Astérisque

MARK CHAIMOVICH

New algorithm for dense subset-sum problem

Astérisque, tome 258 (1999), p. 363-373
<http://www.numdam.org/item?id=AST_1999__258__363_0>

Astérisque **258**, 1999, p. 363–373

NEW ALGORITHM FOR DENSE SUBSET-SUM PROBLEM

by

Mark Chaimovich

Abstract. — A new algorithm for the dense subset-sum problem is derived by using the structural characterization of the set of subset-sums obtained by analytical methods of additive number theory. The algorithm works for a large number of summands (m) with values that are bounded from above. The boundary (ℓ) moderately depends on m. The new algorithm has $O(m^{7/4}/\log^{3/4} m)$ time boundary that is faster than the previously known algorithms the best of which yields $O(m^2/\log^2 m)$.

1. Introduction

Consider the following subset-sum problem (see [**13**]). Let $A = \{a_1, \ldots, a_m\}$, $a_i \in \mathbb{N}$. For $B \subseteq A$, let $S_B = \sum_{a_i \in B} a_i$ and let $A^* = \{S_B \mid B \subseteq A\}$. The problem is to find the maximal subset-sum $S^* \in A^*$ satisfying $S^* \leq M$ for a given target number $M \in \mathbb{N}$.

Although the problem is NP-hard (the partition problem is easily reduced to the SSP), its restriction can be solved in polynomial time. Denote $\ell = \max\{a_i \mid a_i \in A\}$. Introducing restriction $\ell \leq m^\alpha$ where α is some positive real number (or equivalently $m \geq \ell^{1/\alpha}$), one can easily solve problems from this restricted class in $O(m^2\ell)$ time using dynamic programming.

This work belongs to the school of thought that applies analytical methods of number theory to integer programming (see [**8**], [**2**]). It continues the application of a new approach, the main idea of which is as follows: analytical methods enable us to effectively characterize the set A^* of subset-sums as a collection of arithmetic progressions with a common difference (see [**7**], [**12**], [**1**], [**10**]). Once this characterization is obtained, it is quite easy to find the largest element of A^* that is not greater than the given M.

Efficient algorithms have recently been derived using the new approach. In almost linear time (with respect to the number m of summands) they solve the following class

---

1991 ***Mathematics Subject Classification.*** — Primary: 90C10 Alternate: 05A17, 11B25, 68Q25.
Key words and phrases. — Analytical Number Theory, Integer Programming, Subset Sum Problem.

of SSP: the target number M is within a wide range of the mid-point of the interval $[0, S_A]$ and $m > c\ell^{2/3}\log^{1/3}\ell$, $\ell > \ell_0$ when A is a set of distinct summands ([9], [4], [6], [11]) or $m > 6\ell\log\ell$ when A is an arbitrary multi-set without any limitation on the number of distinct summands ([5]). Here and further on $\ell_0, c, c_1, c_2, \ldots$ denote some absolute positive constants.

The latest analytical result ([10]) allows one to apply the algorithm from [9] to problems with density $m > c_1(\ell\log\ell)^{1/2}$. The algorithm from [11] works for density $m > c_2\ell^{1/2}\log\ell$ which is almost the same as in [10]. For $m < \ell^{2/3}$, the time boundary for both algorithms is estimated as $O((\frac{\ell}{m})^2)$, i.e., $O(\frac{m^2}{\log^2 m})$ for the lowest density $(m \sim (\ell\log\ell)^{1/2})$.

This work refines the structural characterization of the set of subset-sums which allows us to use more efficient conditions in the process of determining the structure. These refinements are discussed in Section 2. They lead to the development of a new algorithm which is described in Section 3. It works in $O(m\log m + \min\{\frac{\ell^{5/4}\log^{1/2}\ell}{m^{3/4}}, (\frac{\ell}{m})^2\})$ time which improves [9] and [11] for $m \le \frac{\ell^{3/5}}{\log^{2/5}\ell}$ and yields $O(m^{7/4}/\log^{3/4} m)$ time for $m \sim (\ell\log\ell)^{1/2}$.

2. Refinement of the structural characterization of the set A^* of subset-sums

The following Theorem 2.1 [10] determines the structure of the set A^* of subset-sums for $m > c_1(\ell\log\ell)^{1/2}$ as a long segment of an arithmetic progression.

***Theorem 2.1* (G. Freiman).** — *Let $A = \{a_1, \ldots, a_m\}$ be a set of m integers taken from the segment $[1, \ell]$. Assume that $m > c_1(\ell\log\ell)^{1/2}$ and $\ell > \ell_0$.*
(i) There is an integer d, $1 \le d \le \frac{3\ell}{m}$, such that

$$|A(0,d)| > m - d \tag{1}$$

and

$$\{M : M \equiv 0(\bmod d), |M - \tfrac{1}{2}S_{A(0,d)}| \le c_2dm^2\} \subseteq A^*(0,d),$$

where $A(s,t) = \{a : a \equiv s(\bmod t), a \in A\}$.
(ii) If for all prime numbers p, $2 \le p \le \frac{3\ell}{m}$,

$$|A(0,p)| \le m - \frac{3\ell}{m}, \tag{2}$$

then the assertion (i) of the Theorem holds true with $d = 1$.

Simple consideration shows that verification of condition (2) is crucial for the structural characterization of a set A^* of subset-sums. Algorithms from [9] and [11] use this condition directly ([9]) or indirectly ([11]). Our intention is to replace condition (2) by a condition (or a set of conditions), verification of which is easier in the sense that the number of required operations is smaller. To do this we introduce the notion of *d-full set.* We say that set A is d-full if A^* contains all classes of residues modulo d, i.e., in other words, $A^*(\bmod d) = \{0, 1, \ldots, d-1\}$.

Let us study some properties of d-full sets.

Define $S_{r(\mathrm{mod}\, d)} = \min\{s \in A^*, s \equiv r(\mathrm{mod}\, d)\}$.

Lemma 2.2. — *Let A be a set of integers taken from the segment $[1, \ell]$. Suppose that A is d-full. Then for each r, $0 < r < d$,*

$$S_{r(\mathrm{mod}\, d)} \leq d\ell. \tag{3}$$

Proof. — Assume that for some r condition (3) is not true, i.e., $S_{r(\mathrm{mod}\, d)} > d\ell$. This means that $S_{r(\mathrm{mod}\, d)} = a_{i_1} + a_{i_2} + \cdots + a_{i_k}$ for some $k > d$. Consider the sequence of subset-sums $T_s = \sum_{j=1}^{s} a_{i_j}$, $1 \leq s \leq k$. Obviously, at least two of these sums (assume T_s and T_q, $s < q$) belong to the same residue class modulo d (since $k > d$). Then $T_q - T_s \equiv 0(\mathrm{mod}\, d)$ and subset-sum $T_k - (T_q - T_s) = a_{i_1} + \cdots + a_{i_s} + a_{i_{q+1}} + \cdots + a_{i_k} \equiv r(\mathrm{mod}\, d)$ and this subset-sum is smaller than $S_{r(\mathrm{mod}\, d)}$. This fact contradicts the minimality of $S_{r(\mathrm{mod}\, d)}$. □

Lemma 2.3. — *Suppose that the set A is d-full. Then there is a d-full subset of A with cardinality less than d.*

Proof. — Let us assume that contrary to the Lemma the smallest d-full subset of A has more than $d-1$ elements. Denote this subset by $A' = \{a_1, \ldots, a_k\}$. In fact, $d \not| a_i$ for all i's.

Let B be the multi-set of non-zero residues modulo d in A', that is B is composed with $|A'(i,d)|$ times i for any $1 \leq i < d$. Naturally one has $B^* = (A')^*(\mathrm{mod}\, d)$. Then, as a multi-set, $|B| = \sum_{i=1}^{d-1} |A'(i,d)| \geq d$, by the assumption.

Define a sequence of multi-sets $B_0, B_1, \ldots, B_k$ as follows: B_0 is an empty set and $B_i = \{b_1, \ldots, b_i\}$ for $i > 0$. Note that $0 \in B_i^*$ (since it is the sum of an empty subset), and that

$$B_i^* = B_{i-1}^* + \{0, b_i\} = B_{i-1}^* \cup (B_{i-1}^* + b_i), 1 \leq i \leq k. \tag{4}$$

Thus, obviously, $|B_{i-1}^*| \leq |B_i^*|$.

Taking into account that $|B_0^*| = 1$ and that $|B| = k \geq d$, for some i we have $|B_{i-1}^*| = |B_i^*|$ implying that residue b_i (and element a_i respectively) does not add new residue classes, i.e., $(B \setminus b_i)^* = B^*$. Therefore, $A' \setminus a_i$ is d-full as well as A'. This fact contradicts the assumption that A' is the smallest d-full subset of A and proves the Lemma. □

The next lemma refines the second assertion (ii) of Theorem 2.1.

Lemma 2.4. — *Let A be a set of integers taken from the segment $[1, \ell]$. Assume that $|A| = m > c_1(\ell \log \ell)^{1/2}$, $\ell > \ell_0$, and suppose that A is q-full for each q, $2 \leq q \leq \frac{3\ell}{m}$. Then the assertion (i) of Theorem 2.1 holds with $d = 1$.*

Proof. — Assume that $d > 1$ in Theorem 2.1. By the theorem, a long segment of an arithmetic progression belongs to $A^*(0,d)$. On the other hand, A is d-full (since $d \leq \frac{3\ell}{m}$) and subset-sum $S_{r(\mathrm{mod}\, d)}$ exists for each r, $1 \leq r < d$. Combine a long segment of an arithmetic progression (with difference d) in interval

$$[\tfrac{1}{2} S_{A(0,d)} - c_2 dm^2, \tfrac{1}{2} S_{A(0,d)} + c_2 dm^2]$$

(belonging to $A^*(0,d)$) with subset-sums $S_{1(\text{mod}\, d)}, S_{2(\text{mod}\, d)}, \ldots, S_{d-1(\text{mod}\, d)}$ (these subset-sums are obtained without using elements of $A(0,d)$). Thus we obtain an interval

$$[\tfrac{1}{2}S_{A(0,d)} - c_2 dm^2 + \max\{S_{r(\text{mod}\, d)} : 1 \leq r < d\}, \tfrac{1}{2}S_{A(0,d)} + c_2 dm^2],$$

all integers of which belong to A^*. In fact, if the length of this new interval is sufficiently large ($O(m^2)$, for example), we will obtain the result of Theorem 2.1 with $d' = 1$. Actually, since we are interested only in the case $d > 1$ and since $\max\{S_{r(\text{mod}\, d)} : 1 \leq r < d\} < d\ell = O(dm^2/\log m)$, the length of the obtained interval is

$$O(dm^2 - \max\{S_{r(\text{mod}\, d)} : 1 \leq r < d\}) = O(dm^2 - \frac{dm^2}{\log m}) = O(dm^2)$$

which completes the proof. □

The latest property (Lemma 2.4) shows that in order to obtain a structural characterization of A^*, it is sufficient to verify that set A is q-full for all q's, $2 \leq q \leq \frac{3\ell}{m}$. Clearly, the new condition is weaker than (2): A can be q-full even if $|A(0,q)| > m - \frac{3\ell}{m}$. However, from an algorithmic point of view this new condition is difficult to verify. To correct this we have to use some lemmas which determine different sufficient conditions implying that set A is q-full. We will also show that it is sufficient to verify the prime numbers only.

Lemma 2.5 (**[3]**). — *If p is prime and*

$$\sum_{i=1}^{p-1} |A(i,p)| \geq p - 1 \tag{5}$$

then A is p-full.

The proof of this lemma is presented here because of the difficulty in accessing of reference [**3**].

Proof. — Using the fact that all elements of $A(i,p), i \neq 0$, are relatively prime to p, introduce ring $\mathbb{Z}_p$ of residues mod p. In the following reasoning it is implied that all arithmetic operations, including the operations for computing subset-sums, are operations modulo p in $\mathbb{Z}_p$.

Put, as in the proof of Lemma 2.3, $B = \{b_1, b_2, \ldots, b_k\}$ for the multi-set of non-zero residues modulo p in A and define the sequence of multi-sets $B_0, B_1, \ldots, B_k$ where B_0 is an empty set and $B_i = \{b_1, \ldots, b_i\}$ for $i > 0$.

By the hypothesis, $|B| = \sum_{i=1}^{p-1} |A(i,p)| \geq p - 1$. If for all $i \leq p-1, |B^*_{i-1}| < |B^*_i|$, then $|B^*_i| \geq |B^*_{i-1}| + 1 \geq |B^*_0| + i = i + 1$, i.e., $|B^*_{p-1}| \geq p$, which concludes the proof, since we are dealing with residues modulo p.

Otherwise, the fact that $|B^*_{i-1}| = |B^*_i|$ for some $i < p - 1$ implies that for any $c \in B^*_{i-1}$, $c + b_i$ also belongs to B^*_{i-1}. Continuing this reasoning we obtain $c + rb_i \in B^*_{i-1} \subseteq B^*$ for any r. Recalling that all operations are modulo p and that $\gcd(b_i, p) = 1$, one obtains that all residues modulo p are in B^*, i.e., A is p-full. □

***Lemma 2.6* (Olson [14]).** — *If p is prime and*

$$|\{i : |A(i,p)| \neq 0, 1 \leq i < p\}| > 2p^{1/2} \tag{6}$$

then A is p-full.

***Lemma 2.7* (Theorem 7, Sárkőzy [15]).** — *If p is prime and*

$$(\sum_{i=1}^{p-1} |A(i,p)|)^3 \geq c_5 p \log p \sum_{i=1}^{p-1} |A(i,p)|^2 \tag{7}$$

where $c_5 = 4 \cdot 10^6$, then A is p-full.

Note that condition (7) implies $\sum_{i=1}^{p-1} |A(i,p)| \geq (c_5 p \log p)^{1/2}$ in view of

$$\sum_{i=1}^{p-1} |A(i,p)| \leq \sum_{i=1}^{p-1} |A(i,p)|^2.$$

The next two lemmas show that it is sufficient to verify the prime numbers only.

Lemma 2.8. — *If for prime numbers p, $2 \leq p \leq Q^{1/2}$,*

$$|A(0,p)| \leq m - Q, \tag{8}$$

and for prime numbers p, $Q^{1/2} < p \leq Q$, the set A is p-full, then the set A is t-full for all integers t, $2 \leq t \leq Q$.

Proof. — The proof employs induction for the total number of prime divisors of t.

1. t is prime. Condition (8) ensures that Lemma 2.5 can be applied to all prime numbers $t \leq Q^{1/2}$. For prime numbers $t > Q^{1/2}$, the set A is t-full by definition.
2. For $n > 1$, assume that the Lemma is true for each number whose total number of prime divisors is less than n. Now we are going to prove the Lemma for any integer t having n prime divisors.

   Let $t = p_1 \cdots p_n$ where $p_1 \leq p_2 \leq \cdots \leq p_n$ are the prime divisors of t. One has $p_1 \leq t^{1/2} \leq Q^{1/2}$ and, in view of (8), $|B| = |A \setminus A(0,t)| \geq |A \setminus A(0,p_1)| \geq Q \geq t$.

   Denote $s = t/p_1$. This integer s has $n-1$ prime divisors. By the induction hypothesis, A is s-full. Thus, according to Lemma 2.3, there is $A' \subseteq A$ such that A' is s-full and $|A'| < s$. Put, as in the proof of Lemma 2.5, $B = \{b_1, b_2, \ldots, b_k\}$ for the multi-set of non-zero residues modulo t in A and define $B_i = \{b_1, \ldots, b_i\}$. Without losing generality, assume that the first residues in B corresponds to elements of A'. Thus, $B^*_{|A'|}$ contains all classes of residue modulo s implying $|B^*_{|A'|}| \geq s$. Continue with the same reasoning as in Lemma 2.5.

   Again, if for all $i, |A'| < i \leq t-1, |B^*_{i-1}| < |B^*_i|$, then $|B^*_i| \geq |B^*_{i-1}| + 1 \geq |B^*_{|A'|}| + (i - |A'|) \geq i+1$, i.e., $|B^*_{t-1}| \geq t$, which concludes the proof, since we are dealing with residues modulo t.

   Otherwise, the fact that $|B^*_{i-1}| = |B^*_i|$ for some i, $|A'| < i \leq t-1$ implies that for any $c \in B^*_{i-1}$, $c + b_i \in B^*_{i-1}$. Continuing this reasoning we obtain $c + rb_i \in B^*_{i-1} \subseteq B^*$ for any r. Recalling that $B^*_{|A'|}$ contains $c_1, \ldots, c_s$ - different residues modulo s - we generate s disjoint sequences $c_j + rb_i$. Since each sequence has $r = \frac{t}{s}$ elements modulo t, all sequences together cover the entire set of residues modulo t, i.e., A is t-full.

This concludes the proof that the set A is t-full for all $t \leq Q$. □

Now we can formulate a sufficient condition for a long interval to exist in the set A^* of subset-sums:

Corollary 2.9. — *Let A be a set of integers taken from the segment $[1, \ell]$. Assume that $|A| = m > c_1(\ell \log \ell)^{1/2}$, $\ell > \ell_0$, and suppose that for all primes p, $2 \leq p \leq (\frac{3\ell}{m})^{1/2}$, condition (2) holds and for all primes p, $(\frac{3\ell}{m})^{1/2} < p \leq \frac{3\ell}{m}$, at least one of the conditions (5), (6) or (7) is satisfied. Then A^* contains a long interval: a segment of an arithmetic progression with difference 1 and length $O(m^2)$.*

Proof. — The corollary follows from previously mentioned Lemmas 2.4, 2.5, 2.6, 2.7 and 2.8. □

3. Algorithm

In the previous section we determined a sufficient condition, ensuring the existence of a long interval contained in A^*. In the case where this condition is not satisfied, namely, if for some p_1 either condition (2) (if p_1 is small) or conditions (5), (6) and (7) (if p_1 is large) fail, the process similar to the process described in **[9]** may be applied. This process finds a number d such that an arithmetic progression with difference d belongs to the set of subset-sums. It is implemented in the first step of the algorithm. The second step of the algorithm finds all non-zero residues modulo this d in A^* by using a modification of dynamic programming approach modulo d.

Now we are ready to describe the algorithm.

Notation. — $n_p(i)$, $0 \leq i < p$: the counter of summands belonging to residue class i $\bmod p$ (when all summands of A are verified $n_p(i) = |A(i,p)|$);
$r_p = |\{i \mid 1 \leq i < p, n_p(i) \neq 0\}|$: the counter of different non-zero residues modulo p;
$R_p = \sum_{i=1}^{p-1} n_p(i)$; $R'_p = R_p + n_p(0)$; $S_p = \sum_{i=1}^{p-1} n_p^2(i)$;
$\frac{A(0,p)}{p} = \{a \mid ap \in A(0,p)\}$;
$prevpr(x)$: the prime number preceding x;
$nextpr(x)$: the prime number following x;

In this notation conditions (5), (6) and (7) will take form $R_p \geq p - 1$, $r_p > 2p^{1/2}$ and $R_p^3 \geq (c_5 p \log p) S_p$, respectively.

Algorithm 1.

1. Finding d
   (a) Initialization: $d \leftarrow 1$, $p \leftarrow 2$, $Q \leftarrow \lfloor \frac{3\ell}{m} \rfloor$.
   (b) $R_p \leftarrow 0$.
   For each $a \in A$ where $a \equiv 0 (\bmod d)$, compute $s = \frac{a}{d} - \lfloor \frac{a}{dp} \rfloor p$ and if $s \neq 0$ then advance the counter $R_p \leftarrow R_p + 1$;
   Continue this process until $R_p \geq Q$ or all elements are processed.

If $R_p \geq Q$ then set $p \leftarrow nextpr(p)$;
otherwise set $d \leftarrow dp$, $Q \leftarrow \lfloor \frac{3\ell}{d|A(0,d)|} \rfloor$ and $p \leftarrow 2$.
If $p \leq Q^{1/2}$ return to 1(b);
otherwise set $p \leftarrow prevpr(Q)$ and go to 1(c).

(c) $n_p(i) \leftarrow 0$ $(0 \leq i < p), R_p \leftarrow 0, S_p \leftarrow 0, R'_p \leftarrow 0, r_p \leftarrow 0$.
For each $a \in A$ for which $a \equiv 0 (\bmod d)$ compute $s = \frac{a}{d} - \lfloor \frac{a}{dp} \rfloor p$ and advance the counters:
$n_p(s) \leftarrow n_p(s) + 1$, $R'_p \leftarrow R'_p + 1$;
if $s \neq 0$ then $(R_p \leftarrow R_p + 1, S_p \leftarrow S_p + 2n_p(s) - 1$;
if $n_p(s) = 1$ then $r_p \leftarrow r_p + 1)$;
Continue this process until one of the following inequalities is true:

$$r_p > 2p^{1/2}, \quad R_p \geq p - 1, \quad R_p^3 \geq (c_5 p \log p) S_p, \tag{9}$$

or all elements are processed.
If all elements are processed $(n_p(0) > |A(0,d)| - p)$ then $d \leftarrow dp$.
If $R'_p \geq (\frac{16 c_5 r_p \ell \log \ell}{p})^{1/2}$ then $p \leftarrow prevpr(\min\{p-1, \frac{4 r_p \ell}{p R'_p}\})$;
otherwise $p \leftarrow prevpr(p-1)$.
If $p \geq Q^{1/2}$ return to 1(c); otherwise go to 1(d).

(d) Find $n_d(i)$, $1 \leq i < d$, and r_d for the set A.

2. Finding C – the set of all non-zero residues modulo d in A^*.

Define the sequence of sets $C_0, C_1, \ldots, C_{d-1}$ in the following way: $C_0 = \{0\}$ and, for $i > 0$, $C_i = C_{i-1} + \{0, i, \ldots, n_d(i)i\} (\bmod d)$ if $n_d(i) \neq 0$ or $C_i = C_{i-1}$ if $n_d(i) = 0$. Clearly, $C_{d-1} = C$.

Let v be a vector with d coordinates (numbered from 0 to $d-1$) which represents C_i in the way that if $j \in C_i$ then $v(j) = i$ and if $j \notin C_i$ then $v(j) = -1$.

(a) Initialization: $v \leftarrow (0, -1, \ldots, -1)$.
(b) For all i, $1 \leq i < d$, for which $n_d(i) \neq 0$ do
  for all j, $1 \leq j < d$, for which $0 \leq v(j) < i$ do
    $v(j) \leftarrow i$ and
    for s running from 1 to $n_d(i)$ while $v(j + si(\bmod d)) = -1$
      $v(j + si(\bmod d)) \leftarrow i$.

3. Finding S^*. Define $s \equiv M (\bmod d)$, $0 \leq s < d$.
Find $S^* = M - s + s_0$, where $s_0 = \max\{s_i \mid s_i \in C, s_i \leq s\}$.

To prove the validity of the algorithm we need to ensure that its step 1 finds a proper number d such that a set $\frac{A(0,d)}{d}$ satisfies all the conditions of Corollary 2.9. Indeed, sub-steps 1(b) and 1(c) use the conditions of the corollary. Therefore, the only thing that needs to be proved is the validity of the condition in sub-step 1(c) $\left(R'_p \geq \left(\frac{16 c_5 r_p \ell \log \ell}{p}\right)^{1/2}\right)$ which allows us to skip verification of some p's.

Recall that R'_p is the counter of elements of the set that have been checked for divisibility by p and that we stop the verification process for a particular prime number p once one of the conditions in (9) is satisfied. Therefore, the number of elements that have been checked for a particular p may be small (if many different non-zero residues are found in the beginning of the process) but this value may also be quite large. However, the fact that many elements have been checked for some $p' > Q^{1/2}$ ensures that A is p-full for many p's, namely, for $p > \frac{4r_{p'}\ell}{p'R'_{p'}}$. This is proved in the following lemma.

Lemma 3.1. — *Let B be a set of integers taken from the segment $[1, \ell]$. Assume that there is a prime $p' < \ell^{1/2}$ which satisfies the inequality*

$$|B| \geq \left(\frac{16c_5 r_{p'}\ell\log\ell}{p'}\right)^{1/2}, \tag{10}$$

where $r_{p'} = |\{i : |B(i,p')| \neq 0, 0 \leq i < p'\}|$ and c_5 is the constant from Lemma 2.7. Then, for prime numbers p, $\frac{4r_{p'}\ell}{p'|B|} < p < \ell^{1/2}$, $p \neq p'$, the set B is p-full.

Proof. — We are going to show that condition (7) of Lemma 2.7 is satisfied for all p's from the required interval. From this point on, for convenience we will use r without a subscript to denote $r_{p'}$.

Let $\{b_1, \ldots, b_r\}$ be the set of all classes of residues modulo p' of the set B and let t_i, $1 \leq i \leq r$, be the number of occurrences of residues from class b_i in the set B. Without losing generality, assume that $t_1 \geq t_2 \geq \cdots \geq t_r$. Among the t_i elements which are in the class of b_i modulo p', only $\lceil\frac{\ell}{pp'}\rceil < \frac{2\ell}{pp'}$ elements can belong to the same class of residues modulo p, $p \neq p'$. Therefore, these t_i elements of B belong to at least $\lceil\frac{t_i pp'}{2\ell}\rceil$ different classes of residues modulo p.

To estimate from above the value of $\sum_{i=1}^{p-1}|B(i,p)|^2$ in the left-hand side in (7) we have taken the worst case scenario where the number of different classes of residues modulo p is the smallest possible. For a given $|B|$, this case occurs when each class of residues contains the maximum possible number of elements. Thus, the number of classes is at least $\lceil\frac{t_1 pp'}{2\ell}\rceil$ and each class can include the following number of elements of B: less than $\frac{2\ell r}{pp'}$ elements in $\lceil\frac{t_r pp'}{2\ell}\rceil$ classes, $\frac{2\ell(r-1)}{pp'}$ elements in $\lceil\frac{t_{r-1}pp'}{2\ell}\rceil - \lceil\frac{t_r pp'}{2\ell}\rceil$ classes, …, and $\frac{2\ell}{pp'}$ elements in $\lceil\frac{t_1 pp'}{2\ell}\rceil - \lceil\frac{t_2 pp'}{2\ell}\rceil$ classes. (Recall that $|B| = \sum_{i=1}^{r} t_i$ is being given.) Using these values we can estimate

$$\begin{aligned}\sum_{i=1}^{p-1}|B(i,p)|^2 &\leq \left(\frac{2\ell r}{pp'}\right)^2\left\lceil\frac{t_r pp'}{2\ell}\right\rceil + \left(\frac{2\ell(r-1)}{pp'}\right)^2\left(\left\lceil\frac{t_{r-1}pp'}{2\ell}\right\rceil - \left\lceil\frac{t_r pp'}{2\ell}\right\rceil\right)\\ &\quad + \cdots + \left(\frac{2\ell}{pp'}\right)^2\left(\left\lceil\frac{t_1 pp'}{2\ell}\right\rceil - \left\lceil\frac{t_2 pp'}{2\ell}\right\rceil\right) - |B(0,p)|^2\\ &= \left(\frac{2\ell}{pp'}\right)^2\left(\left\lceil\frac{t_r pp'}{2\ell}\right\rceil(2r-1) + \left\lceil\frac{t_{r-1}pp'}{2\ell}\right\rceil(2r-3)\right.\\ &\quad \left. + \cdots + \left\lceil\frac{t_1 pp'}{2\ell}\right\rceil\cdot 1\right) - |B(0,p)|^2\end{aligned}$$

$$\leq \left(\frac{2\ell}{pp'}\right)^2 \left(\frac{t_r pp'}{2\ell}(2r-1) + \frac{t_{r-1}pp'}{2\ell}(2r-3) + \cdots + \frac{t_1 pp'}{2\ell} + r^2\right) - |B(0,p)|^2$$

$$\leq \left(\frac{2\ell r}{pp'}\right)^2 \cdot \frac{|B|}{r} \cdot \frac{pp'}{2\ell} + \left(\frac{2\ell r}{pp'}\right)^2 - |B(0,p)|^2$$

$$= \frac{2\ell r|B|}{pp'}\left(1 + \frac{2\ell r}{|B|pp'} - \frac{pp'|B(0,p)|^2}{2\ell r|B|}\right)$$

and, taking into account (10) and that $|B| > \frac{4r\ell}{pp'}$, we continue

$$\sum_{i=1}^{p-1} |B(i,p)|^2 \leq \frac{|B|^3}{8c_5 p\log\ell}\left(1 + \frac{1}{2} - \frac{2|B(0,p)|^2}{|B|^2}\right)$$

$$= \frac{(\sum_{i=1}^{p-1}|B(i,p)|)^3}{8c_5 p\log\ell} \cdot \frac{\frac{3}{2} - 2\alpha^2}{(1-\alpha)^3},$$

where $\alpha = \frac{|B(0,p)|}{|B|}$. To prove now the validity of (7) for p it is sufficient to show that $\frac{\frac{3}{2}-2\alpha^2}{(1-\alpha)^3} \leq 8$. It is easy to see that the function in the left-hand side of this inequality increases with α for $\alpha < \frac{2}{3}$ and, therefore, the inequality holds true for $\alpha \leq \frac{1}{2}$. Indeed, since the number of elements in one class of residues modulo p cannot exceed $\frac{2\ell r}{pp'}$ and $|B| > \frac{4\ell r}{pp'}$, $\alpha = \frac{|B(0,p)|}{|B|} \leq \frac{1}{2}$ that concludes the proof. □

The complexity. — Step 1 checks the divisibility of elements a_i by different prime numbers p. Since $a_i \leq \ell$, the number of prime divisors of a_i cannot be more than $\log_2 \ell$. Therefore, the overall number of occurrences where some p divides some element of A is $O(m\log m)$. In order to estimate the number of occurrences where some p does not divide some element of A we need to investigate each part of Step 1 separately.

In Step 1(b), in the worst case, we may find Q elements not divisible by p while verifying this number p. Since this part of Step 1 deals with prime numbers less than $Q^{1/2}$, the number of operations in Step 1(b) where some p does not divide some element of A is $O(Q^{3/2}) = O((\frac{\ell}{m})^{3/2})$. (Recall that $Q \sim \frac{\ell}{m}$.)

In step 1(c), again, no more than p elements not divisible by p may be found. Thus, the number of operations in Step 1(c) where some p does not divide some element of A is limited by $O(Q^2) = O((\frac{\ell}{m})^2)$. In fact, for $m \leq \frac{\ell^{3/5}}{\log^{2/5}\ell}$ this estimate can be improved.

If the number of verified elements is sufficiently large ($R'_p \geq (\frac{16c_5 r_p \ell \log\ell}{p})^{1/2}$) for some p, we are able to skip verification of some numbers according to Lemma 3.1. (The above "skipping" condition supersedes condition $R'_p > \frac{4r_p\ell}{p^2}$ for $p > \ell^{2/5}$ which ensures that the next number to be verified is less than p.)

Let us analyze this situation. The worst scenario (from a complexity point of view) occurs when we do not reach the "skipping" condition during verification. Thus, the number of operations in Step 1(c) where some p does not divide some element of A

is limited by

$$\sum_{p=\lceil Q^{1/2}\rceil}^{\lfloor \ell^{2/5}\rfloor} p + \sum_{p=\lfloor \ell^{2/5}\rfloor+1}^{\lfloor Q\rfloor} \left(\frac{16c_5 r_p \ell \log \ell}{p}\right)^{1/2} = O\left(\int_{Q^{1/2}}^{\ell^{2/5}} x dx + \int_{\ell^{2/5}}^{Q} \frac{(\ell \log \ell)^{1/2}}{x^{1/4}} dx\right).$$

Here we took into consideration the first condition in (9) which implies $r_p \leq 2p^{1/2}$. By keeping after integration only the most significant term in each integral, we obtain complexity

$$O(\ell^{1/2} Q^{3/4} \log^{1/2} \ell) = O\left(\frac{\ell^{5/4} \log^{1/2} \ell}{m^{3/4}}\right). \tag{11}$$

This estimate is obtained assuming $p > \ell^{2/5}$. Observe that p can be greater than $\ell^{2/5}$ only for $m \leq \ell^{3/5}$ since $p \leq Q \sim \frac{\ell}{m}$. Comparing (11) with the first estimate – $O((\frac{\ell}{m})^2)$ – one can see that (11) improves it for $m \leq \frac{\ell^{3/5}}{\log^{2/5} \ell}$.

Combining the results for sub-steps 1(b) and 1(c), one can get the overall complexity of the process that verifies divisibility of elements of A:

$$O\left(m \log m + \min\left\{\left(\frac{\ell}{m}\right)^2, \frac{\ell^{5/4} \log^{1/2} \ell}{m^{3/4}}\right\}\right). \tag{12}$$

This estimate also holds true for the overall complexity of the algorithm, since in the worst scenario both steps 1(d) and 2 have complexity $O(m)$.

In conclusion, the only thing that remains is to analyze the above expression (12). The second term dominates for $m \leq \ell^{2/3} \log^{1/3} \ell$. It is equal to $O(\frac{\ell^{5/4} \log^{1/2} \ell}{m^{3/4}})$ for $m \leq \frac{\ell^{3/5}}{\log^{2/5} \ell}$ and $O((\frac{\ell}{m})^2)$ otherwise. This improves the algorithms from **[9]** and **[11]** for low density $\left(m \leq \frac{\ell^{3/5}}{\log^{2/5} \ell}\right)$. In the worst case $(m \sim (\ell \log \ell)^{1/2})$ time is $O(m^{7/4} / \log^{3/4} m)$.

References

[1] Alon N., and Freiman G. A., *On Sums of Subsets of a Set of Integers*, Combinatorica, **8**, 1988, 305–314.

[2] Buzytsky P., and Freiman G.A., *Analytical Methods in Integer Programming*, Moscow, ZEMJ., (Russian), 1980, 48 pp.

[3] Chaimovich M., *An Efficient Algorithm for the Subset-Sum Problem*, a manuscript, 1988.

[4] Chaimovich M., *Subset-Sum Problems with Different Summands: Computation*, Discrete Applied Mathematics, **27**, 1990, 277–282.

[5] Chaimovich M., *Solving a Value-Independent Knapsack Problem with the Use of Methods of Additive Number Theory*, Congressus Numerantium, **72**, 1990, 115–123.

[6] Chaimovich M., Freiman G.A., and Galil Z., *Solving Dense Subset-Sum Problem by Using Analytical Number Theory*, J. of Complexity, **5**, 1989, 271–282.

[7] Erdős P., and Freiman G., *On Two Additive Problems*, J. Number Theory, **34**, 1990, 1–12.

[8] Freiman G.A., *An Analytical Method of Analysis of Linear Boolean Equations*, Ann. New York Acad. Sci., **337**, 1980, 97–102.

[9] Freiman G.A., *Subset-Sum Problem with Different Summands*, Congressus Numerantium, **70**, 1990, 207–215.

[10] Freiman G.A., *New Analytical Results in Subset-Sum Problem*, Discrete Mathematics, **114**, 1993, 205–218.

[11] Galil Z., and Margalit O., *An Almost Linear-Time Algorithm for the Dense Subset-Sum Problem*, SIAM J. of Computing, **20**, 1991, 1157–1189.

[12] Lipkin E., *On Representation of r-Powers by Subset-Sums*, Acta Arithmetica, **LII**, 1989, 353–366.

[13] Martello S. and Toth T., *The 0-1 Knapsack Problem*, in Combinatorial Optimization, ed: N. Christofides, A.Mingozzi, P. Toth, C.Sandi, Wiley, 1979, 237–279.

[14] Olson J., *An Addition Theorem Modulo p*, J. of Combinatorial Theory, **5**, 1968, 45–52.

[15] Sárkőzy A., *Finite Addition Theorems II*, J. Number Theory, **48**, 1994, 197–218.

M. CHAIMOVICH, 7041 Wolftree Lane, Rockville MD 20852, USA
E-mail : mark.chaimovich@bellatlantic.COM