in A, x_i \in \{0, 1\}, 1 \leq i \leq m, M \in \mathbb{N}$. The function $\phi(x) = e^{-2\pi i x M}$ has period 1; so, we may now write, for $l' = 2l$,

$$I = 2^m \int_{-1/l'}^{1/l'} \phi(x) e^{-2\pi i M x} dx = 2^m \int_{-1/l'}^{1/l'} \phi(x) e^{-2\pi i M x} dx + 2^m \int_{-1/l'}^{1/l'} = 2^m (I_1 + I_2).$$
New analytical results in subset-sum problem

Let \( B^2 = \frac{1}{2} \sum_{j=1}^{m} a_j^2 \) and suppose that
\[
B \gg l \log^{1/2} l.
\]
(5)

For \( l_1 \), we have (see, for example, \[2, pp. 7-8\])
\[
I_1 = \frac{1}{\sqrt{2\pi B^2}} (e^{-M - S_{l/2}^{1/2} B^2} + o(1)).
\]
(6)

Since \( B^2 \leq m^2 \leq l^2 \), we obtain an asymptotic formula for \( l \) when
\[
I_2 \leq \frac{1}{l^2} = o\left( \frac{1}{B} \right).
\]
(7)

Let \( \|x\| \) denote the distance from \( x \) to the nearest integer. Using the inequality
\[
|\varphi(j(x))| \leq e^{-(n/2)\|x\|_2^2},
\]
we have
\[
I_2 \leq \max_{1 \leq s \leq 1 - \frac{1}{l}} |\varphi(j(x))| \leq e^{-(n/2)\min_{1 \leq s \leq 1 - \frac{1}{l}} \|x\|_2^2}.
\]
(8)

If
\[
\min_{1 \leq s \leq 1 - \frac{1}{l}} \sum_{j=1}^{m} \|x_a_j\|^2 > 10 \log l,
\]
(9)

we obtain (7) and, using (6), we get the asymptotic formula
\[
l = \frac{1}{\sqrt{2\pi B^2}} (e^{-M - S_{l/2}^{1/2} B^2} + o(1)).
\]
(10)

The asymptotic formula (10) implies that all integers \( M \) for which
\[
\left| M - \frac{1}{2} S_A \right| < B
\]
belong to \( A^* \). Thus, we have obtained an arithmetic progression of length \( B \), with the difference equal to 1. Remembering that \( B \gg m^{3/2} \) and \( B \gg l \log^{1/2} l \), we see that the desired length \( m^2 \) is not achieved.

To obtain this length, let us take \( A_1 = \{ a_j | a_j \in A, \lfloor m/2 \rfloor < j \leq m \} \). Suppose that the asymptotic formula (10) is valid for \( A_1 \). To obtain it, we have used conditions (5) and (9). The cases where these conditions are not fulfilled for \( A_1 \) will be studied later in Sections 3 and 4.

From (11) we obtain the arithmetic progression \( P_0 \) of length
\[
\gg l \log^{1/2} l,
\]
(12)

with the difference equal to 1, which belongs to \( A^*_1 \). Let now \( P_i = P_{i-1} + \{0, a_i\} \), \( i = 1, \ldots, \lfloor m/2 \rfloor \). Since \( a_i \leq l \), in view of (12), each of the sets \( P_i \) is an arithmetic progression and \( |P_i| - |P_{i-1}| \gg i \). Clearly, \( P_{\lfloor m/2 \rfloor} \subseteq A^* \) and \( |P_{\lfloor m/2 \rfloor}| \gg \sum_{i=1}^{\lfloor m/2 \rfloor} i > m^2/10 \).
3. Proof of Theorem 1.1

We study now the case where (8) does not hold for $A_1$. This means that there exists $x$, $1/l' \leq x \leq 1 - 1/l'$, for which

$$\sum_{j \in (m, 2m]} \| x a_j \|^2 \leq 10 \log l. \quad (13)$$

Take all numbers $a_j \in A_1$ for which $\| x a_j \|^2 \geq 15c_3 \log l/m$. If the number of such $a_j$ were bigger than $m/c_3$, we would have $\sum_{j \in (m, 2m]} \| x a_j \|^2 > 10 \log l$, contrary to (13). We see that, for the set $A_2 = \{ a_j \in A_1 \mid \| x a_j \| < 4c_3^{1/2} \log^{1/2} l/m^{1/2} \}$, we have

$$|A_2| \geq m \left( \frac{1}{2} - \frac{1}{c_3} \right). \quad (14)$$

Let us study the set

$$D = \{ x \mid x \in [1, l], x \in \mathbb{Z}, \| x \| < 4c_3^{1/2} \log^{1/2} l/m^{1/2} \}. \quad (15)$$

Evidently, $A_2 \subset D$.

We introduce the set $D$, having in mind the following goals. We will show later (see (25) and (29)) that $|D| \leq (l \log l/m)^{1/3}$. So, we have, in view of (2) and (14),

$$|A_2| \geq |D|^{2/3} \log^{1/3} |D|. \quad (16)$$

Thus, the density of $A_2$ in $D$, characterized by (16), is substantially higher than the density of $A$ in $[1, l]$, characterized by (2). Remember that in [9] the problem was solved for a set $A$ satisfying condition (2) (see [9], p. 208, (2)), namely,

$$|A| \geq l^{2/3} \log^{1/3} l. \quad (17)$$

Conditions (16) and (17) seem very similar. But there is a difference, too, for the set $D$ is not a segment. So, we build injection $D \rightarrow \mathbb{Z}^2$, sending $D$ into the set of integer points of some convex domain $H \subset \mathbb{R}^2$ which is a natural two-dimensional generalization of a segment. This map $\rho$ transforms our one-dimensional problem into a two-dimensional one, which was solved in [10] for density conditions corresponding to condition (16).

We realize this plan as follows.

Let $Q = (l/\log l)^{1/2}$. For each $x \in [0, 1]$, we have a well-known representation $x = p/q + z$, where $p, q \in \mathbb{Z}$, gcd$(p, q) = 1$, $1 \leq q \leq Q$ and $|z| < 1/qQ$.

Let us study two different cases: Case A when

$$|z| < \frac{\log^{1/2} l}{m^{1/2} l} \quad (18)$$

and Case B when

$$|z| \geq \frac{\log^{1/2} l}{m^{1/2} l}. \quad (19)$$
Case A: Find $p'$ such that $p' p \equiv 1 \pmod{q}$ and $1 \leq p' \leq q - 1$. If

$$x \equiv p's \pmod{q}, \quad -\frac{q}{2} < s \leq \frac{q}{2}$$

then

$$\|zx\| = \left\|\frac{p}{q} x + zx\right\| = \left\|\frac{s}{q} + zx\right\|.$$

Let $h$ be the number of elements of $A_2$ which are not divisible by $q$. If $h < m/c_5$ then the set $A_2$ has a special structure (a large part of it belongs to some arithmetic progression). This situation will be analyzed in the next section. So, we assume that

$$h \geq \frac{m}{c_4}. \quad (21)$$

Let us study the case where

$$q < c_5 \left(\frac{m}{\log l}\right)^{1/2}. \quad (22)$$

From (21) and (22) we obtain $h > m/c_4 > q^2 \log l/c_4 c_3 = c_6 q^2 \log l$ (we take $c_6 = 1/c_4 c_3$).

We suppose that $z > 0$ (the case where $z < 0$ can be investigated in the same way). From (18) and (22) we have $z x < z l < \log^{1/2} l/m^{1/2} < 1/2 q$ and for $q + x$ it follows that $\|s/q + 2x\| > 1/2 q$.

We see that, in the case of (21), $\sum \|x a_j\|^2 \geq h/4 q^2 > (c_6/4) \log l$, in contradiction (13).

We have proved that, for $x$ satisfying (13), we have

$$q \geq \left(\frac{m}{\log l}\right)^{1/2}. \quad (23)$$

Let $D' = \{ x \mid x \in [1, l] \}, x \in \mathbb{Z}, x \equiv p's \pmod{q}, |s| \leq q(4 c_3^{1/2} + 1) \log l/m^{1/2} + 1/2\}$ then, in view of (15), (18) and (20), we have $D \subseteq D'$.

We will now build maps $\rho : D' \rightarrow \mathbb{Z}^2, \rho(x) = (x, s)$, where $s$ is determined by (20), and $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}, \psi(x, y) = x$.

Define the rectangle $H = \{ v \mid v \in \mathbb{R}^2, v = x_1 l_1 + x_2 l_2, 0 \leq x_1 \leq 1, \ 1 \leq x_2 \leq 1\}$, where $l_1 = (l, 0), l_2 = (0, s_0), s_0 = \lfloor q(4 c_3^{1/2} + 1) \log l/m^{1/2} + 1/2\rfloor$, and the lattice $\Gamma = \{ v \mid v \in \mathbb{Z}^2, \ v = k_1 u_1 + k_2 u_2, \ k_1, k_2 \in \mathbb{Z}\}$, where $u_1 = (q, 0), u_2 = (p', 1)$ are the vectors of the basis. We have, evidently,

$$\psi(H \cap \Gamma) \supseteq D. \quad (24)$$

In fact, the point $\rho(x) = (x, s)$ lies in $H$ according to our construction. Additionally, from (20) we have $x = q t + p's$ for some $t \in \mathbb{Z}$ and $(x, s) = t(q, 0) + s(p', 1) = t u_1 + su_2 \in \Gamma$.

Now, using (24), we have

$$|D| \leq |H \cap \Gamma| \leq \left\lfloor \frac{l}{q} \right\rfloor + 1 \left(2s_0 + 1\right) \leq \frac{l \log^{1/2} l}{m^{1/2}}. \quad (25)$$
Case B: Let \( x \equiv -p's \pmod{q} \), where \( pp' \equiv 1 \pmod{q} \), \( 1 \leq p' \leq q - 1 \). Then, for \( x \in D \), we have

\[
\|zx\| = \left\| \frac{p}{q} x + zx \right\| = \left\| -\frac{s}{q} + zx \right\| < \frac{4c^{1/2} \log^{1/2} l}{m^{1/2}}.
\]  

(26)

Define

\[
D_x = \left\{ x \mid x \in \mathbb{Z}, x \equiv -p's \pmod{q}, \left\| -\frac{s}{q} + zx \right\| < \frac{4c^{1/2} \log^{1/2} l}{m^{1/2}} \right\}.
\]

(27)

The numbers of the set \( D_x \) fulfill the condition (26). On the other hand, the number \( x \), fulfilling (26), for which \( x \equiv -p's \pmod{q} \), belongs to one of \( D_{x'} \), \( s' \equiv s \pmod{q} \). In fact, (26) implies the existence of an integer \( t \) for which \( |t - s/q + zx| < 4c^{1/2} \log^{1/2} l/m^{1/2} \), i.e. \( s' = -qt + s \). From (27) we obtain, for \( x \in D_{x'}, \)

\[
\frac{s}{zq} - \frac{4c^{1/2} \log^{1/2} l}{zm^{1/2}} < x < \frac{s}{zq} + \frac{4c^{1/2} \log^{1/2} l}{zm^{1/2}}.
\]

(28)

It is clear from (28) that

\[
zqx - \frac{4c^{1/2} q \log^{1/2} l}{m^{1/2}} \leq s \leq zqx + \frac{4c^{1/2} q \log^{1/2} l}{m^{1/2}}.
\]

Since \( 0 \leq x \leq l \), we have, using (19),

\[
-s_1 \leq -\frac{4c^{1/2} q \log^{1/2} l}{m^{1/2}} \leq s \leq zql + \frac{4c^{1/2} q \log^{1/2} l}{m^{1/2}} < c_qzql = s_1,
\]

and supposing that \( s_1 \in \mathbb{Z} \), we have

\[
D \subseteq \bigcup_{s = -s_1}^{s_1} D_x.
\]

Let us now build maps \( \rho : D \rightarrow \mathbb{Z}^2 \), defining \( \rho(x) = (x, s) \) for \( x \in D_x \), and \( \psi : \mathbb{R}^2 \rightarrow \mathbb{R} \), as in Case A.

Define parallelogram \( H = \{ v \mid v \in \mathbb{R}^2, \ v = x_1 l_1 + x_2 l_2, \ -1 \leq x_1 \leq 1, \ -1 \leq x_2 \leq 1 \} \), where \( l_1 = (4c^{1/2} \log^{1/2} l/zm^{1/2}, 0) \), \( l_2 = (s_1/q, s_1) \) and the lattice \( \Gamma \) with the basis \( u_1 = (q, 0), u_2 = (q - p', 1) \).

Show that \( \psi(H \cap \Gamma) \supseteq D \). Clearly, \( \psi(H) \supseteq [1, l] \supseteq D \). In addition, if \( x \in D \) then \( x \in D_x \) for some \( s, x \equiv -p's \pmod{q} \), i.e. \( x - tq - p's \), \( t \in \mathbb{Z} \), and \( \rho(x) = (x, s) - (tq - p's, t) \in zql^2 \). Thus, \( \psi(\Gamma) \supseteq D \).

In the same way as in Case A, we have

\[
|D| \leq |H \cap \Gamma| \leq \left( \left\lceil \frac{2l_1}{q} \right\rceil + 1 \right) (2s_1 + 1) \leq \frac{l \log^{1/2} l}{m^{1/2}}.
\]

(29)

So, in both cases we have obtained the map \( \rho \) which transforms our problem into a two-dimensional one with increased density (see (16)). To apply the results of [10], we need some additional reasoning.
For each \( a_j \in A_2 \), define \( a'_j = \rho(a_j) \) and \( A' = \{ a'_j \} \). \( A' = \rho(A_2) \subseteq H \cap \Gamma \) since \( A_2 \subseteq D \). Define also \( A'(\Gamma') = \{ a'_j | a'_j \in A' \cap \Gamma', \Gamma' \subseteq \Gamma \} \), \( L_{\Gamma'} = |\Gamma' \cap H| \) and \( K_{\Gamma'} = (c_3/4\varepsilon_3) L_{\Gamma'}^{-1/3} \log^{1/3} L_{\Gamma'} \). Here \( \Gamma' \) is a sublattice of the lattice \( \Gamma \).

Since \( A' = \rho(A_2) \) and \( \rho(D) \subseteq H \cap \Gamma \), we have, in view of (14) and (16) (for sufficiently large \( \varepsilon_3 \)), \( |A'| \geq \varepsilon_8 K_{\Gamma'} \), that expresses the density of \( \rho(A_2) \) in \( \rho(D) \).

Choose now any \( A'_1 \subset A' \) such that \( |A'_1| = \lceil \frac{1}{2} |A'| \rceil \).

We will show that there exists a sublattice \( \Gamma' \subseteq \Gamma \) such that, for every \( \Gamma'' \subset \Gamma' \), the condition

\[
|A'_1(\Gamma'')| < |A'_1(\Gamma')| - K_{\Gamma'} \tag{31}
\]

is satisfied and \( |A'_1(\Gamma')| \geq |A'_1| - 6K_{\Gamma'} \). Suppose that (31) is not satisfied for \( \Gamma' = \Gamma \). We find a sequence \( \Gamma_1 = \Gamma \supset \Gamma_2 \supset \cdots \supset \Gamma_p \) of lattices such that \( |A'_1(\Gamma_i+1)| \geq |A'_1(\Gamma_i)| - K_{\Gamma_i} \), for \( 1 \leq i < p \). Let us use the following lemma.

**Lemma 3.1** (Freiman [10, Lemma 3]). Let \( K \) be a set of integer points of the convex set \( F \subseteq \mathbb{R}^2 \). Suppose that not all points of \( K \) belong to one line and that \( |K| > C \), where \( C \) is a sufficiently large constant. Then, for each \( \Gamma \neq \mathbb{Z}^2 \), either \( K(\Gamma) \leq \frac{1}{2} |K| \) or points of \( K(\Gamma) \) belong to one line.

Applying this lemma with \( H \) and \( \Gamma_i \) instead of \( F \) and \( \mathbb{Z}^2 \), respectively, we have either \( L_{\Gamma_i} \leq \frac{3}{4} L_{\Gamma_{i-1}} \), \( K_{\Gamma_{i-1}} \leq (\frac{3}{4})^{1/3} K_{\Gamma_i} \), or the fact that \( A'_1(\Gamma_i+1) \) is on a line and \( \psi(A'_1(\Gamma_i+1)) \) is an arithmetic progression. The latter possibility will be considered in Section 4.

So, \( \sum_{i=1}^{j} K_{\Gamma_i} \leq K_{\Gamma_j} \left( \frac{1}{2} - \frac{3}{4} / \sqrt[3]{3} \right) \leq 6K_{\Gamma_j} \) for every \( j \). Therefore, \( |A'_1(\Gamma_j)| \geq |A'_1| - 6K_{\Gamma_j} \geq m \). After \( p \leq \log l \) steps, we obtain the lattice \( \Gamma' \), for which (31) is valid. Otherwise, we would obtain, after \( j = \varepsilon_8 \log l \) steps, \( |A'_1(\Gamma_i)| \geq L_{\Gamma_i} \) since \( L_{\Gamma_i} < (\frac{3}{4})^{1} L_{\Gamma_i} \).

Define \( A'_2 = A'_1(\Gamma') \), \( A'_3 = A' \setminus A'_2 \). Now we will use the following theorem [10].

**Notation.** \( M_1 = \frac{1}{2} \sum_{a_j \in A} a_1j, \quad M_2 = \frac{1}{2} \sum_{a_j \in A} a_2j, \quad B_1 = \frac{1}{4} \sum_{a_j \in A} a_1j a_2j, \quad B_2 = \frac{1}{4} \sum_{a_j \in A} a_2j, \quad B = \begin{pmatrix} B_1 & B_2 \\ B_1 & B_2 \end{pmatrix} \).

**Theorem 3.2** (Freiman [10, Corollary 1]). Let \( \Gamma' \) be a lattice and \( H \subseteq \mathbb{R}^2 \) be a convex set which satisfies the conditions \( (0,0) \in H \) and \( |H \cap \Gamma'| = L, \quad L > L_0 \). Assume that \( A \subseteq H \cap \Gamma' \) is a set of vectors for which \( |A| > m_0 = c_9 L^{2/3} (\log L)^{1/3} \) and, for each \( \Gamma'' \subset \Gamma' \), we have \( |A(\Gamma'')| \leq |A| - m_0/c_9 \). Then every vector \( b = (b_1, b_2) \in \Gamma' \) from the domain

\[
(M_1 - b_1, M_2 - b_2) B^{-1} \begin{pmatrix} M_1 - b_1 \\ M_2 - b_2 \end{pmatrix} \leq 1 \tag{32}
\]

belongs to \( A^* \).

Application of Theorem 3.2 to the set \( A'_2 \) is possible in view of (16) and (31). Choose a basis \( e_1, e_2 \) of \( \Gamma' \) as follows. Let \( e_1 = (d_1 q, 0) \) be the nearest to \( (0,0) \) point of axis
Let $x$ belonging to $\Gamma'$ and $e_2 = (t, d_2')$ be the nearest to axis $x$ point of $\Gamma'$. Let $C$ denote the domain defined by (32). Its image $\psi(C)$ consists of short segments of arithmetic progressions. We will use the set $A_2'$ to increase the dimensions of $C$ in such a way that arithmetic progressions will stick together as one long progression. In the same way as before, let us consider the two cases A and B according to the previous choice of map $\rho$.

Case A. First, let us estimate the values of $B_1$ and $B_2$ for the set $A_2'$. For values of the coordinates of the vector, $l_1$ and $l_2$, we have

\[ l_1 \ll l, \quad l_2 \gg \frac{q \log l}{m^{1/2}}. \]

Evidently,

\[ B_2^2 \ll m l^2, \quad B_1 \ll m^{1/2} l \]

and

\[ B_2^2 \ll m \left( \frac{q \log l}{m^{1/2}} \right)^2, \quad B_2 \ll q \log l^{1/2}. \]

Let

\[ R = \sqrt{B_1^2 B_2^2 - B_1^2} = \frac{1}{2} \left( \sum_{a_i, a_j \in \mathcal{A}_2} (a'_1 a'_2 - a'_2 a'_1)^2 \right)^{1/2}. \]

Keeping in view that $\frac{1}{2} |a'_1 a'_2 - a'_2 a'_1| \leq \text{area of a triangle with the vertices } (0, 0), (a'_1, a'_2), \text{ and } (a'_2, a'_1)$ and that all $|\mathcal{A}_2| \gg m$ points of $A_2'$ belong to the lattice $\Gamma'$, with the area of the fundamental parallelogram $qd'_1 d'_2$, we see that

\[ R \gg m^2 q d'_1 d'_2. \]

To obtain an estimate of $B_1$ from below, we have to take $p = l_2/d_2 = q \log l^{1/2}/m^{1/2} d'_2$ numbers $a'_j \in \Gamma'$, with $a'_1$ equal to 0, the same number of $a'_j$, equal to $qd'_1, 2qd'_1, \ldots, \text{ etc.}$ Clearly, $m/p$ is not small since $q \leq Q \ll (l/\log l)^{1/2}$ and we obtain

\[ B_2^2 \gg pd'_1^2 q^2 \left( \frac{m}{p} \right)^3 = m^3 d'_1^2 q^2 p^2, \quad B_1 \gg \frac{m^2 d'_1 d'_2}{(\log l)^{1/2}} \gg l_1 d'_1 d'_2 \log^{1/2} l. \]

To estimate $B_2$, recall that the number of elements not divisible by $q$ is large (see (21)). We have

\[ B_2^2 \gg m d'_2^2, \quad B_2 \gg m^{1/2} d'_2 \gg \frac{q \log l}{m^{1/2}} \gg l_2 \log^{1/2} l, \]

in view of (2), (18) and since $q \ll Q < (l/\log l)^{1/2}$.

Evaluate now the size of the area defined by (32). The values of $b_1$ when $b_2$ is equal to $M_2$ satisfy

\[ B_2^2 (M_1 - b_1)^2 \leq R^2, \quad |M_1 - b_1| \leq \frac{R}{B_2}. \]
Let us study now the points $F=(M_1 + \lambda B_1, M_2 + \lambda B_2)$. From (32) we obtain $(2\lambda^2 B_1 B_2/R^2) (B_1 B_2 - B_{12}) < 1$ and, for $|\lambda| < 1/\sqrt{2}$, this inequality is valid. Thus, vectors $b=(b_1, b_2)$ satisfying
\[ b_2 = M_2 + \lambda B_2, \quad |M_1 + \lambda B_1 - b_1| \leq \frac{R}{2B_2} \] (39)
for $|\lambda| \leq 1/2\sqrt{2}$ are in $(A'_2)^*$, in view of (38).

Let $A' = \{g_1, g_2, \ldots, g_w\}$ and $A'_e = \{g_1, \ldots, g_e\}$. Denote the set of all different residues modulo $\Gamma'$ of $(A'_3,)^*$ by $G$. Now construct the set $G = \{g_j | |G| > |G_i-|\}$. (This construction for a one-dimensional case was described in detail in [9], pp. 10–11). As a result, we obtain a set $G$ and a lattice $\Gamma''$, for which
\[ \Gamma' \subseteq \Gamma'' \subseteq \Gamma, \quad A'_3 \setminus G \subseteq \Gamma'' \quad \Gamma'' \subseteq \Gamma' + G*. \] (40)

Clearly, $|G| < d_1^* d_2^* \leq 1/qm = o(m)$ and $|A'_3 \setminus G| \geq m$.

According to Theorem 3.2, subset sums of $A'_2$ represent all points of $\Gamma'$ in the parallelogram with edges $(R/B_2, 0)$ and $(1/\sqrt{2})(B_1, B_2)$ defined by (39). Now, after adding $\leq d_1^* d_2^*$ vectors of $G$ to $A'_2$, we obtain that $(A'_2 + G)^*$ contains all points of $\Gamma''$ in a similar parallelogram. In fact, adding the vectors of $G^*$ moves the parallelogram not more than $l_2 d_1^* d_2^* = o(R/B_1)$ along the vertical direction and not more than $l_1 d_1^* d_2^* = o(R/B_2)$ along the horizontal direction, in view of (34)–(36). Thus, the order of the size of the intersection $P$ of the initial parallelogram and the moved one is the same as in (39).

So, we have obtained a new lattice $\Gamma'' \subseteq \Gamma$, all points of which in the parallelogram $P$ belong to the set of subset sums. Choose now the basis of lattice $\Gamma''$ in the same way as the basis of $\Gamma'$ was chosen: $e_1 = (qd_1^*, 0)$, $e_2 = (t, d_2^*)$. All vectors of $A'_3 \setminus G$ have integer coordinates in this basis. Now we will take coordinates of all vectors in this new basis. One has to take this into account regarding the values $M_1, B_1, \ldots$ etc.

Now we need the following lemma.

**Lemma 3.3.** Let $a_1 = (a_{11}, a_{12})$, $a_2 = (a_{12}, a_{22}) \in \mathbb{Z}^2$, and $A = \{(x, b) | x \in \mathbb{Z}, a \leq x < a + g\}$. So, $A$ is a segment of integer points of length $g$ (length here means the number of integer points). Let $a_{21} = a_{22}$ and
\[ |a_{12} - a_{11}| = \Delta < g. \] (41)
Then $(A + a_1) \cup (A + a_2)$ is the segment of the length $g + \Delta$.

**Proof.** We can suppose that $a_{12} > a_{11}$.
\[
(A + a_1) \cup (A + a_2) = a_1 + AU(A + a_2 - a_1)
= a_1 + AU(A + (a_{12} - a_{11}, 0))
= a_1 + \{(x, b) | x \in \mathbb{Z}, a \leq x < a + g\}
\cup \{(x, b) | x \in \mathbb{Z}, a + \Delta \leq x < a + \Delta + g\}
= a_1 + \{(x, b) | x \in \mathbb{Z}, a \leq x < a + \Delta + g\}.
\]
Having in view the application of this lemma, take the partition of \( A_3 \setminus G \) into subsets \( F_1, F_2, \ldots, F_s \) according to a relation \( a_2 = a_2' \) of equivalence for two vectors \( a_1, a_2 \in A_3 \setminus G \), i.e. into subsets lying on one horizontal. Denote by \( m_1, m_2, \ldots, m_t \) the numbers of elements in each of these subsets.

We may suppose that \( m_1 \geq m_2 \geq \cdots \geq m_t \). Let us take the maximal value of \( s' \) for which \( m_i > c_{10} m/s' \log l \), where \( c_{10} \) could be chosen sufficiently small. Actually, such \( s' \) exists. In contrast, we would have \( \sum_{i=1}^{s'} m_i \leq (c_{10} m \log l) \sum_{i=1}^{s'} (1/i) \leq 2c_{10} m < |A_3 \setminus G| \), a contradiction. Since \( \sum_{i>s'} m_i \leq (c_{10} m \log l) \sum_{i=1}^{s'} (1/i) \leq 2c_{10} m \),

\[
\sum_{i<s'} m_i \geq m. 
\] (42)

From each subset of \( m_i \) numbers, \( 1 \leq i \leq s' \), we can choose \( m_i/4 \) disjoint pairs \( a_1', a_2' \) of vectors from \( A_3 \setminus G \) with \( A > m_i/3 \). So, in view of (42), we can find \( t \geq \frac{1}{s'} \sum_{i=1}^{s'} m_i \geq m \) pairs with

\[
\sum A \geq \sum_{i=1}^{s'} m_i^2 \geq \frac{(\sum m_i)^2}{s'} \geq m^2/s'. 
\] (43)

From (33), in our basis we have \( s' \leq l_{22}/d_2 \leq q(\log l)^{1/2}/m^{1/2} d_2 \) and, so, \( m^2/s' \geq m^{5/2} d_2/q(\log l)^{1/2} \).

Apply Lemma 3.3 to one of the pairs \( a_1^{(j)}, a_2^{(j)} \), \( 1 \leq j \leq t \), mentioned above and the segments defined by (39) for each value of \( b_2 \). Verifying condition (41), we have, from (35) and (36), \( A = |a_2' - a_1'| < l/q d_1 \leq m^2 d_2/2q(\log l)^{1/2} = o(R/B_2) \).

Applying Lemma 3.3, we obtain new segments with length greater by \( A \) than before. So, applying it \( t \) times, we have, in view of (43), a set of segments with lengths \( m^{5/2} d_2/q(\log l)^{1/2} \) belonging to \( (A')^s \). At each step, parallelogram \( P \) is moved and becomes longer. As a result, we obtain a long, but narrow parallelogram.

From (34) and (2) we have \( B_1 \leq m^{1/2} l/q d_1 = o(m^{5/2} d_2/q(\log l)^{1/2}) \); so, we can take our segments with equal beginnings (throwing out \( \ll B_1 \) points in each segment), equal ends and the same (in order) length. We have obtained a rectangle \( P' \) of length given by (43) and height \( \approx B_2 \).

Now we will use the following lemma.

**Lemma 3.4.** Let \( a_1 = (a_{11}, a_{21}), a_2 = (a_{12}, a_{22}) \in \mathbb{Z}^2 \), and \( A = \{(x, y) | a \leq x < a + g_1, b \leq y < b + g_2 \} \), i.e. \( A \) is a rectangle with sides of length \( g_1 \) and \( g_2 \). Let \( |a_{12} - a_{11}| = A_1 < g_1 \) and \( |a_{22} - a_{21}| = A_2 < g_2 \). Then \( (A + a_1) \cup (A + a_2) \) contains a rectangle of integer points with sides \( g_1 - A_1, g_2 - A_2 \).

**Proof.** Consider, for example, the case \( a_{12} > a_{11} \) and \( a_{22} > a_{21} \). We have

\[
(A + a_1) \cup (A + a_2) = a_1 + A \cup (A + a_2 - a_1) = a_1 + \{(x, y)\},
\]

where \( a + a_{12} - a_{11} \leq x \leq a + g_1 - 1 \), \( b \leq y \leq b + g_2 - 1 + a_{22} - a_{21} \). \( \square \)
We will use Lemma 3.4 to enlarge the height of $P'$. Applying Lemma 3.3, we have used no more than half the vectors from subsets $F_1, \ldots, F_s$. Now we will use the rest of these vectors.

Let us choose pairs $a_i \in F_i, a_j \in F_j, 1 \leq i, j \leq s'$ in such a way that $A_2 \leq l_{22} = q(\log l)^{1/2}m^{1/2}d_2 = o(B_2)$. Since we have at least $c_{10}m/2s'\log l$ vectors in each subset, we can choose $c_{10}m/\log l$ such pairs. Show that $k = 2qd_1/s' < c_{10}m/5\log l$ (applying Lemma 3.4, one can use not more than $k$ pairs $a_i, a_j$ in order to keep the order of length of rectangle $P'$).

In view of (42), there exists $i$ for which $m_i > m/s'$. These $m_i$ points belong to a segment with a length not more than $l/qd_1$. Hence, we have $m/s' \leq l/qd_1$, which provides $2qd_1/s' \leq 2l/m < 2m/c_1\log l < c_{10}m/\log l$ for sufficiently large $c_1$.

Using Lemma 3.4 for $P'$ and $k = 2qd_1/s'$ pairs mentioned above, we obtain a new rectangle $P''$. Its length is $\gg m^2/s'$, since the length of $P'$ is diminishing not more than $k \cdot \max A_1 \leq (2qd_1/s')(l/qd_1) = o(m^2/s')$. On the other hand, the height of $P''$ is larger than $k \cdot \min A_2 \geq (2qd_1/s')(s'/2) = qd_1$.

To complete the consideration of case A, we analyze the set $\psi(P'')$. Let $P''_i, 0 \leq i < qd_1$, be $i$'s horizontal of $P''$. Clearly, $\psi(P''_i)$ are arithmetic progressions with difference $qd_1$. The beginnings of these progressions are

$$r_0, r_0 + d_2p', r_0 + 2d_2p', \ldots, r_0 + (qd_1 - 1)d_2p',$$

where $r_0$ is the beginning of $P''$. It may be assumed that $\gcd(d_4, d_2p') = 1$ (the opposite implies that all elements of $\psi(A_2)$ are divisible by some integer $d > 1$, but we agreed (see (21)) that this possibility will be analyzed in Section 4). Taking this into account, we see that the sequence (44) is the full system or residues modulo $qd_1$. The difference between the beginnings of the progressions, in view of (44), is less than their length $m^2/s'$, namely, $d_2p' \cdot qd_1 = o((m^2/s')qd_1)$ since $s' < l_{22}/d_2 \leq q(\log l)^{1/2}/m^{1/2}d_2$ and $q \leq (l/\log l)^{1/2}$.

We have shown that the union of progressions $\psi(P''_i)$ contains a segment of length $= (m^2/s')qd_1 \gg m^{2/2}/\log^{1/2} l$.

**Case B:** We will analyze this case in the same way as Case A.

Let $e_1 = (qd_1, 0)$ and $e_2 = (t, d_2)$ be the basis of lattice $\Gamma'$. Since domain $C$, defined by (32), does not depend on the choice of a basis (see [10]), we use the basis $\{(1, 0), (l_{12}/l_{22}, 1)\}$ for convenience in our estimation. The fact that integer points in the new basis will not be integer is not essential (we will return to the previous basis when necessary). Let $\overline{B}_1, \overline{B}_2, \overline{B}_{12}$ and $\overline{R}$ correspond to a new basis. The size of $H$, according to the new basis, is characterized by

$$l_{11} = \frac{(\log l)^{1/2}}{2m^{1/2}}, \quad l_{12} = l_{22} = l_{22} = l_{22}.$$
In the same way as in case A, we obtain the estimates

\[
\bar{B}_1 \lesssim \frac{\log^{1/2} l}{zq},
\]

\[
\bar{B}_2 \lesssim m^{1/2} l zq,
\]

\[
\bar{B}_1 \gtrsim \frac{m^{3/2} d'_1 d'_2}{lz},
\]

\[
\bar{B}_2 \gtrsim \frac{m^2 z q d'_1 d'_2}{\log^{1/2} l} \quad \text{for } zq \gtrsim \frac{\log^{1/2} l}{m^{3/2} d'_1},
\]

\[
\bar{B}_2 \gtrsim m^{1/2} d'_2 \quad \text{for } zq < \frac{\log^{1/2} l}{m^{3/2} d'_1},
\]

\[
\bar{R} \gtrsim m^2 q d'_1 d'_2,
\]

and the parallelogram

\[
|\lambda| \lesssim \frac{1}{2\sqrt{2}}, \quad b_2 = M_2 + \lambda \bar{B}_2, \quad |M_1 + \lambda \bar{B}_1 - b_1| \lesssim \frac{\bar{R}}{2\bar{B}_2},
\]

which belongs to the domain defined by (32), the set \(G\) and the lattice \(\Gamma''\), for which (40) holds.

Verifying the possibility of applying Lemma 3.3, we have \(\Delta \lesssim l_{11} = o(\bar{R}/\bar{B}_2)\) (using estimates for \(l_{11}, \bar{R}\) and \(\bar{B}_2\)), which confirms (41). So, taking \(m_i/4\) pairs of vectors with \(d \geq m_i/5\) from \(F_i, t = 1, 2, \ldots, s' < l z q/d_2\), and applying Lemma 3.3, we obtain the parallelogram \(P'\) with length \(\approx m^2/s' \approx m^2 d_2/\|zq\|\).

Now we apply Lemma 3.4 \(k = 2 q d_1/s'\) times. Using an inequality \(k \max \Delta_1 < (2 q d_1/s') \cdot (l_{11}/q d_1) = 2 l_{11}/s'\) and remembering that \(l_{11} \approx \log^{1/2} l / z m^{1/2}\), we have to verify that

\[
\left(\frac{\log l}{z m^{1/2}}\right)^{1/2} = o(m^2)
\]

and this is true in view of (2) and (18).

4. Proof of Theorem 1.1

In previous sections we have analyzed the main cases of the theorem and the segment with length \(\gg m^2\) has been obtained. It remains to study two special situations: (a) for the set \(A_i\) we have \(B < l \log^{1/2} l\), and (b) there exists a number \(d\) which divides at least \((1 - 1/c_5)|A|\) elements of \(A_i\).

The situation (a) implies that less than \(16 \log l\) numbers of segment \([\frac{1}{2} l, l]\) belong to \(A_i\). Taking the sequence of sets \(A_{10} = A_1, A_{11}, A_{12}, \ldots\) for which \(A_{11} = \{a | a \in A_1, \ a \in [1, (1/2^2) l]\}\) and \(|A_{11}| > |A_1| - 16 l \log l\), we obtain, after \(p < \log(l/m)\) steps, the subset \(A_{1p}\), with \(|A_{1p}| \gg m\), and for this subset condition (5) holds.
In situation (b) we construct another sequence of sets, $A_{10} = A_1, A_{11}, \ldots$, for which
$A_{11} = A_{10} \setminus d_i = \{ a | a d_i \in A_{10} \setminus d_i, a \in \mathbb{Z} \}$, where $d_i > 1$ is the integer for which
$|A_{11}| > |A_{10} \setminus d_i| - (l_i/\log l_i)^{1/2}$, where $l_i = \max \{ a | a \in A_{11} \}$. Clearly, $l_i \leq l_{i-1}/d_i$ and
$|A_{11}| > |A_{10}| - \sum_{i=1}^{m} (l_i/\log l_i)^{1/2} > |A_{10}| - 4(l/\log l)^{1/2} > m$. So, for some $p < \log l$, we have
the set $A_{1p}$, elements of which are 'well distributed' among the residues of all integers. This fact confirms all assumptions of Section 3 for the set $A_{1p}$.

The above-mentioned reasoning allows us to obtain the set $A_{1p}$ and a long arithmetic progression with difference $d = d_1 \cdots d_p$, the length of which can be increased by the other elements of $A$ in the same way as was done in Section 2. Clearly, $d \leq 3l/m$ because the number of integers divisible by $d$ is bounded by $\lceil l/d \rceil$ and we have $|A_{1p}| > \frac{1}{3}m$ such integers.

5. Applications and concluding remarks

We have proved that the set of subset-sums $A^*$ contains an arithmetic progression under a rather wide range of conditions. This fact enables us to solve different problems connected with the structure of $A^*$.

For example, let $p(n, r)$, for $r \geq 2$, denote the maximal cardinality of a subset $A \subset \{ 1, 2, \ldots, n \}$ such that there are no $r$-powers of an integer in $A^*$. In [1, 11], it was proved that, for $r \geq 10$ [11] and for $r \geq 6$ [1],

$$p(n, r) = (1 + o(1)) 2^{1/r + 1} n^{(r-1)/(r+1)}.$$  \hspace{1cm} (45)

From Theorem 1.1 we obtain that (45) is true for $r \geq 4$.

In [8, 9] it was shown how the fact of the existence of one arithmetic progression leads to a full description of the structure of $A^*$ and gives a possibility of obtaining algorithms for solving dense SSP ($m \gg l^{2/3} \log^{1/3} l$). It is clear that the new result proving the existence of an arithmetic progression for $m \gg l^{1/2} \log^{1/2} l$ gives the main tool to solve SSP for a wider range of problems. A description of the algorithm, time bound and computational experience will be given in forthcoming papers.

The exponent $\frac{1}{2}$ in condition (2) is the best one to obtain a long arithmetic progression. If we take $m < l^\alpha$, $\alpha < \frac{1}{2}$, then there will be cases where a long arithmetic progression in $A^*$ does not exist. Example: For $0 < \alpha < \frac{1}{2}$, let

$$A = \{ x | x = a + b \lfloor l^{1/\alpha} \rfloor, 0 \geq a, b < \lfloor l^{1/2} \rfloor, a^2 + b^2 \neq 0 \}.$$  

For this set the set $A^*$ has a 'two-dimensional' structure.

The idea to use maps $D \rightarrow \mathbb{R}^n$, where $D \subset \mathbb{R}$, was developed in [6]. An improved version of the main theorem and the method is given in [7]. In this paper the map $D \rightarrow \mathbb{R}^2$ is used. With the use of $D \rightarrow \mathbb{R}^n$, maps of the sets on a line into the spaces of larger dimensions, we hope to obtain a solution of SSP for $m > l^\varepsilon$ for each $\varepsilon > 0$.

Note. During the Israel-French conference on combinatorics (November 1988, Jerusalem), Professor Paul Erdős told me about similar results of Professor András.