Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

A REFINED CLASSIFICATION OF THE SUBSET SUM PROBLEM IN POLYNOMIAL TIME COMPLEXITY

B. SINCHEV[1], A.B. SINCHEV[2], A.M. MUKHANOVA[3]
[1] International University of Information Technology, Almaty, Kazakhstan
E-mail: b.sinchev@iitu.edu.kz
[2] National Information Technologies JSC, Astana, Kazakhstan
E-mail: askar.sinchev@nitec.kz
[3] Kainar Academy, Almaty, Kazakhstan
E-mail: nuraksulu72@mail.ru

Abstract

The problem considered is the selection of at least one subset from a set (array) of distinct positive integers, such that the sum of the subset's elements exactly matches a given target sum (target certificate). According to R. M. Karp, this problem belongs to the class of NP-complete problems. Diophantine equations and an auxiliary problem, which facilitates the solution of the original problem and has independent scientific interest, have been introduced. A novel method has been developed, which includes proven lemmas and theorems. These results enable the development of efficient and straightforward algorithms for solving the subset sum problem. The time and space complexity for selecting the required subsets do not exceed the square of the length of the original set. An analytical framework has been proposed for managing indices within the original set. These algorithms are applicable to solving problems related to the independent set of cardinality *k* and the *k*-vertex cover problem. Additionally, we present examples to confirm claimed results.
It should be noted that the time complexity of sorting an array of integers is proportional to the square of the array's size, and this problem belongs to class P. Therefore, based on the newly developed method, it can be inferred that the subset sum problem, originally classified as NP-complete within the NP class, also belongs to P.

Key words: P, NP-complete class, set, subset, time, space, algorithm, index management system

AMS subject classifications. 11Y16, 68W10, 68Q25

Introduction

In 1971, S. Cook [1] introduced the NP-complete class, which included the Boolean circuit satisfiability problem, the Boolean formula satisfiability problem, and the conjunctive normal form (CNF) satisfiability problem. In 1972, R. M. Karp, in his work [2] proved the NP-completeness of the *k*-independent set problem in directed graphs, the *k*-vertex cover problem, the Hamiltonian path problem in directed graphs, and the subset sum problem. The class NP contains problems that can be

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

verified quickly; in other words, there exists an algorithm that can verify a positive solution to the problem in polynomial time. It has been shown that within the NP class, there exist "hardest" problems, meaning those to which all other NP problems can be reduced. This implies that if even one of these hardest problems can be solved in polynomial time, then every problem in NP can be solved in polynomial time, thus establishing that P=NP. The Boolean circuit satisfiability problem [1], the independent set problem [2], the vertex cover problem [2], the Hamiltonian path problem [2], the subset sum problem [2], and many others belong to the class of NP-complete problems. In general, the NP-completeness of these problems has been proven through reductions involving Boolean circuits, Boolean formulas, conjunctive normal forms, and graphs. It should be noted that the proof of the NP-completeness of the aforementioned problems also includes algorithms for their solution. However, in practical terms, these algorithms are challenging to implement. Therefore, this work is a continuation of the research presented in [3], which investigates an alternative approach to studying combinatorial problems in the NP-complete class, based on set theory, Diophantine equations, and properties of arithmetic means.

Currently, the primary methods for investigating this problem include exponential algorithms that examine subsets of the original set with cardinalities ranging from one to the length of the input [4,5], and the brute force method [6], which explores subsets with cardinalities from one to a constant value that is smaller than an independent of the input length. The computational complexity of the subset sum problem depends on two parameters: the cardinality *n* and the precision *p* (defined as the number of binary digits in the numbers comprising the set).

In [7], this constant and the algorithm execution time were reduced by half, while [8] established an upper bound on the cardinality of the selected subset. A US patent [9] has been acquired. Worth to note that these parameters have been leveraged to improve the efficiency of pseudo polynomial algorithms [10-12]. Therefore, the key factor for the time complexity is the cardinality of the original set and the specific subsets required to solve the given problem.

Problems' statement

Reduction of the subset sum problem to the auxiliary problem

A sorted set of distinct positive integers $X^n = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ with cardinality $n = |X^n|$, and an integer S^k are given. It is required to determine whether it is possible to select at least one subset X^k with cardinality $k = |X^k|$ such that the sum of its elements exactly equals the target certificate S^k.

Henceforth, the superscript of all variables and other quantities will correspond to the cardinality of the set X^n or the subset X^k.

Then, the formal statement of the subset sum problem in parameterized form is as follows:

$$S^k: \exists X^k \subseteq X^n, \sum_{x_i \in X^k} x_i = S^k, \quad (1)$$

where $k = 2m \vee 2m+1, k \leq n,$ for even $(k = 2m)$ and odd $(k = 2m+1)$.

Subsets X^k are selected based on the combination function:

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

$$C_n^k = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}. \quad (2)$$

It should be noted that, in addition to the selected subsets X^k, there remain unselected subsets X^{n-k} with cardinality $n-k = |X^{n-k}|$. By the symmetry property $C_n^k = C_n^{n-k}$ it follows that:

$$X^{n-k} = X^n \setminus X^k. \quad (3)$$

Based on the properties of the combination function(2), we will find the partial sums $z_k = \sum_k x_k,\ x_k \in X^n$ and the discrete range:

$$z_k \in [z_{min}^k, z_{max}^k], \quad (4)$$

where $z_{min}^k = \sum_{i=1}^k x_i$, $z_{max}^k = \sum_{i=n-k+1}^n x_i$, $x_i \in X^n$.

Let's introduce the set of consecutive natural numbers N^n ={1,2,3,…,n} with cardinality $n = |N^n|$ (without loss of generality, zero can be included in the set N^n, resulting in N^n ={0,1,2,3,…,n-1}). Then, the parameterized formulation of the subset sum problem for $N^k \subseteq N^n$ with cardinality $k = |N^k|$ and a given index certificate s^k is as follows:

$$s^k: \exists N^k \subseteq N^n, \sum_{n_i \in N^k} n_i = s^k. \quad (5)$$

Note that this set of consecutive natural numbers $N^k \subseteq N^n$ are the indices of the elements of the initial set. Accordingly, s^k is the corresponding index certificate for certificate S^k , i.e., the sum of the indices of the elements in the subset X^k whose sum equals S^k.

The auxiliary problem(5) eliminates the precision parameter p (defined as the number of binary digits in the numbers comprising the original set) from the computational complexity of problem(1), thereby simplifying the solution of problem(1). The auxiliary problem (5) is of independent scientific interest. Elements of the subset N^k are determined based on the combination function(2). Each subset N^kconsists of k elements from the set N^n.

Therefore, we find the values $s_{min}^k = \sum_1^k n_i, n_i \in N^n$, $s_{max}^k = \sum_{n-k+1}^n n_i$, $n_i \in N^n$. The possible range of the index certificate s^k, corresponding to a subset from subsets $N^k \subseteq N^n$ is:

$$s^k \in [s_{min}^k, s_{max}^k], \quad (6)$$

which is equivalent to: $s_i^k \in \{ s_{min}^k,\ s_{min}^k + 1, \dots, s_{max}^k \}$.

Note that the range(6) describes unique index certificates s_i^k. Next, we find the value:

$$m^k = s_{max}^k - s_{min}^k + 1 = kn - \frac{(k-1)k}{2} - \frac{k(k+1)}{2} + 1 = kn - k^2 + 1. \quad (7)$$

Equation(7) defines only the number of unique index certificates: $s_i^k,\ i = 1,2,\dots,m^k$.

***Lemma*1.** Let the certificate S^k of problem(1) belongs to the discrete range(4). Then, there exists at least one subset $N^k \subseteq N^n$ with cardinality k and an index certificate s^k such that the auxiliary problem(5) is solvable.

Proof. Satisfying the first condition of the lemma implies that the sum of the k elements of the subset X^k with cardinality k in problem(1) is determined as follows:

$$x_i + x_j + \dots + x_m + x_l = S^k, i \neq j \neq \dots \neq m \neq l;\ (x_i, x_j, \dots, x_m, x_l) \in X^k \subseteq X^n, \quad (8)$$

Since the certificate S^k belongs to the discrete range(4) and all indices $i \neq j \neq \dots \neq m \neq l$ of the elements of the subset $X^k \subseteq X^n$ are chosen based on the combination function(2), we have,

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

$\sum_{x_i \in X^k} x_i = S^k$. From equation(8), the index certificate s^k for the auxiliary problem(5) is easily determined as follows:

$$n_i + n_j + \cdots + n_m + n_l = s^k, \quad (9)$$

where in case of a sorted set N^n, the indices can be written as:

$i = n_i, j = n_j, \dots, m = n_m, n_l = l, (n_i, n_j, \dots, n_m, n_l) \in N^k \subseteq N^n$.

Solving the Diophantine equation(9) with a given index certificate s^k allows us to find the indices of the subset N^k with cardinality k. These indices will coincide with both the indices of the elements in the subset X^k and the indices of the variables in equation(9). Thus, the solvability of the original problem(1) implies the solvability of the auxiliary problem(5), specifically, the fulfillment of the condition: $\sum_{n_i \in N^k} n_i = s^k$. Conversely, if the sum of the elements of a subset X^k is exactly S^k, this implies that the subset contains exactly k numbers corresponding to the indices of the subset N^k and k variables in equation(8).

Solution approaches

***Lemma*2.** Let k=2 and $S^k \in [z^k_{min}, z^k_{max}]$. Then the time to select the subset $X^k$and the required space satisfy the following conditions:

$$T \leq O(kn) < O(n^2), \ \mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right). \quad (10)$$

Proof. Based on the combination function(2), subsets X^k with cardinality $k = 2$ are represented as a two-dimensional triangular array of order (n-1)x(n-1):

$$X^2 = \begin{Bmatrix} x_1 + x_2 & x_1 + x_3 & \dots & \dots & x_1 + x_{n-1} & x_1 + x_n \\ & x_2 + x_3 & x_2 + x_4 & \dots & x_2 + x_{n-1} & x_2 + x_n \\ & & \dots & \dots & \dots & \dots \\ & & & x_{n-2} + x_{n-1} & x_{n-2} + x_n & \\ & & & & x_{n-1} + x_n & \end{Bmatrix}. \quad (11)$$

Here, each subset X^2 consists of two elements: $X^2 = \{x_i, x_j\}$. *Algorithm for generating the array* (11): It is sufficient to add the element x_1 to the elements of the set X^n (which is represented as a one-dimensional array), starting from the second element, resulting in pairs $(x_1 + x_2) \in X^2$ through $(x_1 + x_n) \in X^2$. Then, add the element x_2 to the elements of the set X^n starting from the third element, producing pairs $(x_2 + x_3) \in X^2$ through $(x_{n-2} + x_{n-1} \ \ x_{n-2} + x_n) \in X^2$. Continue this process until the last element is reached, resulting in $(x_{n-1} + x_n) \in X^2$.

The discrete values in range (4) are directly derived from the values of the elements in array (11). The number of elements in array(11) is $C_n^2 = \frac{(n-1)n}{2}$. In particular, $x_{12} = x_1 + x_2, i = 1, j = 2, \dots, x_{ij} = x_{n-1} + x_n, i = n-1, j = n, \ X^2 = \{x_1, x_2\}, \dots, X^2 = \{x_{n-1}, x_n\}$.

Array(11), with respect to the indices i, j of the elements x_{ij} in the subset X^2, is structured as follows:

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

$$N^2 = \begin{Bmatrix} 12 & 13 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots 1\,n-1 & 1\,n \\ & 23 & 24 \dots \dots \dots \dots \dots \dots \dots 2\,n-1 & 2\,n \\ & & \dots \dots \dots \dots \dots \dots \dots \dots & \\ & & n-2 & n-1 \quad n-2\,n \\ & & & n-1\,n \end{Bmatrix}. \quad (12)$$

Here, the indices of the two-dimensional triangular array N^2 are selected from the set of consecutive natural numbers $N^n = \{1,2, \dots, n\}$ with cardinality $n = |N^n|$. It should be noted that there is a one-to-one correspondence between arrays(11) and (12). According to the condition $S^2 \in [z_{min}^2, z_{max}^2]$ stated in Lemma1 the certificate S^2 belongs to the discrete range(4), as function(2) generates all the combinations necessary to form the complete set of subsets X^2. Otherwise, the problem (1) has no solution. This implies that for the given cardinality k there exists an element x_{ij} in array(6) that equals the certificate $x_{ij} = S^2, x_{ij} \in X^2$ and the indices i, j are fixed. For this element, the condition $\sum_{x_i \in X^2} x_i = x_{ij} = x_i + x_j = S^2, (x_i, x_j) \in X^n$, is satisfied. Consequently, the subset sum problem (1) is resolved.

Next, we will prove inequalities (10). Based on array (11), we introduce a two-dimensional triangular array of index certificates:

$$s^2 = \begin{Bmatrix} 1+2 & 1+3 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots 1+(n-1) & 1+n \\ & 2+3 & 2+4 \dots \dots \dots \dots \dots \dots 2+(n-1) & 2+n \\ & & \dots \dots \dots \dots \dots \dots \dots \dots & \\ & & n-2+(n-1) & (n-2)+n \\ & & & (n-1)+n \end{Bmatrix}. \quad (13)$$

From array(13), we extract the unique index certificates:: 3, 4,…, $1+(n-1)$, $1+n, 2+n, \dots, (n-1)+n$.

Thus, this relationship includes the first row and the last column of array(13), as other elements along the diagonal are repeated.

It has been established that problem(1) is solvable and that there exists a subset X^2. This indicates that there exists an element $x_{ij} = x_i + x_j = S^2, x_{ij} \in X^2 \vee (x_i, x_j) \in X^n$ with determined indices i and j. Consequently, the required index certificate is $s^2 = i + j$. These indices represent solutions to the Diophantine equation(9) for finding elements on one of the diagonals of matrix(12).

To find the required subset N^2, it is sufficient to perform a sequential examine of the elements along the identified diagonal of matrix(12). The maximum number of solutions to the Diophantine equation(9) will be less than or equal to $n/2$ - the maximum number of elements on the diagonal of matrix(12) with index certificate $s^2 = 1 + n$ and $T \leq O\left(\frac{n}{2}\right)$. In other words, the time required to select the subset X^2, that describes the subset N^2, is determined by the number of elements on the identified diagonal of matrix(12) for a given index certificate s^2. Let's note that range(6) describes only the unique index certificates $s_i^k = s_{min}^k \vee s_{min}^k + 1 \vee \dots \vee s_{max}^k$. Equation(7) defines only the number of unique index certificates: $s_i^k, i = 1,2, \dots, m^k$. In conclusion, taking into account the derived inequality $T \leq O\left(\frac{n}{2}\right)$, equation(7), the sequential examination of subsets N^2, we have that $T \leq$

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

$O\left(\frac{n}{2}\right) \le O(m^k) = O((n-k)k+1) \le O(kn) < O(n^2),\ \mathbb{S} \le O\left(\frac{(n-1)*n}{2}\right)$. These inequalities define the time complexity and required space in general terms.

In addition to describing the subsets N^k the Diophantine equation (17) also serves as a criterion for the existence of duplicate elements within the subset X^2. Thus, it determines one or more subsets X^k for problem(1).

The subsets X^3 and N^3 can be obtained after applying *Lemma*2 for $k = 3$:

$X^3 = X^2 \cup x_l\ , N^3 = N^2 \cup n_l,\ S^3 = S^2 + x_l,\ s^3 = s^2 + n_l,$
$x_l \in X^n, n_l \in N^n, x_l \notin X^2, n_l \notin N^2.$ (14)

Next, consider problem(1) with the initial set X^n of large cardinality. In this case, the Vandermonde convolution $\sum_{r=0}^{k} C_n^r\, C_m^{k-r} = C_{n+m}^k$ is applicable, and we will present some transformations. Let the initial set be:

$X^n = X^{n_1} \cup X^{n_2}, n_1 + n_2 = n$ (15)

and the problem (1) is then solved.

Note that the following relationships are valid if:

$k = 2m+1 \le n:\ X^k = \bigcup_m X_m^2 \cup x_l, N^k = \bigcup_m N_m^2 \cup n_l,\ S^k = S^{2m} + x_l, s^k = s^{2m} + n_l\ , x_l \in X^n, x_l \notin \bigcup_m X_m^2\ , n_l \in N^n, n_l \notin \bigcup_m N_m^2 \quad \vee$
$k = 2m \le n: X^k = \bigcup_m X_m^2 \setminus x_l, N^k = \bigcup_m N_m^2 \setminus n_l,\ S^k = S^{2m} - x_l, s^k = s^{2m} - n_l\ , x_l \in X^n, x_l \notin \bigcup_m X_m^2\ , n_l \in N^n,\ n_l \notin \bigcup_m N_m^2\ .$ (16)

It is essential to ensure the integrity of the values in all the aforementioned formulas.

Let the cardinality of the subsets X^k and N^k be $k = 2m \le n$. We introduce a Diophantine equation:

$N^2: n_i + n_j = s^2, (\ n_i, n_j) \in N^n$ (17)

and introduce the notations and formulas considering the Diophantine equation (17):

$N_m^2 = \{(i,j) \in N^n, i + j = s^2, s^2 \in [s_{min}^2, s_{max}^2]\}$ (18)

$X_m^2 = \left\{\left(x_i\ , x_j\right) \in X^n, (i,j) \in N_m^2\ , x_i\ + x_j = S^2, S^2 =\ z^2 \in [z_{min}^2, z_{max}^2]\right\},$ (19)

where, from the range(6), we have that the unique index certificate s^2 varies from s_{min}^2 to s_{max}^2 in steps of one. The subsets N_m^2 and X_m^2, consist of two indices $(i,j) \in N^n$ and two elements $(x_i\ , x_j) \in X^n$, respectively. The value m determines the number of subsets N_m^2 of cardinality 2, and this value can take on values $m = 1 \vee 2 \vee \ldots \vee m = k/2$.

A new solution to problem(1) is proposed, based on unique index certificates $s^2 = s_{min}^2 = 1 + 2 \vee s^2 = s_{min}^2 + 1 \vee \ldots \vee s^2 =\ s_{max}^2 = n - 1 + n$ and the Diophantine equation(17).

Theorem1: Let $S^k \in [z_{min}^k, z_{max}^k]$, and set X^n be a set with cardinality $k = 2m$. The subsets N_m^2 , corresponding to each index certificate $s^2 \in [s_{min}^2, s_{max}^2]$, are solutions to the Diophantine equation(17). Then, for a given value of m there exist subsets $N^k = \bigcup_{m=1}^{m} N_m^2$, $X^k = \bigcup_{m=1}^{m} X_m^2$ such that: $\sum_{(x_i,x_j)\in \bigcup_m X_m^2} (x_i + x_j) = S^k \equiv \sum_{x_i \in X^k} x_i$ with S^k falling within the range(4). The time complexity and space requirements are as follows: $T \le O(m^k) \le O(kn) < O(n^2),\ \mathbb{S} \le O\left(\frac{(n-1)*n}{2}\right)$.

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

Proof. The sequential application of relations(18) for each index certificate s^2 and each subset N_m^2, associated with each solution of the Diophantine equation(17) and subset X_m^2 forms the subsets $\bigcup_m N_m^2$, $\bigcup_m X_m^2$ and the following equality holds: $\sum_{(x_i,x_j)\in\bigcup_m X_m^2}(x_i+x_j)=S^k$. Here, the indices of the found subsets N_m^2 are distinct. This property follows from the characteristics of the solutions to the Diophantine equation (17). Due to certificate S^k belongs to the range(4), and considering *Lemma*1 and the last equality, we obtain the solution to the original problem(1). As noted above, sequentially examining the subsets N_m^2 allows us to determine the time and space requirements, which are given by the formulas: $T\le O(m^k)\le O(nk)\le O(n^2),\ \mathbb{S}\le O\left(\frac{(n-1)*n}{2}\right)$.

***Corollary 1*:** 1. If the Diophantine equation (17) is unsolvable for a given index certificate s^k, from the symmetry property of the combination function(2), there will always be a certificate $\sum_{i=1}^n x_i - S^k$ with the possibility of applying the formulas(3). 2. If the relations(16) hold, then the theorem is applicable for odd k when using the estimates $T\le O(m^k)\le O(kn)\le O(n^2)$.

For a clear demonstration of the results of Theorem1, it is sufficient to consider the selection of subsets X^k, N^k with cardinality $k=2$ in problems(1), (5) and their combination $\bigcup_m X_m^2$, $\bigcup_m N_m^2$ when taking into account the relations(16).

The theorem is valid for odd k. For an odd $k=2m+1$, the certificate S^3 is defined as $S^3=S^2+x_l$, where $x_l\in X^n, x_l\notin X^k$ is added to the certificate $\left[\frac{S^k}{m}\right]$. Note that the integer divisibility of the quantities is always taken into account $S^k/m, \frac{s^k}{2}/m$ and, without loss of generality, subtraction can also be used instead of division.

Theorem 2: Let the conditions of Theorem1 be satisfied, and let the Vandermonde convolution hold. Then, considering the partition(15), the subset $X^k\subset X^n$ is determined based on the formula: $\sum_{r=0}^k C_{n_1}^r C_{n_2}^{k-r}=C_{n_1+n_2}^k=C_n^k$.

Proof. Indeed, for each set X^{n_1} and X^{n_2}, Theorem1 holds with certificates S^{k_1} and S^{k_2} (where $S^k=S^{k_1}+S^{k_2}, k_1+k_2=k$), respectively. Moreover, $X^{k_1}\subseteq X^{n_1}, X^{k_2}\subseteq X^{n_2}$, $X^{k_1}\cup X^{k_2}=X^k$.

The simplest way to partition the original set X^n into two subsets X^{n_1}, X^{n_2} with cardinalities n_1, n_2 is to divide n by 2. Similarly, to partition the subset X^k into two subsets X^{k_1}, X^{k_2} with cardinalities $k_1\le n_1, k_2\le n_2$, one should divide k by 2, taking into account their evenness and oddness.

The advantage of Theorem2 lies in the ability to generate all subsets N_m^2 and X_m^2 without using the certificate S^k, as well as in its ability to utilize the certificate S^k for the combined sets $\bigcup_m N_m^2$, $\bigcup_m X_m^2$.

Thus, it is important to note that the correspondence between the certificate S^k and the index certificate s^k must be represented as a functional relationship $s^k=f(S^k)$ for $k\ge 3$, while $k<3$ this unambiguous relationship is established through matrices(11) and (12).

Based on the theorems of means (the properties of the arithmetic mean and the combination function(2)), we introduce the following formulas with standard rounding rules to the nearest integer,

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

considering that the certificate S^k and the index certificate s^k belong to the discrete ranges(4) and (6), respectively:

$$k = \frac{(s^k_{min}+s^k_{max})/2}{\sum_{i=1}^{n} i} n \vee k = \frac{(z^k_{min}+z^k_{max})/2}{\sum_{i=1}^{n} x_i} n, \quad (20)$$

$$s^k = \frac{s^k \sum_{i=1}^{n} i}{\sum_{i=1}^{n} x_i} \vee s^k = \frac{(\sum_{i=1}^{n} i)*(z^k_{min}+z^k_{max})/2}{\sum_{i=1}^{n} x_i}. \quad (21)$$

In formulas (20) and (21), the arithmetic means can be replaced by the certificate S^k and the index certificate s^k. However, the precise determination of the cardinality k and the index certificate s^k is obtained through solving the Diophantine equation(9).

The proven lemmas and theorems demonstrate the practical applicability of the new method (new algorithm) for solving the subset sum problem, in comparison to existing exponential [4,5], pseudo polynomial [10,11,12] algorithms, and brute-force methods [6].

Practical implementation of the novel method for solving the Subset Sum Problem

*Example*1. Given the sets $X^8 = \{10,14,17,20,36,38,43,47\}$, $N^8 = \{1,2,\dots,8\}$ (one−dimensional arrays), and the certificate $S^2 = 57$. The index certificate s^2 for $S^2 = 57$ can be computed as follows: since $S^2 \in [24, 90]$ and $k = 2$, we have $s^2 = 9$. We verify the results: according to formulas(20), (21) and range(6) $s^2 \in [3, 15]$, we find $k = \frac{(z^k_{min}+z^k_{max})/2}{\sum_{i=1}^{n} x_i} n = \frac{(24+90)2}{225} 8 = 2{,}02$ or $= \frac{(s^k_{min}+s^k_{max})/2}{\sum_{i=1}^{n} i} n = \frac{(3+15)/2}{36} 8 = 2, s^2 = \frac{3+15}{2} = 9$. Thus, for this index certificate s^2 the solution to the Diophantine equation(17) defines the subsets $N^2 = \{1,8\} \vee \{2,7\} \vee \{3,6\} \vee \{4,5\}$. By sequentially examining the identified subsets N^2, we find subsets $N^2 = \{1,8\}, \{2,7\} \leftrightarrow x_1 + x_8 = 57 \vee x_2 + x_7 = 57$, $X^2 = \{x_1, x_8\} \vee \{x_2, x_7\}$, $T \le O\left(\frac{n}{2}\right) = O(4) \le O(m_2) = O(15) \le O(nk) = O(16) < O(64)$, $\mathbb{S} \le O\left(\frac{(n-1)*n}{2}\right) = 28$.

*Example*2. Given same set X^8 with the certificate $S^k = 100$. The index certificate s^3 for $S^3 = 100$ can be computed as follows: since $S^3 \in [41, 128]$ and $k = 3$, we have $s^3 = s^2 + l$. We verify the results: according to formulas(20), (21) and range (6) $s^3 \in [6, 21]$, we find $k = \frac{(z^k_{min}+z^k_{max})/2}{\sum_{i=1}^{n} x_i} n = \frac{\frac{(41+128)}{2}}{225} 8 = 3{,}004$ or $= \frac{(s^k_{min}+s^k_{max})/2}{\sum_{i=1}^{n} i} n = \frac{\frac{6+21}{2}}{36} 8 = 3$. The solutions to equation(17) are the subsets N^2, where $s^2 = 8$. According to relation(16), by sequentially examining the identified subsets N^2, we find subsets $N^2 = \{1,7\} \vee \{3,5\} \rightarrow X^2 = \{x_1, x_7\} \vee \{x_3, x_5\} \rightarrow S^2 = 53 \rightarrow S^3 = S^2 + x_l = 53 + 47 = 100$. Answer: $N^3 = N_1^2 + l \leftrightarrow X^3 = \{x_1, x_7, x_8\} \vee \{x_3, x_5, x_8\}$,
$T \le O(nk) = O(24) < O(64)$, $\mathbb{S} \le O\left(\frac{(n-1)*n}{2}\right) = 28$.

The second method follows from relation(16): $S^4 = 114 \rightarrow S^2 = 57, s^2 = 9$. We use the calculations from Example1: For the index certificate s^2, the solution to the Diophantine equation(17)

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

defines the subsets $N^2 = \{1,8\} \vee \{2,7\} \vee \{3,6\} \vee \{4,5\}$. By sequentially examining the identified subsets N^2, we find subsets $N^2 = \{1,8\}, \{2,7\} \leftrightarrow x_1 + x_8 = 57 \vee x_2 + x_7 = 57$, $X^2 = \{x_1, x_8\} \vee \{x_2, x_7\}$. According to relation(16), we obtain: $k = 2m \leq n$: $X^3 = \bigcup_m X_m^2 \setminus x_l$, $N^3 = \bigcup_m N_m^2 \setminus n_l$, $S^3 = S^4 - x_l$, $S^3 = 114 - 14 = 100, s^3 = s^4 - n_l$, $s^3 = 18 - 2 = 16, x_2 = 14, x_2 \in X^n, x_2 \notin \bigcup_m X_m^2, n_2 = 2, n_2 \in N^n, n_2 \notin \bigcup_m N_m^2, X^4 = \{x_1, x_8\} \cup \{x_2, x_7\}, N^2 = \{1,8\} \cup \{2,7\}$.

Answer: $X^3 = \bigcup_m X_m^2 \setminus x_2 = \{x_1, x_8, x_7\}$, $N^3 = \bigcup_m N_m^2 \setminus n_2 = \{n_1, n_8, n\}$, $T \leq O(nk) = O(24) < O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

*Example*3. Given same set X^8 with the certificate $S^k = 113$. First, determine the cardinality *k* of the subset X^k based on the certificate $S^k = 113$ belonging to range(4), where $S^k \in [61,164]$ and *k*=4. Alternatively, according to formula(20), we have: $k = \frac{(61+164)8}{2*225} = 4{,}017$, after rounding, $k = 4$. To apply Theorem1 and solve the Diophantine equation(17), S^k needs to be divided by 2, yielding $S^2 = 57$ or 56. Consequently, we have $s^2 = 1 + 8 = 9$. Solving the Diophantine equation(17) for s^2 gives: $N^2 = \{1,8\} \vee \{2,7\} \vee \{3,6\} \vee \{4,5\}$. Verification of the condition: $\sum_{(x_i,x_j)\in X^2}(x_i + x_j) = S^2 = 57$ determines: $X_2^2 = \{10,47\} \vee \{14,43\}$. Perform similar steps for: $S^2 = 56$, resulting in: $N_2^2 = \{4,5\}$ and $\sum_{(x_i,x_j)\in X^2}(x_i + x_j) = S^2 = 56$. According to Theorem1, the final result is: $N^4 = \bigcup_2 N_2^2 = \{1,8\} \cup \{4,5\} \vee N^4 = \bigcup_2 N_2^2 = \{2,7\} \cup \{4,5\}$. These subsets describe: $X^4 = \bigcup_2 X_2^2 = \{x_1, x_8\} \cup \{x_4, x_5\} \vee \{x_2, x_7\} \cup \{x_4, x_5\}$, $X^4 = \{x_1, x_8, x_4, x_5\} \vee \{x_2, x_7, x_4, x_5\}$, $T \leq O(nk) = O(32) < O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

*Example*4. Given the set $X^8 = \{10,14,17,20,36,38,43,47\}$ and the certificate $S^4 = 120$ since there is no solution to the Diophantine equation(17) for s^4, an alternative certificate satisfying the conditions of Theorem1 must be found. Specifically, $\sum_{i=1}^n x_i$- $S^4 = 225 - 120 = 105$. Thus, $S^2 = \frac{105}{2}, S^2 = 53 \vee 52$. Given that $s^2 = 8$, $N^4 = \bigcup_2 N_2^2 = \{1,7\} \cup \{2,6\} \vee N^4 = \bigcup_2 N_2^2 = \{3,5\} \cup \{2,6\}$. The result is: $N^4 = N^n \setminus \bigcup_2 N_2^2 = N^n \setminus (\{3,5\} \cup \{2,6\}) = \{1,4,7,8\}$, $X^4 = \{x_1, x_4, x_7, x_8\} \vee N^4 = N^n \setminus \bigcup_2 N_2^2 = N^n \setminus (\{1,7\} \cup \{2,6\}) = \{3,4,5,8\}$: $X^4 = \{x_1, x_4, x_7, x_8\} \vee X^4 = \{x_3, x_4, x_5, x_8\}, T \leq O(nk) = O(32) < O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

*Example*5. Given the set $X^8 = \{10,14,17,20,36,38,43,47\}$ and the certificate $S^5 = 120$. According to formula(20), we have: $k = \frac{(97+184)8}{2*225} = 4{,}99 \rightarrow k = 5, k = \frac{(15+30)/2}{36}8 = 5$. Solving the Diophantine equation(17) for $s^2 = 8$, gives: $N^2 = \{1,7\} \vee \{2,6\} \vee \{3,5\}$ and $N^6 = \{1,7\} \cup \{2,6\} \cup \{3,5\}$, $S^6 = 158$, $S^5 = S^6 - x_6 = 158 - 38 = 120 \rightarrow N^5 = \{1,7\} \cup \{3,5\} \cup \{2\} \rightarrow X^5 = \{x_1, x_7, x_3, x_5, x_2\}, T \leq O(nk) = O(40) < O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

*Example*6. Given the set $X^8 = \{10,14,17,20,36,38,43,47\}$ and the certificate $S^6 = 168$. Here: 168/3=56. According to Example3, we have: $S^2 = 56$ and $s^2 = 9, N^2 = \{1,8\} \vee \{2,7\} \vee \{3,6\} \vee \{4,5\}$, $N^6 = \{1,8\} \cup \{3,6\} \cup \{4,5\} \vee \{2,7\} \cup \{3,6\} \cup \{4,5\}, X^6 = \{x_1, x_8, x_3, x_6, x_4, x_5\} \vee \{x_2, x_7, x_3, x_6, x_4, x_5\}$, $T \leq O(nk) = O(48) < O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

*Example*7. Consider the set $X^{27} = X^{13} \cup X^{14}$ and it is necessary to find a subset $X^5 \subset X^{27}$. according to formula (2). In this case, the Vandermonde convolution is given by: $C_{13}^4 C_{14}^1 + C_{14}^4 C_{13}^1 + C_{13}^2 C_{14}^3 + C_{14}^3 C_{13}^2 + C_{13}^5 C_{14}^0 + C_{14}^5 C_{13}^0 = C_{27}^5$. For each combination function, Theorem1 can be applied based on the examples provided above. This significantly reduces the number of combinations, as the Diophantine equation (17) is used. The Vandermonde formula facilitates the parallelization of the computation process with the potential use of Theorem1.

*Example*8. The validity of the theoretical and practical results obtained using Theorem1 and the Diophantine equation(17) for the sets $X^8 = \{10,14,17,20,36,38,43,47\}$, $N^8 = \{1,2,\dots,8\}$ is demonstrated by matrix(11), which has elements $x_{ij} = x_i + x_j = S^2$; matrix (12), which contains elements $n_{ij} = (n_i, n_j)$; and matrix(13), which represents the sum of two indices such that $n_i + n_j = s^2$:

$$X^2 = \begin{Bmatrix} 24 & 27 & 30 & 46 & 48 & 53 & 57 \\ & 31 & 34 & 50 & 52 & 57 & 61 \\ & & 37 & 53 & 55 & 60 & 64 \\ & & & 56 & 58 & 63 & 67 \\ & & & & 74 & 79 & 83 \\ & & & & & 81 & 85 \\ & & & & & & 90 \end{Bmatrix}, N^2 = \begin{Bmatrix} 1,2 & 1,3 & 1,4 & 1,5 & 1,6 & 1,7 & 1,8 \\ & 2,3 & 2,4 & 2,5 & 2,6 & 2,7 & 2,8 \\ & & 3,4 & 3,5 & 3,6 & 3,7 & 3,8 \\ & & & 4,5 & 4,6 & 4,7 & 4,8 \\ & & & & 5,6 & 5,7 & 5,8 \\ & & & & & 6,7 & 6,8 \\ & & & & & & 7,8 \end{Bmatrix}, s^2 = \begin{Bmatrix} 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ & 5 & 6 & 7 & 8 & 9 & 10 \\ & & 7 & 8 & 9 & 10 & 11 \\ & & & 9 & 10 & 11 & 12 \\ & & & & 11 & 12 & 13 \\ & & & & & 13 & 14 \\ & & & & & & 15 \end{Bmatrix}. \quad (22)$$

Indeed, the use of matrices (22) facilitates the combined application of Theorem1 and Lemma2 for certificates $S^8 = 225$ and $S^7 = 189$, among others. The sum of the indices of the set N^8 is given by: $s^8 = \frac{n(n+1)}{2} = 36$. The solutions to the Diophantine equation(17) with an index certificate 36/4=9 are the subsets $N_m^2 = \{1,8\} \vee \{2,7\} \vee \{3,6\} \vee \{4,5\}$. Their combination is given by: $N^8 = \bigcup_{m=1}^4 N_m^2$, accordingly, the solution to the problem(1) is: $X^8 = \bigcup_{m=1}^4 X_m^2$. From the first matrix of matrices(22), by sequential examination, we have: $S^8 = 57 + 57 + 55 + 56 = 225$. For S^7 using the relationships(16), we get: $X^7 = \bigcup_{m=1}^4 X_m^2 \setminus x_l = 225 - 36 = 189, x_5 = 36$. The complexity is: $T \leq O(nk) = O(64)$, $\mathbb{S} \leq O\left(\frac{(n-1)*n}{2}\right) = 28$.

In some examples, the index *m* is omitted in the subsets N_m^2 and X_m^2 for simplifying the notation of the formulas.

It is important to emphasize that Theorem1 has been examined for applications involving the selection of subsets X^k with arbitrary cardinalities $k = 2 \vee 3 \vee 4 \vee 5 \vee 6 \vee 7 \vee 8$ from an initial set with any size n. This underscores the general applicability of Theorem1 and Lemma2.

Conclusion

The subset sum problem, classified as NP-complete, is considered. Diophantine equations and an auxiliary problem are introduced to facilitate solving the original problem, which also holds independent scientific interest. Lemmas and theorems are proven, enabling the development of efficient and straightforward algorithms for solving the subset sum problem. The time required to select the necessary subsets and the space needed do not exceed the square of the length of the input data. An analytical framework for working with indices of the initial set is developed. The proposed

Vol. 15 No. 3 (2024):301-311
DOI: https://doi.org/10.61841/turcomat.v15i3.14804

algorithms are applicable to solving independent set problems of size *k* and *k*-vertex cover problems. Examples are provided to illustrate the high efficiency of the new method for solving the subset sum problem.

It should be noted that the time required to sort an array of integers is proportional to the square of the dimension of the one-dimensional array describing the initial set of distinct positive integers, and this task belongs to the class P. Therefore, based on the newly developed method, it can be assumed that the subset sum problem, which is NP-complete and belongs to the NP class, also belongs to the P class.

References

[1] S.A. Cook. The complexity of theorem-proving procedures // STOC, 1971, pp.151–158.
[2] R.M. Karp. Reducibility among combinatorial problems // Complexity of Computer Computations, 1972, IBM Research Symposia Series, pp.85–103.
[3] B. Sinchev, A.B. Sinchev, A.M. Mukhanova. Algorithm based on the subset sum problem for high performance computing // Springer Link Proceedings of Ninth International Congress on Information and Communication Technology, 20 Jul 2024, pp.627-637
[4] E. Horowitz, S. Sanni. Computing Partitions with Application to the Knapsack Problem // Journal of the ACM(JACM), 1974, T21, pp.277-292
[5] R. Schroeppel, A. Shamir. A $T=O(2^{n/2})$, $S=O(2^{n/4})$ Algorithm for Certain NP-Complete Problem // SIAM Journal on Computing, 1981, Vol.10, № 3, pp.456-464
[6] Н. Вирт. Алгоритмы и структуры данных. Пер. с англ. – М.: Мир, 2006.
[7] B. Sinchev, A. Sinchev, J. Akzhanova, Y. Issekeshev, A. Mukhanova. Polynomial time algorithms for solving NP-complete problems // News of the National Academy of Sciences of Kazakhstan, Series of Geology and Technical Sciences, Volume 3, Number 441, 2020, pp.97-101
[8] Б. Синчев. О полиномиальной разрешимости класса NP-complete // International journal of information and communication technologies, Том 2, Вып. 8, 2021, pp.67-71
[9] B. Sinchev, A. B. Sinchev, Z.A. Akzhanova, Y. Issekeshev. Computing network architecture for reducing a computing operation time and memory usage associated with determining, from a set of data elements, a subset of at least two data elements, associated with a target computing operation result. // Patent USPTO 10,860,317 B2, 2020, 34p.
[10] Konstantinos Koiliaris, Chao Xu. A Faster pseudopolynomial time algorithm for subset sum // arXiv:1610.04712v2[cs.Ds], 8 Jan 2017, 18p.
[11] K. Bringmann. A near-linear pseudopolynomial time algorithm for subset sum // arXiv:1610.04712v2[cs.Ds], 8 Jan 2017, 18p.
[12] A. Lincoln, V. Williams, JR Wang, R. Williams. Deterministic Time-Space Tradeoffs for k-SUM // arXiv preprint arXiv:1605.07285