Efficient Approximation Algorithms for the SUBSET-SUMS EQUALITY Problem*

Cristina Bazgan[1] Miklos Santha[2] Zsolt Tuza[3]

[1] Université Paris-Sud, LRI, bât.490, F–91405 Orsay, France, bazgan@lri.fr
[2] CNRS, URA 410, Université Paris-Sud, LRI, F–91405 Orsay, France, santha@lri.fr
[3] Computer and Automation Institute, Hungarian Academy of Sciences, H–1111 Budapest, Kende u.13–17, Hungary, tuza@sztaki.hu

Abstract. We investigate the problem of finding two nonempty disjoint subsets of a set of n positive integers, with the objective that the sums of the numbers in the two subsets be as close as possible. In two versions of this problem, the quality of a solution is measured by the ratio and the difference of the two partial sums, respectively.
Answering a problem of Woeginger and Yu (1992) in the affirmative, we give a fully polynomial-time approximation scheme for the case where the value to be optimized is the ratio between the sums of the numbers in the two sets. On the other hand, we show that in the case where the value of a solution is the positive difference between the two partial sums, the problem is not 2^{n^k}-approximable in polynomial time unless P=NP, for any constant k. In the positive direction, we give a polynomial-time algorithm that finds two subsets for which the difference of the two sums does not exceed $K/n^{\Omega(\log n)}$, where K is the greatest number in the instance.

1 Introduction

KNAPSACK is a well known problem which was shown to be NP-complete in 1972 by Karp [3]. It remains NP-complete even if the size of each object is equal to its value. This particular case is called the SUBSET-SUM problem. Ibarra and Kim [2], gave a fully polynomial-time approximation scheme for the optimization problem associated with KNAPSACK which, therefore, applies to SUBSET-SUM as well. The most efficient fully polynomial-time approximation scheme known for the SUBSET-SUM problem is due to Kellerer et al. [4]. The running time of their algorithm is $O(min\{n/\varepsilon, n + (1/\varepsilon)^2 log(1/\varepsilon)\})$, and the space required is $O(n + 1/\varepsilon)$, where n is the number of the integers and ε the accuracy.

The input to an instance of SUBSET-SUM is a set of n positive integers $a_1, \ldots, a_n$ and another positive integer b. The question is to decide if there

* This research was supported by the ESPRIT Working Group RAND2 n^o 21726 and by the bilateral project Balaton, grant numbers 97140 (APAPE, France) and F-36/96 (TéT Alapítvány, Hungary)

exists a set of numbers (a subset of $\{a_1, \ldots, a_n\}$) whose sum is equal to b. In the optimization version the goal is to find a set of numbers whose sum is as large as possible under the constraint that it does not exceed b.

Woeginger and Yu [7] introduced a related problem, called SUBSET-SUMS EQUALITY. Given n positive integers, the question is to decide if there exist two disjoint nonempty subsets whose sums are equal. They also defined a related optimization problem that we call SUBSET-SUMS RATIO; it requires to find two disjoint subsets with the ratio of their sums being as close to 1 as possible. In the same paper they proved the NP-completeness of SUBSET-SUMS EQUALITY, and gave a polynomial-time 1.324-approximation algorithm for SUBSET-SUMS RATIO. They left as an open question to decide whether this problem has a polynomial-time approximation scheme.

In this paper we answer their question in the affirmative, by showing the stronger assertion that actually SUBSET-SUMS RATIO has a *fully polynomial-time* approximation scheme.

The problems defined by Woeginger and Yu have some interesting special instances. Consider the case where the sum of the n numbers is less than $2^n - 1$. It is immediately seen by the pigeonhole principle that there always exist two disjoint nonempty subsets whose sums are equal. Nonetheless, no polynomial-time algorithm is known so far to find two such subsets effectively. We call this latter problem PIGEONHOLE SUBSET-SUMS. This problem is a well known member of what Meggido and Papadimitriou [5,6] call the class $TFNP$ of total functions. This class contains function problems associated with languages in NP where, for every instance of the problem, a solution is guaranteed to exist. Other examples in the class are FACTORING, SECOND HAMILTONIAN CYCLE and HAPPYNET.

Many functions in $TFNP$ (like the examples quoted above) have a challenging intermediate status between FP and FNP, the function classes associated with P and NP. Although these problems are not NP-hard unless NP=co-NP, no polynomial-time algorithm is known for them.

Although the polynomial-time solvability of PIGEONHOLE SUBSET-SUMS still remains open, we will show that in a sense this problem is much better approximable in polynomial time than SUBSET-SUMS EQUALITY. For this purpose, we define a further related optimization problem that we call SUBSET-SUMS DIFFERENCE. Here the value of a solution is the positive difference between the sums of the two sets plus 1. The same problem, with the additional constraint that the sum of the numbers is less than $2^n - 1$, is called PIGEONHOLE SUBSET-SUMS DIFFERENCE.

The existence of a fully polynomial-time approximation scheme for SUBSET-SUMS RATIO implies that, for any constant k, there is a polynomial-time $2^n/n^k$-approximation algorithm for PIGEONHOLE SUBSET-SUMS DIFFERENCE. We will show an even stronger result, giving a polynomial-time $2^n/n^{\Omega(\log n)}$-approximation for this problem. This will follow from a more general theorem: we will show that SUBSET-SUMS DIFFERENCE has a polynomial-time $K/n^{\Omega(\log n)}$-approximation algorithm where K is the largest number in the input. On the