Astérisque

MARK CHAIMOVICH

New structural approach to integer programming : a survey

Astérisque, tome 258 (1999), p. 341-362

<http://www.numdam.org/item?id=AST_1999__258__341_0>

Astérisque **258**, 1999, p. 341–362

NEW STRUCTURAL APPROACH TO INTEGER PROGRAMMING: A SURVEY

by

Mark Chaimovich

Abstract. — The survey discusses a new approach to Integer Programming which is based on the structural characterization of problems using methods of additive number theory. This structural characterization allows one to design algorithms which are applicable in a narrower, yet still wide, domain of problems, and substantially improve the time boundary of existing algorithms. The new algorithms are polynomial for the class of problems in which they are applicable, and even linear ($O(m)$) for a wide class of the Subset-Sum and Value-Independent Knapsack problems. Previously known polynomial time algorithms for the same classes of problems are at least two orders of magnitude slower.

1. Introduction

This survey considers a recently developed approach to Integer Programming (IP) which is based on the application of analytical methods of Additive Number Theory. Elaborated by G. Freiman in the early 1980's, this new approach was developed by N. Alon, P. Buzytsky, M. Chaimovich, P. Erdős, G. Freiman, Z. Galil, E. Lipkin and O. Margalit (in alphabetical order).

In general, the number of Integer Programming models is vast and they have numerous applications; only a few of them – Subset-Sum (one and multi-dimensional), Value-Independent Knapsack and k-Partition problems – were investigated using the new structural approach. Theorems from analytical number theory allow one to characterize the structure of the domain of solutions for a wide class of problems and to design efficient algorithms for these problems. These new algorithms substantially improve the time boundary of existing algorithms. They are polynomial for the class of problems in which they are applicable, and even linear ($O(m)$) for certain classes of the Subset-Sum and Value-Independent Knapsack problems. That is at least two orders of magnitude faster than previously known polynomial time algorithms for the

1991 ***Mathematics Subject Classification.*** — Primary: 90C02 Alternate: 05A17, 11B25, 68Q25.
Key words and phrases. — Analytical Number Theory, Integer Programming, Subset Sum Problem.

same classes of problems. This fact allows one to solve problems with a much larger number of variables.

This article is organized into several parts. In section 2 the general idea for development of an analytical approach to Integer Programming is considered. Sections 3 and 4 deal with the Subset-Sum Problem (SSP). The first of them provides a detailed, structural analysis of the problem including an example of the analytical theorem while the second describes algorithms for solving SSP based on this structural analysis. Proofs of the validity of the algorithms are not provided in this survey, however, they may be found in the references. Section 5 describes the application of the structural approach to multi-dimensional Subset-Sum, Value-Independent Knapsack and k-Partition problems. (Only the main theorems and outlines of the algorithms are presented.) In the conclusion possible directions for future research are discussed.

2. General idea of the application of the structural approach to IP

In this section the main idea of the structural approach is described. We begin with a simple example that illustrates the approach. Further, the concept of density is discussed, this explains how the structural characterization of the problem may be obtained. We conclude the section with a short history of the research in the field of structural characterization.

2.1. A simple illustration of the structural approach. — In order to understand a structural approach to IP, consider the problem of feasibility of a single boolean equation. Given an integer m, an integral vector $(a_1, a_2, \ldots, a_m)$ and an integer N, does equation

$$a_1x_1 + a_2x_2 + \cdots + a_mx_m = N \tag{1}$$

have any solutions for $x_i \in \{0,1\}$ for all i? To illustrate the approach, we use the following concrete equation

$$7x_1 + 8x_2 + 14x_3 + 15x_4 + 22x_5 + 28x_6 + 56x_7 = 75, \tag{2}$$

i.e. $m = 7$, $(a_1, \ldots, a_7) = (7, 8, 14, 15, 22, 28, 56)$ and $N = 75$.

Dynamic programming approach

Denoting $S_0 = \{0\}$ and $S_k = \{b | b = \sum_{j=1}^{k} a_j x_j, x_j \in \{0,1\}\}$ for $1 \leq k \leq 7$, we have $S_k = S_{k-1} + \{0, a_k\} = \{b | b \in S_{k-1} \text{ or } b - a_k \in S_{k-1}\}$. Thus, having S_7 – the set of all possible values of the linear form in the left-hand side of (2),– it remains only to check if $N = 75 \in S_7$. In fact,

$$\begin{aligned} S_1 &= \{0, 7\}, \\ S_2 &= \{0, 7, 8, 15\}, \\ S_3 &= \{0, 7, 8, 14, 15, 21, 22, 29\}, \\ &\ldots \end{aligned}$$

and so on. Finally,

$$S_7 = \{0,7,8,14,15,21,22,23,28,29,30,35,36,37,42,43,44,45,49,50,\\ 51,52,56,57,58,59,63,64,65,66,70,71,72,73,77,78,79,80,\dots\},$$

i.e., $75 \notin S_7$ and equation (2) does not have a solution.

Structural approach

We characterize the structure of S_7 without explicitly enumerating it. Observe, that some of the coefficients of the equation are divisible by 7: $a_1 \equiv a_3 \equiv a_6 \equiv a_7 \equiv 0(\mathrm{mod}\,7)$. Then, for $b \in S_7$ we have $b \equiv 8x_2 + 15x_4 + 22x_5 \equiv x_2 + x_4 + x_5(\mathrm{mod}\,7)$, i.e.,

$$b \equiv 0,1,2,3(\mathrm{mod}\,7). \tag{3}$$

However, $75 \equiv 5(\mathrm{mod}\,7)$, so, the equation does not have a solution.

Condition (3) determines a necessary condition for solvability equation (2). In order to obtain a sufficient condition let us analyze the same equation with another right-hand side:

$$7x_1 + 8x_2 + 14x_3 + 15x_4 + 22x_5 + 28x_6 + 56x_7 = 79.$$

Clearly, $79 \equiv 2(\mathrm{mod}\,7)$, so it can belong to S_7 according to (3). To confirm that it really belongs to S_7, consider a linear form

$$L = 7x_1 + 14x_3 + 28x_6 + 56x_7 = 7(x_1 + 2x_3 + 4x_6 + 8x_7).$$

The linear form $L' = x_1 + 2x_3 + 4x_6 + 8x_7$ can take all values from 0 to 15, thus, the linear form L can, correspondingly, take values of the form $7t$, where $0 \le t \le 15$. When we combine these values with the other coefficients $(8, 15, 22)$, we have

$$\begin{aligned} S_7 = \{b \mid\ & b \equiv 0(\mathrm{mod}\,7), 0 \le b \le 7 \cdot 15, \ \textit{or}\\ & b \equiv 1(\mathrm{mod}\,7), 8 \le b \le 22 + 7 \cdot 15, \ \textit{or}\\ & b \equiv 2(\mathrm{mod}\,7), 23 \le b \le 37 + 7 \cdot 15, \ \textit{or}\\ & b \equiv 3(\mathrm{mod}\,7), 45 \le b \le 45 + 7 \cdot 15\}. \end{aligned} \tag{4}$$

Here 8, 23, 45 are the smallest numbers with residues 1, 2, 3 modulo 7 that can be represented by the linear form in the left-hand side of the equation. Since $79 \equiv 2(\mathrm{mod}\,7)$ and $79 = 23 + 7 \cdot 8$, the answer is that the equation has at least one solution.

Observe that the above consideration determines the structure of the set of possible values of a linear form on the left-hand side of an equation as a collection of arithmetic progressions with a common difference. This fact allows one to solve the problem immediately for each right-hand side. One can suppose that this example was especially selected to illustrate the approach and that would be true. However the situation obtained can be generalized: for a wide class of problems we can always determine the structure.

To obtain a general structural characterization of the IP problem (in the same way that (4) was obtained for a concrete equation), a specific analytical theorem must be proven. Of course, certain conditions have to be imposed on the coefficients in order to obtain such a characterization. These conditions follow directly from the analytical

theorem. Once we have the conditions, it is possible to go to the next step – to design algorithms to verify these conditions and to obtain the structure.

Indeed, the structure obtained and the conditions of its existence provide an understanding of why some problems are easy and others are very hard for various enumerative algorithms. To confirm this statement consider the following problem which was investigated by R. Jeroslow (1974) [**19**]: maximize x_1 satisfying $2x_1 + 2x_2 + \cdots + 2x_n = n$ where n is odd. Although this problem is by nature trivial, it requires almost complete enumeration using different enumerative techniques. (Branch and Bound, for example, is one of them.) The secret is the fact that the constraint has no solutions, however, we must verify all possibilities to confirm this fact. The structural approach allows one to obtain an answer for this problem in no time.

2.2. Concept of density and its use in structural characterization. — In order to apply analytical methods to solve an IP problem, it is necessary for the problem to have a *high density.* To explain the notion of "density" and its importance in the application of the analytical approach to IP, let us consider again the feasibility of equation (1).

Let $\ell = \max\limits_{1\le i\le m} a_i$. The linear function on the left in (1) has a domain of size 2^m and a range of size $m\ell$. Since the domain size represents the overall number of "solutions" for all possible values of the right-hand side, the ratio $\frac{2^m}{m\ell}$ represents the average number of "solutions" for a value from the range. We say that this ratio characterizes the density of the problem. The density of other IP problems can be defined similarly.

In the case of equation (1), the density condition means that $\ell = o(\frac{2^m}{m})$ or $\frac{2^m}{m\ell} \to \infty$. Currently, algorithms are still not capable of handling this density. The only situation that has been investigated is $\ell = O(\frac{m^2}{\log m})$. The conjecture of G. Freiman is that the new approach can be refined to handle the case $\ell = O(m^c)$ for any positive constant c.

To highlight basic features of the approach, we present some non-strict considerations resulting from probability theory. In view of

$$\int_0^1 e^{2\pi i\alpha b} d\alpha = \begin{cases} 0 & \text{for } b \in \mathbb{Z}, b \neq 0, \\ 1 & \text{for } b = 0, \end{cases}$$

it is easy to verify that the number of solutions of (1) can be expressed by the integral

$$(5)\quad J(N) = \int_0^1 \prod_{j=1}^m (1 + e^{2\pi i\alpha a_j}) e^{-2\pi i\alpha N} d\alpha = 2^m \int_0^1 \prod_{j=1}^m (\tfrac{1}{2} + \tfrac{1}{2} e^{2\pi i\alpha a_j}) e^{-2\pi i\alpha N} d\alpha.$$

One may look at $\frac{1}{2} + \frac{1}{2}e^{2\pi i\alpha a_j}$ as the characteristic function of a random variable ξ_j taking values 0 and a_j with probabilities equal to $\frac{1}{2}$. Then the value of integral (5) is equal to the probability $P(\zeta = N)$, where $\zeta = \xi_1 + \cdots + \xi_m$ is a random variable with mathematical expectation $M = \frac{1}{2}\sum_{j=1}^m a_j$ and dispersion $\sigma^2 = \frac{1}{4}\sum_{j=1}^m a_j^2$. Assuming that the local limit theorem can be applied, the variable ζ has asymptotically normal

distribution; we therefore have

$$J(N) \sim \frac{2^m}{\sqrt{2\pi\sigma^2}} e^{-\frac{(M-N)^2}{2\sigma^2}}, \tag{6}$$

which implies the existence of solutions for equation (1) for right-hand sides N in a wide interval of the mathematical expectation M.

As a rule, a local limit theorem is not always available. In spite of this difficulty, the precise analysis of integral (5) (see [**11**], [**20**], [**1**], for example) confirms the validity of the asymptotic formula (6) for sufficiently dense equations whose coefficients satisfy some distributive properties. These distributive properties require that not "too many" coefficients have one common divisor. Non-compliance with this requirement provides a special structure for the range of the linear form on the left in (1).

Note that application of the analytical approach to a specific IP model requires one to prove the model's own structural theorem. Alternatively, one can reduce the new problem to another problem for which the structural theorem is already proved.

2.3. Historical background. — The possibility of using analytical methods for solving IP problems was shown for the first time by G. Freiman in 1980 [**12**] (see also P. Buzytsky and G. Freiman [**3**]). However, at that time, his concepts did not provide an explicit structural characterization of the problem. Only recently has determination of a precise structure for some IP problems become possible on the basis of methods proposed by G. Freiman and P. Erdős in [**11**].

The first works investigating structural characterization of IP using analytical methods were concentrated on the following Subset-Sum Problem (SSP) with *different summands*: Given a set A of positive integers and a number N, find (a) $z = \max\{S_B = \sum_{a\in B} \mid S_B \leq N, B \subseteq A\}$ and (b) subset $B \subset A$ such that $S_B = z$.

The authors proved the analytical theorems, showing that a set of subset-sums around the middle sum may be characterized as a collection of long arithmetic progressions with a common difference. P. Erdős and G. Freiman [**11**] assumed very dense inputs ($m > \frac{\ell}{3}$) and a very small interval. E. Lipkin [**20**] improved the density ($m > \ell^{4/5+\varepsilon}$) and enlarged the interval size. N. Alon and G. Freiman [**1**] further improved the density ($m > \ell^{2/3+\varepsilon}$) but used a small interval (like in [**11**]). Later, G. Freiman [**17**] proved the same result for sets with density $m > c(\ell \log \ell)^{1/2}$. All of these characterization theorems used analytic number theory and hold true for sufficiently large values of ℓ. M. Chaimovich ([**5**] and [**9**]) shows the existence of an arithmetic progression in subset sums for sets with density $m > g(\ell)\ell^{2/3}\log^{1/3}\ell$, $\ell \geq 155$, where $g(\ell)$ is some function depending on ℓ, $1.9 < g(\ell) < 2.5$. The proof is done with exact computation of all constants, which allows one to use the result in practical algorithms.

A. Sárközy [**22**] has independently obtained an arithmetic progression for sets with the same density as [**17**]; he used algebra and combinatorial methods. However, his proof is not constructive and therefore it may not be applicable to algorithmic design. Z. Galil and O. Margalit [**18**], using elementary number theoretic facts only (in contrast to A. Sárközy's approach), have explicitly constructed a long arithmetic

progression in subset-sums. They achieved density $m = O(\ell^{1/2} \log \ell)$ (slightly weaker than [**17**] and [**22**]) and matched the interval size of [**20**].

To complete the discussion of the results related to SSP with different summands, we mention that G. Freiman has shown in [**17**] that the density $m = \Omega(\ell^{1/2})$ is the lowest density for which the characterization of the structure of subset-sums is an arithmetic progression. For sets of different summands with $m < \ell^{1/2}$ the structure is more complicated. G. Freiman conjectures that this structure is multi-dimensional in the sense that it is formed by a few, relatively short, arithmetic progressions, and that each of these arithmetic progressions can be viewed as a "dimension" of the structure.

The first algorithms solving the SSP were derived by G. Freiman [**13**], [**14**] and M. Chaimovich [**5**]. They solve in linear time problem (a) which finds the maximal sum but not the subset. In comparison, dynamic programming solves the same problem in $O(m^2\ell)$ time which is two orders of magnitude slower. Solving problem (b) with this approach (see [**10**]) takes $O(\ell^2 \log \ell)$ time. The algorithm of Z. Galil and O. Margalit [**18**] solves both problems (a) and (b) – finding the maximal sum and the subset. It reaches $O(\ell \log \ell)$ time improving [**10**] by one order of magnitude.

The SSP with repeated summands (relaxing the restriction that the summands must be distinct) was considered by M. Chaimovich in [**4**], [**6**]. The existence of a long arithmetic progression in a set of subset-sums was proved for $m > 6\ell \log \ell$. This estimate is the best possible apart from a logarithmic factor and a constant.

Investigation of the multi-dimensional SSP, where vectors take place of the integers, was begun by G. Freiman [**16**]. He has shown that for two-dimensional problems an integral lattice takes the place of an arithmetic progression in the structural characterization. M. Chaimovich [**8**] extended this result for an arbitrary number of dimensions. For multi-dimensional problems the time boundary of the new algorithm is more impressive than the one dimensional one: for n-dimensional SSP it reaches $O(m^2)$ time instead of $O(m^{n+2})$ in dynamic programming.

Another problem investigated recently by using the structural approach was a k-partition problem (KPP): Given a set A of positive integers and k positive target numbers $N_1 \leq N_2 \leq \cdots \leq N_k$ such that $\sum_{i=1}^{k} N_i = S_A$, find a partition of A into k subsets $(B_1, \ldots, B_k), \bigcup_{j=1}^{k} B_j = A$, whose sums are closest to the target numbers in the sense that they minimize (or maximize) an appropriate objective function z.

This problem is especially hard to solve using traditional methods. It was solved by dynamic programming (see [**21**]) in $O(m^{2k})$ time. Applying the structural approach to it (M. Chaimovich [**7**]) gives $O(m^{1+1/(k-1)})$ time for sufficiently dense input sets. The gain is considerable for a fixed k as well as for k increasing with ℓ (its value is bounded by $\ell^{1/4}$).

3. Analytical method for structural analysis of the Subset-Sum Problem.

In this section we provide a detailed explanation of the structural characterization of the set of subset-sums. First we determine sufficient conditions of the existence of a long interval in a set of subset-sums (Theorem 3.1). Next we elaborate the structure

of the set of subset-sums in the case where sufficient conditions are not fulfilled. Our consideration is based on the proofs from [**1**] and [**9**].

3.1. Existence of a long interval in a set of subset-sums – sufficient conditions. — For a set X define

$$m_X = |X|, \quad \ell_X = \max\{x \in X\}, \quad S_X = \sum_{x\in X} x, \quad \sigma_X = \tfrac{1}{2}(\sum_{x\in X} x^2)^{1/2},$$
$$X^* = \{z \mid \exists Y \subseteq X, S_Y = z\}, \quad X(s,q) = \{x \in X \mid x \equiv s(\mathrm{mod}\, q)\}.$$

Note: In the following theorem and further on $c_0, c_1, c_2, \ldots$, always denote absolute positive constants. We will also omit the subscript identifying the set if it is clear from the context which set is being discussed.

Theorem 3.1. — *Let A be a set of positive integers, such that*

$$m > c_1 \ell^{2/3} \log^{1/3} \ell > c_0. \tag{7}$$

Suppose that for all integers q the inequality

$$|A(0,q)| \le m - \ell^{2/3} \log^{1/3} \ell \tag{8}$$

is true. Then all integers N for which

$$\left|N - \tfrac{1}{2}S_A\right| \le c_2\sigma_A \tag{9}$$

belong to the set of subset-sums of A, i.e., $\left[\frac{1}{2}S_A - c_2\sigma_A, \frac{1}{2}S_A + c_2\sigma_A\right] \subseteq A^*$.

General idea of the proof. — The fact that an integer N belongs to the set of subset-sums is equivalent to the existence of a solution of a linear equation (1).

For $1 \le j \le m$ define $\varphi_j(\alpha) = \frac{1}{2}\left(1 + e^{2\pi i \alpha a_j}\right)$ and $\varphi(\alpha) = \prod_{j=1}^{m} \varphi_j(\alpha)$. As mentioned on page 344, the number of solutions to equation (1) can be expressed by the integral $J(N) = 2^m \int_0^1 \varphi(\alpha) e^{-2\pi i \alpha N} d\alpha$, thus, it is necessary to show that $J(N) \ge 1$ whenever N satisfies (9). In order to do this we can prove the asymptotic formula

$$J(N) = (1 + o(1)) \frac{2^m}{\sqrt{2\pi\sigma^2}} e^{-\frac{(M-N)^2}{2\sigma^2}}, \tag{10}$$

for the number of solutions of equation (1).

Let us analyze the nature of conditions of the theorem. A restriction (9) on number N ensures that the exponent in (10) is not too small. This restriction is necessary to obtain an asymptotic formula, but not to prove the existence of a solution and/or a structure, therefore, we will relax this restriction below.

Condition (7) represents the density of a problem in the sense that the number of combinations of unknowns is large with respect to a range of possible values of the linear form. This condition can be strengthened to $m > c\ell^{1/2} \log^{1/2} \ell$ (see [**17**] and [**22**]) , but then the proof becomes quite complicated. In any case we need a condition of density to ensure the existence of the structure.

Finally, condition (8) is a condition of distribution. Its validity is necessary to obtain an asymptotic formula, but it is not necessary to obtain a structural characterization. The influence of a distribution of summands on a structure will be studied in the next paragraph.

Let us define $F_N(\alpha) = \varphi(\alpha)e^{-2\pi i\alpha N}$ and $L = 2\ell$. Observing that $F_N(\alpha)$ is a periodic function with a period equal to 1, one can write

$$(11)\quad J(N) = 2^m \int_{-\frac{1}{L}}^{1-\frac{1}{L}} F_N(\alpha)d\alpha = 2^m \left(\int_{-\frac{1}{L}}^{\frac{1}{L}} + \int_{\frac{1}{L}}^{1-\frac{1}{L}}\right) \geq 2^m \left(\left|\int_{-\frac{1}{L}}^{\frac{1}{L}}\right| - \left|\int_{\frac{1}{L}}^{1-\frac{1}{L}}\right|\right).$$

Note that for a sufficiently high density, the first integral on the right side of the equation in (11) provides the major part of the asymptotic formula for $J(N)$; the second integral forms the error term. The proof estimates these two integrals separately. It shows that

$$\int_{-\frac{1}{L}}^{\frac{1}{L}} F_N(\alpha)d\alpha = (1+o(1))\frac{2^m}{\sqrt{2\pi\sigma_A^2}}e^{-\frac{(M-N)^2}{2\sigma_A^2}},$$

and that

$$\left|\int_{\frac{1}{L}}^{1-\frac{1}{L}} F_N(\alpha)d\alpha\right| = o(\frac{1}{\sigma_A}).$$

In this survey we omit the detailed explanation of the integrals' evaluation.

Enlarging the interval. — According to Theorem 3.1, the interval $[\frac{1}{2}S_A - c_2\sigma_A, \frac{1}{2}S_A + c_2\sigma_A]$ belongs to the set of subset-sums. The length of the interval may be easily estimated to be at least $\sigma_A = \Omega(\ell\log^{1/2}\ell) \gg \ell$. However, this length is "small" relative to the range of subset sums which is $S_A > \frac{1}{2}m^2 \gg \ell^{4/3}$ (for our density).

Now we are going to show that this interval may be larger without considerably enlarging density.

Take the set A with $m_A = c_3 m_0$ where $m_0 = c_1\ell^{1/2}\log^{1/2}\ell$ (the density required by the theorem) and $c_3 \geq 2$. Let $1 \leq a_1 < a_2 < \cdots < a_m$ where $a_i \in A$ and denote $A' = \{a_j\}_{j=1}^{m_0}$. Suppose also that A' satisfies a condition similar to condition (8) of Theorem 3.1 so that Theorem 3.1 can be applied to set A'. According to the theorem, interval $I = [\frac{1}{2}S_{A'} - c_2\sigma_{A'}, \frac{1}{2}S_{A'} + c_2\sigma_{A'}]$, which is longer than ℓ, belongs to the set of subset sums of A'.

Return now to the original set A. Denote $N_i = N - \sum_{j=1}^{i} a_{m_0+j}$ for $1 \leq i \leq m-m_0$. Clearly, $0 < N_i - N_{i+1} \leq \ell$ and, whenever $N \in [\frac{1}{2}S_{A'}, S_A - \frac{1}{2}S_{A'}]$, for some i_0 we will have $N_{i_0} \in I$, which means that the entire interval $[\frac{1}{2}S_{A'}, S_A - \frac{1}{2}S_{A'}]$ belongs to the set of subset sums of A.

To estimate the length of this long interval we recall that set A' consists of m_0 smallest elements of A such that $S_{A'} \leq \frac{1}{c_3}S_A$. Thus, the length of the interval in the set of subset sums is at least $(1 - \frac{1}{c_3})S_A = O(S_A)$.

3.2. Elaborating on the structure of the set of subset sums. — To elaborate on the structure of the set of subset sums, we consider the case where condition (8) is

not satisfied for the set A. We will show that in this situation we have an arithmetic progression (instead of an interval) belonging to the set of subset sums.

The situation when an arithmetic progression is obtained is unusual situation. It is characterized by the fact that many elements of A are divisible by the same integer Q.

To refine the structure of the set of subset sums we manipulate it with those elements of A which are not divisible by this integer Q, i.e., have non zero residues modulo Q.

The results from [**15**] and [**9**] are used in the presentation of the section.

Arithmetic progression. — If condition (8) is true for all q's, then an arithmetic progression with the difference $Q = 1$ beginning before $s = \frac{1}{2}S_A - \sigma_A$ and having length more than $h = 2\sigma_A \gg \ell$ belongs to the set of subset sums.

Assuming that (8) fails for some integer q, we construct a sequence of sets $A_0, \dots, A_p$ and a sequence of integers $q_0, \dots, q_p$ in the following way:

Assign $A_0 = A, q_0 = 1$, and assume that set A_i has already been found. Introduce also $q_i' = \prod_{j=0}^{i} q_j$. The integer q_{i+1} will be an integer such that

$$|A_i \setminus A_i(0, q_{i+1})| \le \ell_{A_i}^{2/3} \log^{1/3} \ell_{A_i}. \tag{12}$$

If such an integer q_{i+1} exists, construct $A_{i+1} = \frac{A_i(0,q_{i+1})}{q_{i+1}} = \frac{A(0,q_{i+1}')}{q_{i+1}'}$; if such an integer q_{i+1} does not exist, set $p = i$ and $Q = q_p'$. This Q is not large, it is less than $\frac{3\ell}{2m}$ (see [**15**]).

Consider the set A_p obtained at the end of the process. It may be shown that for the set A_p, all the conditions of Theorem 3.1 are true. Therefore, apply Theorem 3.1 to A_p in order to arrive at

$$\left[\tfrac{1}{2}S_{A_p} - c_2\sigma_{A_p}, \tfrac{1}{2}S_{A_p} + c_2\sigma_{A_p}\right] \subseteq A_p^*.$$

Recalling that $A(0, Q) = \{aQ \mid a \in A_p\}$, we obtain a long segment of a progression with difference Q being contained in $(A(0, Q))^*$.

Refining the structure using residues. — In the previous paragraph it was shown that an arithmetic progression with a small difference ($Q \le \frac{3\ell}{2m}$) belongs to the set of subset sums. Furthermore, these subset sums (elements of the arithmetic progression) may be constructed using only the elements of A that have zero residue modulo Q (the difference of the arithmetic progression). The next step is to try to use the remaining elements of A (with non-zero residues modulo Q) to refine the structure by "filling" the "holes" in the progression. To do this we need some properties of subset sums. In this survey we will only list these properties; the proofs may be found in [**15**], [**9**].

Properties of subset sums modulo integer q. — Consider ring $\mathbb{Z}_q$ of residues mod q. For $d \in \mathbb{Z}_q$, $d \mid q$, define $H_d = \{0, d, 2d, \dots, (\frac{q}{d} - 1)d\}$, and for $r \in \mathbb{Z}_q$ define $H_d(r) = r + H_d$.

(a) Let C be a set of elements of the ring $\mathbb{Z}_q$. If an element $b \in \mathbb{Z}_q$ is such that $C = C + \{0, b\}$, then the set C has the following structure: for each $r \in C$, we have $H_d(r) \subseteq C$, where $d = \gcd(b, q)$; i.e., $C = \bigcup_{r \in C} H_d(r)$.

(b) Let set $C \subset \mathbb{Z}_q$ have the following structure: $C = \bigcup_{r \in C} H_d(r)$ for some d, $d|q$. Then for any $b \in \mathbb{Z}_q$, the set $C + \{0, b\}$ has the same structure.

Refining the structure. — We continue from the following point: an integer $Q \leq 3\ell_A/2m_A$ is found, such that a long arithmetic progression with the difference Q belongs to $(A(0,Q))^*$. Let $A \setminus A(0,Q) = \{b_1, \ldots, b_w\}$, and define a sequence of numbers $d_0, \ldots, d_w$ in the following way.

Let $B_i = \{b_1, \ldots, b_i\}$ and C_i be the set of the smallest non-negative residues modulo Q of $B_i^* \cup \{0\}$, i.e., $C_0 = \{0\}$ and $C_i = C_{i-1} + \{0, b_i\} (\mathrm{mod}\, Q)$. Let $d_0 = Q$. If the numbers $d_0, \ldots, d_{i-1}$ have already been determined, take $d_i = d_{i-1}$ when $|C_i| > |C_{i-1}|$ and $d_i = \gcd(d_{i-1}, b_i)$ when $|C_i| = |C_{i-1}|$. In this way the numbers d_i and sets C_i possess property (a), i.e., for any $c \in C_i$ we have $H_{d_i}(c) \in C_i$. At the end of the process we obtain the set C of all non-zero residues modulo Q which may be represented by subset sums A^*. This set has the following structure: $C = \cup_{c \in C} H_{d_w}(c)$ where $d_w \leq Q$. Combining the set C with the previously obtained arithmetic progression we conclude that the structure of the set of subset sums may be characterized as *a collection of long arithmetic progressions with a common difference.*

Relaxing the condition of distribution. — Working with residues allows not only the refining of the structure as was shown above, it also provides the way to relax the condition (8) of distribution in Theorem 3.1. Indeed, looking on the structure of set C above, one can see that the result of the previous paragraph is a collection of arithmetic progressions with a difference $d_w \leq Q$. It might be shown that $d_w \neq 1$ only if $|A(0, d_w)| > m - d_w$. This means that condition (8) might be replaced by condition

$$|A(0,q)| \leq m - q \tag{13}$$

and we would still get a long interval belonging to A^*.

3.3. Reducing density. — There are a few ways to achieve an arithmetic progression in a set of subset sums for lower density, namely, for $m > c(\ell \log \ell)^{1/2}$.

Analytical approach

G. Freiman in **[17]** proves that the asymptotic formula (10) is still valid for the lower density if the elements of A are "well distributed". "Bad distribution" in his consideration means one of the following:
(1) there are too many small elements in A.
(2) there are too many elements in A divisible by one number q.
(3) there are too many elements in A belonging to a two-dimensional structure.
All these situations, where the asymptotic formula is not valid, are investigated separately and an arithmetic progression is constructed for each of them. The third case is of special interest because its analysis shows the possibility of future improvements. Let us outline main points of this thought.

Build injection $A \xrightarrow{\rho} \mathbb{Z}^2$. This map ρ transforms our one-dimensional problem into a two-dimensional one. (Recall that dense two-dimensional problems were solved in **[16]**.) G. Freiman shows that, in the case that the asymptotic formula does not work, there is a rectangle $H \subset \mathbb{Z}^2$ which contains images of most elements of A such that

for this rectangle, the density condition of [16] holds. Thus, subset sums of these elements represent all integer points of a lattice – a two-dimensional analogue of an arithmetic progression. Now, transforming back to one dimension, we get a collection of short arithmetic progressions, the union of which forms a long one.

Finite addition approach

A. Sárkőzy arrives at an arithmetic progression for the same density ($m > c(\ell \log \ell)^{1/2}$) using a different approach. He proves a sequence of theorems that leads to the existence of an arithmetic progression (we formulate his result using the notation of this survey).

Theorem 3.2. — *Let $\ell > 2500$ and $|A| = m > 200(\ell \log \ell)^{1/2}$. Then there are integers d, y, z that*

$$1 \leq d < 10^4 \frac{\ell}{m}, \quad z > 7^{-1} \cdot 10^{-4} m^2, \quad y < 7 \cdot 10^4 \frac{\ell}{m^2} z,$$

and

$$\{M : M \equiv 0 (\mathrm{mod}\, d), yd \leq M \leq zd\} \subseteq A^*.$$

A. Sárkőzy ([22]) shows that this theorem is the best possible apart of the constants and a logarithmic factor in the density constraint. However, the proof of this theorem does not lead to an explicit way of calculating a difference d of an arithmetic progression for a specific instance of a set A.

Algorithmic approach

Z. Galil and O. Margalit ([18]) obtain almost the same density ($m > c\ell^{1/2} \log \ell$) while explicitly constructing a progression. We will discuss this approach in the next section whilst explaining their algorithm.

4. Algorithms for the Subset-Sum Problem based on the structural characterization

This section is dedicated to algorithms for solving SSP. The first algorithm using the structural approach is due to G. Freiman [14]. Using structural characterization from [1] (density $m > \ell^{2/3+\varepsilon}$) this algorithm solves SSP (finding the maximal sum but not the subset) in $O(\frac{\ell^{5/3}}{m} + m \log^2 m)$. In [15] G. Freiman improved this algorithm obtaining a linear time algorithm for the same density of problems. This algorithm also works perfectly for lower density (up to $m > c(\ell \log \ell)^{1/2}$)) but then it is not linear. Its time grows and becomes $O(m^2 / \log m)$ for the lowest density. This algorithm was improved by M. Chaimovich (see [5], [9]) using the same idea but more complicated technique for verifying the divisibility of the summands.

Z. Galil and O. Margalit [18] use another technique. Their algorithm finds both the maximal sum S_B and the subset B. Its time is $O(m)$ for the high density ($m > c\ell^{3/4} \log \ell$) and $O(\ell \log \ell) = O(m^2 / \log m)$ for the lower one ($m \sim \ell^{1/2} \log \ell$). Moreover, this algorithm provides an elementary proof of the structural characterization theorem by explicit construction of the desired structure.

In this survey two algorithms are presented. The first of them ([**15**]) is based on the analytical theorem. We discuss also the methods that were used in [**5**] and [**9**] in order to improve the algorithm. The second algorithm is created by Z. Galil and O. Margalit [**18**]. We will only present descriptions of the algorithms and their estimated complexities (for detailed proofs the reader may refer to cited articles).

4.1. Algorithm for finding the maximal subset sum S_B. — The main idea of the first algorithm is to find the difference Q of the arithmetic progression in the set of subset sums. Based on the analytical theorem (as in Theorem 3.1), finding this difference requires verification of a condition similar to condition (12). Verification is done in the same way as was explained on page 349. In Algorithm 1 condition (14) is used. It may be shown that only prime numbers $q < 3\ell/2m$ must be verified. Once Q is found, elements of A with non-zero residues modulo this Q allow one to "fulfill" the "holes" in the progression and to complete the construction of the structure.

The algorithm does not require that all summands are different but that the amount of the different ones is large enough (see [**5**]). In the algorithm below, the number of different elements in a multi-set X is denoted by $\overline{m}_X$. Recall also that N is the target number and z is the maximal subset sum that does not exceed N.

Algorithm 1

1. Finding Q.
   (a) Initialization. $q_0 \leftarrow 1, A_0 \leftarrow A, t_0 \leftarrow \lfloor \frac{3\ell_A}{2\overline{m}_A} \rfloor, i \leftarrow 0$.
   (b) Find the smallest prime number q_{i+1} such that $2 \le q_{i+1} \le t_i$ and

$$|A_i \setminus A_i(0, q_{i+1})| < t_i. \tag{14}$$

   If such a number q_{i+1} exists, compute $A_{i+1} \leftarrow \frac{A_i(0,q_{i+1})}{q_{i+1}}$, $t_{i+1} \leftarrow \lfloor \frac{3\ell_{A_{i+1}}}{2\overline{m}_{A_{i+1}}} \rfloor$ and continue to next $i (i \leftarrow i+1)$.
   (c) If such a number q_{i+1} does not exist, set $p \leftarrow i$ and compute $Q \leftarrow \prod_{j=0}^{p} q_j$. If $Q = 1$, then go to step 3.
2. Finding C. Let $G = \{b_1, \ldots, b_k\} = A \backslash A(0, Q)$.
   (a) Initialization. $d_0 \leftarrow Q, C_0 \leftarrow \{0\}, i \leftarrow 0$.
   (b) Computing C_{i+1} and d_{i+1}.
   If b_{i+1} is divisible by d_i, then $C_{i+1} \leftarrow C_i$ and $d_{i+1} \leftarrow d_i$.
   Otherwise, compute C_{i+1} explicitly
   $(C_{i+1} = C_i + \{0, b_{i+1}\} (\mathrm{mod}\, Q) = \{s \mid s \in C_i \text{ or } s - b_{i+1} (\mathrm{mod}\, Q) \in C_i\})$.
   If $|C_{i+1}| = |C_i|$ then $d_{i+1} \leftarrow \gcd(d_i, b_{i+1})$; otherwise, set $d_{i+1} \leftarrow d_i$.
   If $|C_{i+1}| = Q$ or $i = k$ go to step 3; otherwise, continue to next i $(i \leftarrow i+1)$.
3. If $Q = 1$ or $|C| = Q$ then $z = \lfloor N \rfloor$, otherwise compute $r = N - \lfloor \frac{N}{Q} \rfloor \cdot Q$ and $z = N - r + \max\{r_i \in C | r_i \le r\}$.

The complexity of Algorithm 1 is $O((\frac{\ell}{\overline{m}})^2 + m \log^2 m)$ which is $O(m \log^2 m)$ for $\ell = O(\overline{m}^{3/2} \log \overline{m})$.

To improve the time boundary of Algorithm 1 to $O(m \log m)$ M. Chaimovich uses condition $|A_i \setminus A_i(0, q_{i+1})| < \left(\prod_{q \le p \le (11\ell)/(8m), p \text{ is prime}} \frac{p}{p-1}\right) \cdot \frac{11\ell}{8m}$ (see [5]) and condition $|A_i \setminus A_i(0, q_{i+1})| < \frac{t_i^{5/3}}{q_{i+1}^{2/3}}$ (see [9]) instead of (14). In both versions gain (comparing with Algorithm 1) is achieved owing to the fact that as soon as prime number q is verified, there is no need to return and to check it again.

To obtain linear complexity (for slightly higher density) the following modification may be used:

Algorithm 1A

1. Let A' be a multi-set consisting of the $\overline{m}_{A'} = \frac{\sqrt{8\overline{m}}}{\log^2 \overline{m}}$ first different elements of A. Find a number Q applying the process from step 1 of Algorithm 1 to set A'.
2. Execute steps 2 and 3 of Algorithm 1.

4.2. Algorithm for finding the optimal subset (Z. Galil and O. Margalit)

Z. Galil and O. Margalit [18] solve the SSP by constructing a long arithmetic progression belonging to the set of subset sums. To do this they partition the input set A into three parts: $A = A_1 \cup A_2 \cup A_3$. First, they construct A_1 – a small set satisfying $(A_1)^*(\mathrm{mod}\, d) = A^*(\mathrm{mod}\, d)$ for every small enough integer d. Set A_2 consists of a number of the smallest elements of $A \setminus A_1$. These elements are used to construct the segment of the progression longer than ℓ. A_3 contains the remaining elements of A. They are used to extend the progression.

The algorithm is based on two main processes. The first of them reduces the problem to the case where $A^*(\mathrm{mod}\, d) = [0, d)$ for every small enough integer d. This constitutes Step 1 of Algorithm 2. Logically this process is similar to the first step of Algorithm 1 and results in the number d_0 such that the set $A' = \frac{A(0,d_0)}{d_0}$ possesses the above mentioned property. The technique used in this algorithm is different than the one used in Algorithm 1. It allows us to obtain the linear time boundary. The same method is employed again in Step 3 when we apply it to set $A'_{(m/4)}$ – the set of $\frac{m}{4}$ smallest elements of A' – in order to construct the subset A'_1.

The second process, used in Algorithm 2, provides a way to construct an arithmetic progression belonging to the set of subset sums. This constitutes Step 5 of Algorithm 2. This process is based on the following simple consideration: Given a set A of μ distinct integers in an interval of length $\lambda \le \ell$, consider the sets P_i of pairs with difference i, i.e., $P_i = \{(a, b) \in A \times A | a - b = i\}$. There are $\Omega(\mu^2)$ pairs (a, b) and thus (by pigeon-hole argument) there are many pairs with the same difference. We first take many P_i's that contain large enough number of pairs. Taking $k - j$ pairs from P_ρ and j pairs from P_σ gives a sequence of subsets $D_j \subset A$ with $S_{D_j} = k\rho + j(\sigma - \rho)$ - an arithmetic progression. (Observe that the pairs in each P_i are disjointed, but the pairs in different P_i's may intersect. So, only some of the pairs from P_i's may be used in our construction, some pairs have to be "deleted" in order to restore the disjointness property.)

The arithmetic progressions generated this way are still too short, but it is possible to generate many of them and then combine them in order to create a longer arithmetic

progression. Starting with the progression of minimal difference, $(i+1)$-st progression is inductively combined with the previously obtained arithmetic progression of the first i progressions. An element of a combined progression is the sum of an appropriate element from each of the two progressions.

As the full description of the process is quite complicated we will omit it in this survey. Thus, we are ready to outline steps of the algorithm.

Algorithm 2

1. Let $t = \lceil \frac{103\ell}{m} \log_2 \ell \rceil$. $A_{(i)}$ denotes the set of i smallest elements of A and q stands for a power of prime. We find an integer $d_0 < t$ such that $|A \setminus A(0, d_0)| < d_0$ and $|A' \setminus A'(0, d)| > d$ for each $d < t$ where $A' = A(0, d_0)/d_0$.
   (a) Compute $G_1 = \{q : 1 < q < t, |A_{(2t)} \setminus A_{(2t)}(0, q)| < t\}$ by verifying all prime powers from 1 to t.
   (b) Compute $G_2 = \{q : q \in G_1, |A_{(b)} \setminus A_{(b)}(0, q)| < t\}$, where $b = 3t \log_2 \ell$ by verifying all elements of G_1.
   (c) Compute $G_3 = \{q : q \in G_2, |A \setminus A(0, q)| < t\}$ by computing $|A(i, \mathrm{lcm}(G_2))|$ for all $i \in [0, \mathrm{lcm}(G_2))$ and using elements of G_2 as candidates for G_3.
   (d) Compute $G_4 = \{d : 1 < d < t, d \equiv 0 (\mathrm{mod}(\mathrm{lcm}(G_3))), |A \setminus A(0, d)| < d\}$ using elements of G_3 as candidates for G_4.
   (e) Compute $d_0 = \max(G_4)$.
2. Use dynamic programming modulo d_0 to compute $A^*(\mathrm{mod}\, d_0)$. In computing the set $A^*(\mathrm{mod}\, d_0)$ keep a subset $C_i \subseteq A \setminus A(0, d_0)$ for each $i \in A^*(\mathrm{mod}\, d_0)$ such that $S_{C_i} \equiv i (\mathrm{mod}\, d_0)$ and $S_{C_i} < \ell d_0$. Also compute the function $f_{d_0}(i) = \max\{j | 0 \leq j \leq i \text{ and } j \in A^*(\mathrm{mod}\, d_0)\}$. (The use of this function will be clarified in step 9.)
3. Reduce the problem to another one by taking $A' = \frac{A(0,d_0)}{d_0}$ instead of A. Apply sub-steps (a)-(d) of Step 1 of the algorithm to $A'_{(m/4)}$ (the first smallest $\frac{m}{4}$ elements of A') and construct $A'_1 = A'_{(m/4)} \cup (\cup_{d \in G'_4} C'_d)$ where G'_4 is the set obtained in sub-step (d) of the second application of Step 1 and C'_d are d elements from $A' \setminus A'(0, d)$.
4. Defining $\lambda = \lceil 64\ell^{1/2} \log_2 \ell \frac{4S_{A'}}{m^2} \rceil$ and $\mu = \lceil 15\ell^{1/2} \log_2 \ell \rceil$ choose $A'_2 \subseteq A' \setminus A'_1$ which contains μ elements where each one is less than $\frac{4S_{A'}}{m}$ and lies in a sub-interval of length λ. This is done by taking elements of $A' \setminus A'_1$ smaller than $\frac{4S_{A'}}{m}$, splitting them into sub-intervals of length λ and choosing the most dense sub-interval.
5. Using the elements of A'_2 obtain the sequence of subsets $\{B'_i\}_{i=0}^{2\ell-1}$ such that their sums form an arithmetic progression with a small difference – $S_{B'_i} = s_0 + ig_r$, $g_r < t$. (A detailed description of the process may be found in section 4 of [**18**].)
6. Using dynamic programming, build a sequence $\{E'_i\}_{i=0}^{g_r-1}$ of subsets of A'_1 such that $S_{E'_i} \equiv i (\mathrm{mod}\, g_r)$ and $S_{E'_i} < \ell' g_r$.
7. Construct sets $F'_i = E'_{i(\mathrm{mod}\, g_r)} + B'_j$, where $j = \ell' + (i - S_{E'_{i(\mathrm{mod}\, g_r)}})/g_r$, for $0 \leq i < \ell g_r$. (Note that $S_{F'_i} = s_0 + \ell' g_r + i$.)
8. Compute all the prefix sums of the set $A'_3 = A' \setminus (A'_1 \cup A'_2)$.
9. Given a target number N, the following sub-steps are executed.

(a) Denoting $r_0 = f_{d_0}(N(\mathrm{mod}\, d_0))$ (for definition of $f_{d_0}(i)$ see step 2), compute $S_B = d_0\lfloor N/d_0 \rfloor + r_0$.
(b) Compute $N' = (S_B - S_{C_{r_0}})/d_0 - s_0 - \ell' g_r$.
(c) Using a binary search on the set of prefix sums of A'_3 find

$$n = \max\{i | S_{A'_{3(i)}} \leq N'\}.$$

(d) The desired subset is $B_? = C_{r_0} \cup d_0 A'_{3(n)} \cup d_0 F'_{N'-S_{A'_{3(n)}}}$.

Observe that the first eight steps of the algorithm are preprocessing steps that may be performed only once in the case that SSP is solved many times for the same set A and different target numbers.

As mentioned before, Algorithm 2 finds the maximal sum and the optimal subset in $O(m + (\frac{\ell}{m}\log\ell)^2 + \frac{S_A \ell^{1/2}}{m^2}\log^2\ell)$ time. (Note that the last term of the expression is required for finding the optimal subset only.) This gives $O(m)$ time for $m > c\ell^{3/4}\log\ell$ and $O(\ell\log\ell)$ time for $m \sim c\ell^{1/2}\log\ell$.

5. Application of an analytical structural approach to other IP models

5.1. Value-Independent Knapsack. — The Value-Independent Knapsack Problem (VIKP) is IP problem of the form: maximize $z = \sum_{i=1}^r a_i x_i$ subject to $\sum_{i=1}^r a_i x_i \leq N$, where $0 \leq x_i \leq n_i$, $x_i \in \mathbb{Z}$, $i = 1, 2, \ldots, r$, and all coefficients are integers (see [**2**]). One can reformulate the VIKP as the SSP with a multi-set A, containing element a_i exactly n_i times (for each i, $1 \leq i \leq r$), and a target number N.

In view of the fact that structural analysis of the SSP was done assuming that the elements are distinct ([**15**], [**18**]) or assuming that the number of distinct elements is sufficiently large ([**5**]), the VIKP requires its own structural analysis. This analysis was done in [**4**] and [**6**] proving the structural characterization of the VIKP for $\ell = O(\frac{m^{3/4}}{\log m})$ and $\ell = O(\frac{m}{\log m})$ respectively. In this survey we formulate the structural characterization as it was done in [**6**] and give a short sketch of the algorithm presented there.

Structural Characterization. — For convenience, we will view A as at a set of pairs of positive integers such that the elements of the pair are an integer and the number of its appearances in A respectively. Thus, we write $A = \{(a'_i, n_i) \mid 1 \leq i \leq r\}$, where $\{a'_1, \ldots, a'_r\}$ is the set of distinct elements of A and n_i is the numbers of appearances of a'_i in A. Define also $t = \max\{n_i \mid 1 \leq i \leq r\}$.

Using this notation, the existence of an arithmetic progression in the set of subset sums was proved in [**6**] for $m \geq \min\{6\ell\log\ell, 9(\ell t)^{2/3}\log^{1/3}(\ell t)\}$.

Indeed, the estimation $m > 6\ell\log\ell$ is the best possible apart from a logarithmic factor and a constant: Let $A = \{(\ell + 1 - i, t) \mid 1 \leq i \leq r\}$ for some integers ℓ, t, r. Clearly, $m = |A| = rt$. A^* consists of rt disjoint intervals (each of which is not longer than $\frac{r^2 t}{2}$) whenever $\frac{r^2 t}{2} \leq \ell - r$, i.e., $m^2 + m \leq \ell t$. We therefore do not have a long arithmetic progression for $m \asymp (\ell t)^{1/2}$ and for $m \asymp \ell$ when $t \asymp \ell$.

Sketch of the algorithm. — One can see two important parts in each of the algorithms (see Algorithm 1) based on the new approach. The first part finds the difference Q of an arithmetic progression in A^*, and the second explicitly constructs subset sums with non-zero residues mod Q, i.e., $A^*(\mathrm{mod}\, Q)$. The same is true for VIKP. In this survey only the main considerations of the algorithm will be presented. Details can be found in [**6**] or in [**9**].

The step that finds the difference of the progression employs two ideas in order to reduce the number of operations of the algorithm. First, the elements of A are grouped in order to present A as a list of pairs (a', n) - an element and the number of its appearances in A. These pairs are sorted such that the most frequent elements appear first.

Second, the difference Q of the progression is found using three different methods depending on the number r of distinct summands in A. If this number is large enough, the method similar to Step 1 of Algorithm 1 is used. Otherwise, Q is determined as the greatest common divisor of the k most frequent elements of A (elements that appear more than $\frac{2\ell}{k} + 1$ times each).

Construction of the subset sums with non-zero residues modulo Q is done using the same technique as in Algorithm 1.

Precise analysis of the steps of the algorithm shows that it carries out the solution in $O(Qr_Q + m)$ time, where r_Q is the number of different residues mod Q of A. In the worst scenario, the first term of the expression dominates and, taking into account that the number of different residues mod Q is limited by the number of different elements of A, we have a $O(\ell^{3/2} \log^{1/2} \ell + m)$ time algorithm.

However, the algorithm becomes linear if (a) $r > c(\ell \log \ell)^{1/2}$ or $r = O(\frac{m}{\ell})$; (b) $k \asymp r$, where k is the number used for calculation of Q (this condition means that there are not many elements with very small number of appearances implying $Q = O(\ell^{1/2})$). Therefore, linear time is not achieved for the following special case: The number of different elements of A is neither large nor very small and all elements with a large number of appearances belongs to one arithmetic progression with a sufficiently large difference. The number of these elements is extremely small in relation to the number of elements with a small number of appearances.

5.2. Multi-dimensional Subset-Sum Problem. — This paragraph is concerned with the multi-dimensional Subset-Sum problem which is a particular case of the multi-dimensional Knapsack Problem. Recall its definition ([**8**]): Let $\mathcal{A}$ be a set of n-dimensional non-zero integral vectors taken from the convex hull $\mathcal{D}$, i.e.,

$$\mathcal{A} = \{\overline{a}_i = (a_{1i}, \ldots, a_{ni})^t\}_{i=1}^m \subseteq ((\mathbb{Z}^n \cap \mathcal{D}) \setminus \{\overline{0}\}).$$

(The notation $(\cdot)^t$ means the transpose of a vector $(\cdot)$, i.e., $\overline{a}_i$'s are viewed as column-vectors.) The problem is: for the given target vector $\overline{b} \in \mathbb{N}^n$, find the vector $\overline{z} \in \mathcal{A}^*$ satisfying $\overline{z} \leq \overline{b}$ and having maximal length, where a partial order on n-dimensional vector space is defined in any appropriate way.

The two-dimensional SSP was investigated by G. Freiman in [**16**]. It has been found that in this case a lattice becomes a basic element of the structure and takes the place

of an arithmetic progression – a basic element of the structure in the one-dimensional case. Further, this result was extended to n dimensions by M. Chaimovich [8].

Two-dimensional SSP. — Let $\Gamma_U = \{\overline{v} \mid \overline{v} = k_1\overline{u}_1 + k_2\overline{u}_2, k_j \in \mathbb{Z}, \overline{u}_j \in U\}$ denote the lattice generated by the set $U = \{\overline{u}_1, \overline{u}_2\} \subseteq \mathbb{Z}^2$ of linearly independent integral vectors. Hereafter the subscript is omitted whenever it is clear from the context which lattice is being considered and let V_Γ denote the number of integer points in the fundamental parallelogram of Γ.

Two vectors $\overline{v}_1, \overline{v}_2$ are congruent modulo Γ (written as $\overline{v}_1 \equiv \overline{v}_2 (\mathrm{mod}\, \Gamma)$) if $\overline{v}_1 - \overline{v}_2 \in \Gamma$. Two sets are congruent modulo Γ (written $\mathcal{A}_1 \equiv \mathcal{A}_2 (\mathrm{mod}\, \Gamma)$) if for each vector $\overline{v}_1 \in \mathcal{A}_1$ there is a vector $\overline{v}_2 \in \mathcal{A}_2$ congruent to $\overline{v}_1$ and inversely for each vector $\overline{v}_2 \in \mathcal{A}_2$ there is a vector $\overline{v}_1 \in \mathcal{A}_1$ congruent to $\overline{v}_2$. In addition, $\mathcal{A}(\Gamma) = \mathcal{A} \cap \Gamma$, and $\overline{b} \in \mathcal{A} (\mathrm{mod}\, \Gamma)$ means that there is $\overline{v} \in \mathcal{A}$ congruent to $\overline{b}$.

For a given $\mathcal{A}$ define $B_j^2 = \frac{1}{4}\sum_{i=1}^m a_{ji}^2$, $j = 1, 2$; $B_{12} = \frac{1}{4}\sum_{i=1}^m a_{1i}a_{2i}$.

Using this notation the following theorem (Theorem 2 [16]) gives a structural characterization for a two-dimensional case.

Theorem 5.1. — *Let $\mathcal{A} \subseteq D \cap \mathbb{Z}^2$ be a set of two-dimensional integral vectors where D is a convex set with $|D \cap \mathbb{Z}^2| = \ell$, $|\mathcal{A}| \geq c_1 \ell^{2/3} \log^{1/3} \ell$, $\ell > \ell_0$. Suppose that for each line "a" containing zero*

$$|\mathcal{A} \cap a| \leq \tfrac{1}{2}|\mathcal{A}|. \tag{15}$$

Then (i) there is the lattice Γ_0 with $V_{\Gamma_0} = O(\frac{\ell}{m})$ and the subset $H \subseteq \mathcal{A}$ such that $|H| \leq V_{\Gamma_0}$ and $\mathcal{A}^* \equiv H^* (\mathrm{mod}\, \Gamma_0)$ and (ii) for convex hull $\mathcal{F}$ defined by

$$\mathcal{F} = \{\overline{v} : |(\tfrac{1}{2}S_{\mathcal{A}} - \overline{v})^t \begin{pmatrix} B_1^2 & B_{12} \\ B_{12} & B_2^2 \end{pmatrix}^{-1} (\tfrac{1}{2}S_{\mathcal{A}} - \overline{v})| \leq c_2\},$$

vector $\overline{b}$ belongs to $\mathcal{A}^* \cap \mathcal{F}$ if and only if $\overline{b} \in \mathcal{F}$ and $\overline{b} \equiv H^* (\mathrm{mod}\, \Gamma_0)$.

According to this theorem, the structural characterization of the set of two-dimensional subset sums is quite simple: a collection of all points from certain classes of residues modulo lattice (including zero residue class) within a two-dimensional convex hull in the wide vicinity of the mid-point $\frac{1}{2}S_{\mathcal{A}}$.

The proof of this result is too complicated to be presented in this survey. First of all, the case where all vectors are taken from the rectangle with edges parallel to axes is investigated and for this case the asymptotic formula for the number of representations by the set of two-dimensional subset sums is obtained. In this step the condition for validity of the asymptotic formula is determined: not too many vectors may belong to a one integer lattice.

Further, this result is extended replacing a rectangle by an arbitrary convex set $\mathcal{D}$. This is done by applying to the set $\mathcal{D}$ a certain transformation which is invariant with regard to the integer lattice. In addition, the image of $\mathcal{D}$ is contained in the rectangle and the number of integer points in this rectangle is of the same order as in $\mathcal{D}$.

Finally, the case where the asymptotic formula is not true or, in other words, where most of the vectors belong to some lattice Γ is considered. This is done in the similar way as in the one-dimensional case.

Observe that condition (15) is crucial for obtaining two-dimensional structural characterization. If this condition is not satisfied, the problem is actually one-dimensional because most of its vectors lie on one line.

The last step of the proof provides a simple algorithmic way to construct the structure and to find the solution to the problem. Precise analysis of the algorithm (not presented here) shows that its time boundary is $O(m^2 \log m)$. For very dense problems ($\ell = O(m)$), the time boundary of the new algorithm is more impressive. It is $O(m \log^2 m)$ – almost linear.

n-dimensional SSP. — Analysis of the n-dimensional SSP is quite similar to the two-dimensional case. The difficulty in the generalization lies in the complexity of the geometry of an arbitrary number of dimensions compared to the geometry of two dimensions. However, the structural characterization of the set of n-dimensional subset sums, explicitly determined by the algorithm, seems to be quite simple: it consists of a collection of all points with certain classes of residues modulo lattice within an n-dimensional convex body.

The density condition for the n-dimensional case is $m \geq (n\ell^{n-1} \log \ell)^{1/n}$, $n > 2$, and the time boundary of the algorithm becomes $O(m^{2+1/(n-1)} \log^n m)$ or even $O(m \log^n m)$ for very dense problems ($\ell = O(m)$).

5.3. The k-Partition Problem. — A structural approach for solving the k-partition problem (KPP) was studied in [**7**] (see page 346 for problem definition). Although the proposed method works for a wide spectrum of objective functions, the author chooses as an objective function the function $z = \max\limits_j \frac{S_{B_j}}{N_j}$. Under this objective function the problem can also be viewed as a problem of scheduling independent tasks on uniform machines so as to minimize an end (make-span) time (see [**21**] for scheduling problem definition).

The solution is based on the reduction of the k-partition problem to a sequence of dense SSP and on the structural characterization of SSP by a collection of arithmetic progressions. As a result, the proposed algorithm solves the problem in $O(k\ell \log \ell)$ time which is considerably faster than previously known polynomial algorithms (dynamic programming, [**21**]) that achieved $O(m^{2k-1}\ell)$ time only.

In this survey general concepts of the reduction process and of the algorithm are presented.

General concepts. — Let $z^* = \min z$ be a value of the objective function for the optimal partition $(B'_1, \ldots, B'_k)$, i.e., $z^* = \max\limits_j \frac{S_{B'_j}}{N_j}$. If $z^* = 1$ we say that KPP is *exactly solvable.* This is equivalent to the existence of a partition $(B_1, \ldots, B_k)$ with $S_{B_j} = N_j$ for all j, $1 \leq j \leq k$.

Suppose that a KPP $(A, N_1, \dots, N_k)$ has an exact solution. Consider the sequence of the following $(k-1)$ Subset-Sum problems:

$$(A, N_1, S_A - N_1), (A \backslash B_1, N_2, S_{A\backslash B_1} - N_2), \dots,$$
$$(A \backslash \bigcup_{i=1}^{k-2} B_i, N_{k-1}, S_{A\backslash \cup_{i=1}^{k-2} B_i} - N_{k-1}),$$

where B_i is some solution of the i-th SSP. Assuming that the first SSP is already solved and $S_{B_1} = N_1$, it is not necessarily true that the remaining SSPs are still exactly solvable. This is because elements which are necessary to find an exact solution for the second SSP could already have been used in B_1 – the solution for the first problem. In other words, the solution B_1, which was chosen from the set of all possible exact solutions of SSP, can be "bad"; thus the rest of SSP will not be exactly solvable.

To overcome this difficulty, a certain subset $C \subset A$, for which SSP $(C, N_1, S_C - N_1)$ has an exact solution, is defined. This subset is created such that selection of any one of these solutions ensures the existence of exact solutions for all subsequent problems. In that way, KPP can be replaced by solving a sequence of SSPs.

Some conditions must be imposed on multi-set A in order to ensure successful application of this method. Recall that an exact solution of SSP in a wide interval of target numbers N is ensured by condition (8), i.e., we have "many" non-zero residues for each modulo q. To solve KPP, it is natural to strengthen condition (8), requiring as many non-zero residues for each modulo q as we need for exact solvability of all $(k-1)$ SSP: the condition (8) becomes

$$|A \setminus A(q)| \geq (k-1)^2 \frac{4\ell_A}{\overline{m}_A} \log_2 2\ell_A, \tag{16}$$

where $\overline{m}_A$ again stands for the number of different elements of A.

Indeed, multi-set C, mentioned above, and from which subset B_1 is chosen, includes the amount of non-zero residues needed to solve one problem only, leaving the rest to be used when solving subsequent problems.

In fact, in addition to condition (16) the density relation

$$\overline{m} > c_1 (k(k-1)\ell \log \ell)^{2/3}, \tag{17}$$

must be imposed on A in order to ensure the possibility of the reduction. Condition (17) and the trivial inequality $\ell \geq \overline{m}$ restrict the values of k for which the class of KPP, solved using the above method, is not empty. Namely, $k < \frac{\ell^{1/4}}{c_1^{3/2} \log^{1/2} \ell}$.

The situation where condition (16) fails for some integer q, can be viewed in a similar way to the way it was handled in the case of SSP. It can be shown that there exists an integer q_0 such that multi-set $A' = \frac{A(0,q_0)}{q_0}$ satisfies conditions similar to conditions (17) and (16), and that KPP $(A(0,q_0), N_1', \dots, N_k')$, where $q_0 | N_i'$, has an exact solution.

Introduce the set Q_{q_0} of k-tuples $(s_1, \dots, s_k)$ of residues modulo q_0 which can be represented by $A \backslash A(0,q_0)$: for each $(s_1, \dots, s_k) \in Q_{q_0}$ there is a partition $(G_1, \dots, G_k)$ of $A \backslash A(0,q_0)$ such that $s_j \equiv S_{G_j} (\mathrm{mod}\, q_0)$, $0 \leq s_j < q_0$.

Combining the solution of the above mentioned KPP $(A(0,q_0),N'_1,\ldots,N'_k)$ and k-tuples from Q_{q_0}, it may be concluded that the original KPP is exactly solvable if and only if there is a k-tuple $(n_1,\ldots,n_k) \in Q_{q_0}$ such that $n_j \equiv N_j (\mathrm{mod}\, q_0), 0 \leq n_j < q_0, 1 \leq j \leq k$.

Thus, to determine if KPP is exactly solvable, it is sufficient to find q_0, and also to verify that the k-tuple of residues of target numbers modulo q_0 can be represented by a partition of $A \setminus A(0,q_0)$.

Finally, it is necessary to describe how to find the solution of a KPP, that is not exactly solvable. For each partition $(B_1,\ldots,B_k)$ of A, we have $S_{B_j} = N_j + d_j$ and $z = \max\limits_j \frac{S_{B_j}}{N_j} = \max\limits_j (1 + \frac{d_j}{N_j})$. The goal is to find a certain set of deviations $\{d^*_j\}$ which will minimize the objective function z.

For $(s_1,\ldots,s_k) \in Q_{q_0}$ define

$$(18)\quad z(s_1,\ldots,s_k) = \min\left\{ \max_j (1+\frac{d'_j}{N_j}) \,\middle|\, \sum_{j=1}^{k} d'_j = 0, d'_j + N_j \equiv s_j (\mathrm{mod}\, q_0), 1 \leq j \leq k, \right\}$$

where the minimum of the function is taken over all possible sets $\{d'_j\}$. It is shown in [7] that an optimal set of deviations $\{d^*_j\}$ is the set which minimizes this function $z^* = \min\{z(s_1,\ldots,s_k) \mid (s_1,\ldots,s_k) \in Q_{q_0}\}$.

Once we have this set of deviations, we obtain the KPP $(A, N_1 + d^*_1, \ldots, N_k + d^*_k)$ which has an exact solution. To solve this problem, we construct a new problem $(A', N'_1, \ldots, N'_k)$ where $A' = \frac{A(0,q_0)}{q_0}$ and $N'_j = \frac{N_j + d^*_j - S_{G_j}}{q_0}$ and solve it by solving $k-1$ subsequent SSPs. (Algorithm 2 from page 354 may be used to solve each SSP.)

Let $(G'_1,\ldots,G'_k)$ be a solution of this KPP. Then $(G_1 \cup q_0 G'_1, \ldots, G_k \cup q_0 G'_k)$ is a solution of the original KPP, since $S_{G_j \cup q_0 G'_j} = S_{G_j} + q_0 N'_j = N_j + d^*_j$.

The complexity of the algorithm. — The complexity of the algorithm is evaluated (details can be found in [7]) as

$$(19)\qquad O(k^3 \frac{\ell}{\overline{m}} q_0^{k-1} \log \ell + k\ell \log \ell).$$

Indeed, for $q_0 = O(\frac{\ell}{\overline{m}})$ and $\ell = O(\frac{m^{3/2}}{k^2 \log \overline{m}})$ (see (17)), the first term in (19) dominates the second one and the time is $o(\overline{m}^{k/2})$. However, in the case where $q_0 < \left(\frac{\overline{m}}{k^2}\right)^{1/(k-1)}$, the dominant term in (19) is the second term and we obtain an almost linear time algorithm with $O(k\ell \log \ell)$ time. This time remains polynomial, even if k increases with ℓ. Observe also that $q_0 < \left(\frac{\overline{m}}{k^2}\right)^{1/(k-1)}$ is always satisfied for a dense 3-partition problem ($k = 3$) and for problems with sufficiently high density, namely, for $\overline{m} > k^{2/k} \ell^{1-1/k}$.

6. Conclusion

There are several other directions which deserve to be explored in order to proceed with this new approach.

One of them is to study structural characterization of subset sums for problems with lower density. G. Freiman conjectures that analytical techniques can be refined to handle the case $\ell = O(m^c)$ for any constant c.

This characterization will allow us to derive new algorithms. Recall that the structural approach, contrary to classical methods, works for problems with a large number of variables. There is, therefore, a gap between an upper boundary of classical algorithms and a lower boundary of the existing algorithms based on the new approach. The purpose of an algorithmic design, from the operational point of view, is to overlap this gap. From the theoretical point of view, the future algorithms will allow us to verify the conjecture: is it true that certain IP problems that are NP-hard have a less than exponential time solution for dense instances?

The other direction of the development is to analyze additional IP problems and to extend new methods to them. These efforts can proceed in two ways. One is to work directly on other specific problems and try to characterize their structure. The other is to reduce a problem to the SSP (one or multi-dimensional) as was done for the k-partition problem. In order to do this we need density-preserving reductions that yield instances of the SSP that are sufficiently dense.

References

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

[2] Balas E. and Zemel E., *An Algorithm for Large Zero-One Knapsack Problems*, Operations Research, **28**, 1980, 1130–1154.

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

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

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

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

[7] Chaimovich M., *Fast Exact and Approximate Algorithms for k-Partition and Scheduling Independent Tasks*, Discrete Mathematics, **114**, 1993, 87–103.

[8] Chaimovich M., *On Solving Dense n-Dimensional Subset-Sum Problem*, Congressus Numerantium, **84**, 1992, 41–50.

[9] Chaimovich M., *Analytical Methods of Number Theory in Integer Programming*, Ph. D. Thesis, Tel-Aviv University, Israel, 1991.

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

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

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

[13] Freiman G.A., *What is the Structure of K if $K + K$ is Small?*, in Lecture Notes in Mathematics, **1240**, 1987, 109–134.

[14] Freiman G.A., *On Extremal Additive Problems of Paul Erdős*, ARS Combinatoria, **26B**, 1988, 93–114.

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

[16] Freiman G.A., *On Solvability of a System of Two Boolean Linear Equations*, The Proceedings of the Number Theory Conference, New York, 1989.

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

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

[19] Jeroslow R.G., *Trivial Integer Programs Unsolvable by Branch and Bound*, Mathematical Programming, **6**, 1974, 105–109.

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

[21] Sahni S.K., *Algorithms for Scheduling Independent Tasks*, J. ACM, **23**, 1976, 116–127.

[22] 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