ON A SOLUTION OF "P VS NP" MILLENIUM PRIZE PROBLEM BASED ON THE SUBSET SUM PROBLEM

B. Sinchev* **A. B. Sinchev†** **A. M. Mukhanova‡**

Abstract

Given a set of distinct non-negative integers X^n and a target certificate S parametrized in: $\exists X^k \subseteq X^n$, $\sum_{x_i \in X^k} x_i = S$ $(k = |X^k|, n = |X^n|)$. We present a polynomial solution of the subset sum problem with time complexity $T \leq O(n^2)$ and space complexity $\mathbb{S} \leq O(n^2)$, so that P = NP.

Keywords. P, NP and NP-complete class, set, subset, potency, time, space

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

1 Introduction

We consider the subset sum problem: given a set of distinct non-negative integers X^n, and a value sum (certificate) S, determine if there is a subset X^k of the given set with a sum equal to the given sum S.

For this problem, exponential [1,2], pseudopolymial [3,4,5,6,7] algorithms and exhaustive search methods based on the "divide and conquer" principle [8] have been developed. The complexity of algorithms was considered in [9,10,11].

It is proved that the subset sum problem belongs to the NP-complete class problems. According to the well-known theorem that if there is a polynomial solution to the NP-complete class problem then $P = NP$.

To the best of our knowledge, the proposed polynomial solution with time complexity $T \leq O(n^2)$ and space complexity $\mathbb{S} \leq O(n^2)$ is the fastest general algorithm for this problem and is reduced to the problem of X^n elements' index selection based on the solution of auxiliary problem of a one-to-one correspondence of the certificate S (of the initial problem) with the newly proposed index certificate s_k.

The paper is organized as follows:

- given lemma 1 specifies a newly proposed s_k;
- given lemma 2 identifies a one-to-one correspondence of the index certificate s_k with the certificate S resulted with selection of subsets N^k among which there is subset X^k- a solution to initial problem;

*International University of Information Technology, Almaty, Kazakhstan: sinchev@mail.ru
†National Information Technologies JSC, Astana, Kazakhstan : askar.sinchev@gmail.com
‡Almaty University of Technology, Almaty, Kazakhstan: a.mukhanova@atu.kz

- given theorem estimates selection time for finding subsets X^k based on triangle two-dimensional matrixes (index combinations) and concatenation operator;
- additionally, in this paper we present: arguments on potency of subset X^k, new combination algorithms and example to confirm claimed results.

The proposed algorithms can be computed in the form of a separate independent module (as a software solution for various problems) and implemented in the form of chips (as a hardware solution).

2 Solution of the subset sum problem

Given a set of distinct non-negative integers X^n and a target certificate S parametrized in:

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

Subset X^k is selected using combination function:

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

We propose a new approach to solve initial problem(1) by solving auxiliary problem:
Let us introduce a sorted set of natural numbers $N^n = \{1, 2, 3, \ldots, n\}$ or $N^n = \{0, 1, 2, 3, \ldots, n-1\}$ with potency $n = |N^n|$. Then the auxiliary problem statement is:

$$s_k : \exists N^k \subseteq N^n, \sum\nolimits_{n_i \in N^k} n_i = s_k \tag{3}$$

where s_k is index certificate.

Auxiliary problem(3) excludes the accuracy parameter p (number of bits of integers X^n) from the computational complexity of problem(1), thus facilitates the solution of problem(1). The subsets N^k are determined based on the combination function(2). Each subset N^k consists of k elements of the set N^n.

According to the combination function(2), we have the range:

$$s_k \epsilon [s_k^{min}, s_k^{max}], \tag{4}$$

and the number of unique index certificates:

$$m_k = s_k^{max} - s_k^{min} + 1 \tag{5}$$

where $s_k^{min} = \sum_{i=1}^{k} i = \frac{(k+1)k}{2}$, $s_k^{max} = \sum_{i=n-k+1}^{n} i = kn - \frac{(k-1)k}{2}$.The lower index characterizes the potency k. Range(4) defines all values of the index certificate from s_k^{min} to s_k^{max}, as well as their quantity.

Lemma1 Let the condition $\sum_{x_i \in X^k} x_i = S$ of the initial problem(1) is met. Then there exist one or more subsets $N^k \subseteq \ N^n$ with potency k and index certificate s_k so that, there is a solution to auxiliary problem(3).

Proof. Meeting the first condition of Lemma 1 means that we have the sum of k elements of subset X^k:

$$x_i + x_j + ... + x_g + x_h = S, i \neq j ... \neq g \neq h; x_i, x_j ..., x_g, x_h \in X^k \subseteq X^n. \tag{6}$$

Indexes $i \neq j \neq ... \neq g \neq h$ of the subsets $X^k \subseteq X^n$ are selected using combination function(2). From the equality(6) we find index certificate s_k of the auxiliary problem(3):

$$n_i + n_j + ... + n_g + n_h = s_k; i = n_j, j = n_j, ..., g = n_g, h = n_h, N^k \subseteq N^n. \quad (7)$$

With the combination function(2) we find indexes of the subset N^k matching indexes of the subset X^k. Thus, the solvability of the initial problem(1) results with solvability of the auxiliary problem(3), namely meeting the condition: $\sum_{n_i \in N^k} n_i = s_k$.

Lemma2 There is index certificate s_k corresponding to certificate S so that the equality $\sum_{n_i \in N^k} n_i = s_k$ is met. Then within the subsets N^k there exist one or more subsets N^k which describe indexes of at least one subset X^k to meet the equality: $\sum_{X_i \in X^k} x_i = S$.

Proof. According to the equalities (6) and (7) of the Lemma1 there is a one-to-one correspondence of the certificate S and the index certificate s_k with potency k of the subsets X^k and N^k. The correspondence we present as:

$$S \sim s_k \quad (8)$$

The second condition of Lemma2 means that $\exists N^k$ which describe indexes of subsets X^k, where at least one subset X^k exists so that the condition $\sum_{x_i \in X^k} x_i = S$ is met.

Notice that the correspondence(8) is difficult to represent in the form of a functional dependence $s_k = f(S)$. The advantages of the lemmas(1,2) and the auxiliary problem(3) are:

1. N^k describes all subsets X^k;
2. the selection time of the subset X^k decreases;
3. the size of the required space $\mathbb{S}$ is determined.

New combination algorithms Let's introduce a counter function:

$$q = q(s_k, N^k) \quad (9)$$

which counts the quantity of subsets N^k.

Theorem Let a set N^n and index certificate s_k are given. Then the selection time of the subset $N^k \subseteq N^n$. and the required space satisfy the conditions: $T \leq O(q(s_k, N^k)) \leq O(kn)$, $\mathbb{S} \leq O(\frac{(n-1)*n}{2})$.

Proof. Initially, we examine all subsets N^k (of potency k) whose indexes vary from (1,2,..,k) to (n-k+1, n-k+2,. . . ,n-1, n) and are representable in the form of two-dimensional triangular matrixes of orders (n-k+1)x(n-k+1) up to (1x1) with the use of concatenation operator $\oplus$, since the set N^n is a one-dimensional array. Then index certificate s_k will be equal to the sum of indexes of the element of one of the diagonals of these matrixes and the quantity of elements is determined by the counter function(9). In other words, the quantity of subsets N^k is equal to the number of elements of the diagonal corresponding to index certificate s_k and the selection time of these subsets satisfies the condition $T \leq O(q(s_k, N^k))$. In addition to condition(7), condition(6) must be satisfied, so that $q(S, X^k) \leq q(s_k, N^k)$. Next we find value $m_k = s_k^{max} - s_k^{min} + 1 = kn - \frac{(k-1)k}{2} - \frac{(k+1)k}{2} + 1$, form these formulas we have the estimation $m_k < kn$. Then, according to the proposed approach, we obtain $T \leq O(q(s_k, N^k)) \leq O(m_k) \leq O(kn), \mathbb{S} \leq O(\frac{(n-1)*n}{2})$. Here the expression $\frac{(n-1)*n}{2}$ defines the maximum size of the required space.

Corollary. Let $k = 2$ and set N^n is given. Then $\exists\ s_2 = s_2^*$ so that:
$T \le O(q(s_2^*, N^2)) \le O(\frac{n}{2}) \le O(m_k) \le O(kn)$, $\mathbb{S} = O(\frac{(n-1)n}{2})$.

Proof. We introduce an algorithm for generating two-dimensional triangular matrices that describe all subsets N^k with k=2. It is enough to use concatenation operator $\oplus$ and add a number 1 to the elements of the set N^n starting from the second element and to the end; then add a number 2 to the elements of this set, starting from the third element to the end, and so on - until we get the last element $(n-1n)$, that is:

$$\begin{array}{cccccccc} 12 & 13 & \dots & 1 & n-1 & 1 & n \\ & 23 & 24\dots & 2 & n-1 & 2 & n \\ & & & & \dots & & \\ & & n-2 & n-1 & & n-2 & n \\ & & & & & n-1 & n \end{array}$$

Quantity of subsets N^2 matches the value of the combination function C_n^2. This matrix describes all the necessary indexes by which all subsets of X^2 can be found, and also defines index certificates $s_2 = n_i + n_j, n_i, n_j \in N^n$. With k=2, the maximization problem can be formulated as: $max_{s_k \in [s_k^{min}, s_k^{max}]} q(s_k, N^k) = q(s_k^*, N^k)$ then the selection time of the subset X^k satisfies the inequality $T \le O(\frac{n}{2}) \le O(kn)$ and the required space is $\mathbb{S} = O(\frac{(n-1)n}{2})$.

It is easy to develop algorithms for generating combinations for k3. In the case k=3, the maximization problem can be formulated as: $max_{(s_k \in [s_k^{min}, s_k^{max}])} q(s_k, N^k) = q(s_k*, N^k)$,then we have $T \le O(n) \le O(kn)$, $\mathbb{S} \le O(\frac{(n-1)n}{2})$ using a recursive way to store two-dimensional matrixes.

Similar results can obtained with $k \ge 4$ with possible additional consideration of the symmetry property - $C_n^k = C_n^{n-k}$, Pascal's Rule - substitution of indexes - $C_n^k C_k^{n-m} = C_n^m C_m^{n-k}$, Vandermonde's convolution - $\sum_{r=0}^{k} C_n^r C_m^{k-r} = C_{n+m}^k$ and recurrent formula $C_n^k = \frac{n-k+1}{k} C_n^{k-1}$.

Using the correspondence (8) and the range (4) we have the range for the certificate S:

$$S \in [S_k^{min}, S_k^{max}] \tag{10}$$

where based on combination function(2) we determine values $S_k^{min} = \sum_{i=1}^{k} x_i, S_k^{max} = \sum_{i=n-k+1}^{n} x_i, x_i \in X^n$, and it is assumed that the set X^n is sorted in ascending order.

Minimum potency k_* is selected from the inequality:

$$S \le S_{k_*}^{max} \tag{11}$$

Argument1 Let the certificate S be given. Potency k_* of the subsets $X^{k_*} \subseteq X^n, N^{k_*} \subseteq N^n$ are determined from the inequality(11) and index certificate s_{k_*} is determined from the correspondence(8). Then, among the quantity of subsets N^{k_*} found from $\sum_{n_i \in N^k} n_i = s_{k*}$ there is at least one subset N^k which describes subset X^k so that $\sum_{x_i \in X^k} x_i = S$.

Proof. The fulfillment of these equalities ensures the solvability of problems (3) and (1). In this case, the potency k_* of subsets X^{k_*}, N^{k_*} is found from the solution of the minimization problem $S_{k_*}^{max} = S_k^{max} min$ k for $k \ge 2$ taking into account the inequality(11). To ensure the correspondence(8), it is enough to use the relation $[\frac{(n+1)nk_*}{2n}]$. The rounding process finds the index certificate $s_{k_*} = [\frac{(n+1)k_*}{2}]$ or $s_{k_*} = [\frac{(n+1)k_*}{2}] + 1$. Thus, the solution to the problem $\sum_{n_i \in N^k} n_i = s_{k_*}$ defines quantity of subsets N^{k_*} which describe all subsets X^{k_*}, among which there is a subset X^k satisfying the equality $\sum_{x_i \in X^k} x_i = S$. The possible value of the index certificate $s_{k_*} = [\frac{(n+1)k_*}{2}] - 1$ takes into account the combinatorics of

the problems being solved. In the case of $S = S_{k_*}^{max}$, the required subsets X^{k_*} are found and there is no need to solve the auxiliary problem(3).

The maximum potency k^* is selected from the inequality:

$$S \geq S_{k^*}^{min} \tag{12}$$

Argument2 Let the certificate S be given. Potency k^* of the subsets $X^{k^*} \subseteq X^n, N^{k^*} \subseteq N^n$ are determined from the inequality(12) and index certificate s_{k^*} is determined from the correspondence(8). Then, among the quantity of subsets N^{k^*} found from $\sum_{n_i \in N^k} n_i = s_{k^*}$ there is at least one subset N^k which describes subset X^k so that $\sum_{x_i \in X^k} x_i = S$.

Proof. Here we solve the maximization problem $S_{k_*}^{min} = S_k^{min} max$ k with $k \geq 2$ and taking into account inequality(12). Index certificate is found from the condition $s_{k^*} \leq \frac{(k^*+1)k^*}{2}$. In case $S = S_{k^*}^{min}$ the required subsets X^{k^*} are found and there is no need to solve the auxiliary problem(3).

Remark. Considering that there are approaches for choosing elements of the set X^n based on arithmetic progression, Fibonacci numbers, recurrence relations, subexponential functions and others, and assuming the possibility that the set X^n consists of elements of the same order, we can carry out the following values:

$$\sum_{n_i \in N^k} n_i = \mathcal{N} \tag{13}$$

$$\sum_{x_i \in X^n} x_i = \mathcal{S} \tag{14}$$

Argument3 With given certificate S of the subset sum problem(1) index certificate s of the auxiliary problem(3) is found by the formula:

$$S = [\frac{\mathcal{N}S}{\mathcal{S}}] - 1 \vee [\frac{\mathcal{N}S}{\mathcal{S}}] \vee [\frac{\mathcal{N}S}{\mathcal{S}}] + 1. \tag{15}$$

Proof. Formula(13) defines the sum of all indexes of the set N^n, formula(14) – the sum of all elements of the set X^n. Then the relation $\frac{S}{s} \cong \frac{\mathcal{S}}{\mathcal{N}}$ is true according to the mean-value theorems. Taking into account the integer divisibility and the combinatorics of problems (1) and (3), we obtain the required formula (15).

Validation of claimed results. Example. Sets $X^8 = \{10, 14, 17, 20, 36, 38, 43, 47\}$, $N^8 = \{1, 2, \ldots, 8\}$ and certificate $S = 60$ are given. Based on any argument (1, 2, 3) it is not difficult to calculate the index certificate s_2 with $S = 60$ because $S \in [24, 90]$ and $k = 2$, then we have $s_2 = 10$. From Corollary we have: $N^2 = \{2, 8\} \vee \{3, 7\} \vee \{4, 6\}$. Among the quantity of subsets N^2 there is subset $N^2 = \{3, 7\} \Leftrightarrow x_3 + x_7 = 60$, $X^2 = \{17, 43\}$, $T \leq O(q(s_2, N^2)) = O(3) \leq O(m_2) = O(15) < O(2n) = O(16)$, $\mathbb{S} \leq O(\frac{(n-1)*n}{2}) = 28$.
For the indicated set X^8, examples with certificates S=57, S=99, S=100, S=105, S=112,S=113,S=120, S=168 were additionally solved to confirm claimed results.

Conclusions and future work. The article provides lemmas, a theorem, and arguments to solve the subset sum problem based on an auxiliary problem. On the basis of the results obtained, an initial set indexes' supervisor engine has been developed. The proposed algorithms can be easily implemented as software and/or hardware solution in

a variety of applications including: scheduling[12], queries in databases[13], graph problems[14] and others.

In a view of the fact that the linear (or quadratic) solvability of the subset sum problem from the NP-complete class is proved, therefore, based on the well-known theorem (stating that if some NP-complete problem is solvable in polynomial time, then P = NP), the equality of classes P and NP is claimed.

Further work directions will be focused on:

1. partition of an initial set into subsets (using Vandermonde's convolution and symmetry property);
2. applying combination function properties;
3. optimization of combination algorithms;
4. calculating process paralleling etc.

References

[1] E. Horowitz, S. Sanni. Computing Partitions with Application to the Knapsack Problem //Journal of the ACM(JACM), 1974, T21, pp.277-292

[2] R. Schroeppel, A. Shamir A T=O(2n/2), S=O(2n/4) Algorithm for Certain NP-Complete Problem // SIAM Journal on Computing, 1981, Vol.10, № 3, pp.456-464

[3] Richard Bellman. Notes on the theory of dynamic programming iv - maximization over discrete sets. //Naval Research Logistics Quarterly, 3(1-2):67–70, 1956.

[4] David Pisinger. Linear time algorithms for knapsack problems with bounded weights. //Journal of Algorithms, 33(1):1 – 14, 1999

[5] Konstantinos Koiliaris, Chao Xu. A Faster pseudopolynomial time algorithm for subset sum. To appear in SODA '17, 2017. //arXiv:1610.04712v2[cs.Ds] 8 Jan 2017.-18p.

[6] Karl Bringmann. A near-linear pseudopolynomial time algorithm for subset sum. To appear in SODA '17, 2017. //arXiv:1610.04712v2[cs.Ds] 8 Jan 2017.-18p.

[7] A Lincoln, VV Williams, JR Wang, RR Williams. Deterministic Time-Space Tradeoffs for k-SUM //arXiv preprint arXiv:1605.07285

[8] N. Wirth. Algorithms and Data Structures.Russian translate – .: Mir, 2006.

[9] R. M. Karp. Reducibility among combinatorial problems. Springer, 1972.

[10] Cobham A. The intrinsic computational difficulty of functions. //In Proceedinges of the Congress for logic, methodology and philosophy of science.-NorthHoiLand, 1964.P.24-30.

[11] Egmonds J. Parths, treers and flowers. // Canadian Journal of mathematics. -1965. Vol.17. –P.449-467.

[12] Xiangtong Qi. Coordinated logistics scheduling for in-house production and outsourcing. //Automation Science and Engineering, IEEE Transactions on, 5(1):188–192, Jan 2008.

[13] Quoc Trung Tran, Chee-Yong Chan, and Guoping Wang. Evaluation of set-based queries with aggregation constraints. //In Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011, pages 1495–1504, 2011.

[14] Venkatesan Guruswami, Yury Makarychev, Prasad Raghavendra, David Steurer, and Yuan Zhou. Finding almost-perfect graph bisections. //In Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings, pages 321–337, 2011.