Equal Sum Subsets: Complexity of Variations

Mark Cieliebak*, Stephan Eidenbenz*, Aris Pagourtzis*, and Konrad Schlude*

Abstract

We start an investigation into the complexity of variations of the EQUAL SUM SUBSETS problem, a basic problem in which we are given a set of numbers and are asked to find two disjoint subsets of the numbers that add up to the same sum. While EQUAL SUM SUBSETS is known to be NP-complete, only very few studies have investigated the complexity of its variations. In this paper, we show NP-completeness for two very natural variations, namely FACTOR-r SUM SUBSETS, where we need to find two subsets such that the ratio of their sums is exactly r, and k EQUAL SUM SUBSETS, where we need to find k subsets of equal sum. In an effort to gain an intuitive understanding of what makes a variation of EQUAL SUM SUBSETS NP-hard, we study several variations of EQUAL SUM SUBSETS in which we introduce additional requirements that a solution must fulfill (e.g., the cardinalities of the two sets must differ by exactly one), and prove NP-hardness for these variations. Finally, we investigate and show NP-hardness for the EQUAL SUM SUBSETS FROM TWO SETS problem and its variations, where we are given two sets and we need to find two subsets of equal sum. Our results leave us with a family of NP-complete problems that gives insight on the sphere of NP-completeness around EQUAL SUM SUBSETS.

1 Introduction

The problem PARTITION, which asks whether there exists a subset A' of a given set A of numbers such that the elements of A' add up to exactly one half of the total sum of all numbers of A, is one of the basic combinatorial problems and has long been known to be NP-complete [4]. We are interested in a variation of PARTITION which we call EQUAL SUM SUBSETS. EQUAL SUM SUBSETS simply asks for two disjoint subsets of a given set of numbers that add up to the same total. In order to give a formal definition of EQUAL SUM SUBSETS, we denote the sum of the elements of a set X of integers by $\mathrm{sum}(X)$, i.e. $\mathrm{sum}(X) := \sum_{x \in X} x$.

Definition (EQUAL SUM SUBSETS). *Given a set[1] of n numbers $A = \{a_1, \ldots, a_n\}$, are there two disjoint nonempty subsets $X, Y \subseteq A$ such that $sum(X) = sum(Y)$?*

EQUAL SUM SUBSETS is a very natural problem that is known to be NP-complete [9]. There also exists an FPTAS for an optimization version of EQUAL SUM SUBSETS, in which the ratio of the sums of the two disjoint subsets is to be minimized [1]. Moreover, the problem has been studied in a restricted version, in which the sum of the n elements is at most $2^n - 1$, in the context of function problems [6].

*Institute of Theoretical Computer Science, ETH Zürich, {cielieba, eidenben, pagour, schlude}@inf.ethz.ch

[1] We do not allow multi-sets here as the problem is trivially solvable if the same number exists more than once in the input.

While Partition, Equal Sum Subsets and variations have numerous applications in production planning and scheduling (see [5] for a survey), our interest for Equal Sum Subsets comes from computational biology. We briefly illustrate this connection (more details can be found in [2]). In the PartialDigest problem we are given a multiset D of distances and are asked to find coordinates of points on a line such that D is exactly the multiset of all pair-wise distances of these points. PartialDigest is a basic problem from DNA sequencing. Neither a polynomial-time algorithm nor a proof of NP-completeness is known for this problem. We have defined an optimization variation of PartialDigest and proved its NP-hardness using a reduction from Equal Sum Subsets.

In this paper, we study the computational complexity of a number of variations of Equal Sum Subsets. After fixing some notation for large numbers that we will use in some of our proofs (Section 2), we study a first set of Equal Sum Subsets variations that we call Factor-r Sum Subsets (for any rational $r > 0$): Given a set of numbers $A = \{a_1, \ldots, a_n\}$, are there disjoint subsets $X, Y \subseteq A$ such that $\text{sum}(X) = r \cdot \text{sum}(Y)$? Factor-$r$ Sum Subsets is a very natural variation of Equal Sum Subsets. In Section 3, we show that Factor-r Sum Subsets is NP-complete for any factor $r > 0$ by giving two reductions from One-in-Three 3-Satisfiability, one that works for all $r > 0$ except $r = 1$ and $r = 2$, and one that works for the case $r = 2$ and uses an argument about the connectivity of Boolean formulas. The case for $r = 1$ is equivalent to Equal Sum Subsets.

In Section 4, we study a second generalization of Equal Sum Subsets, namely k Equal Sum Subsets, in which we need to find k (disjoint) subsets of equal sum from a given set of numbers. k Equal Sum Subsets is a variation of Equal Sum Subsets with an importance of its own. We show that k Equal Sum Subsets is NP-complete for any integer $k \geq 3$ by proposing a reduction from Alternating Partition, which is an NP-complete variation of Partition [3]. The NP-completeness for the case $k = 2$ follows directly from the NP-completeness of Equal Sum Subsets.

In our effort to gain an intuitive understanding of what makes a variation NP-hard, we study variations of Equal Sum Subsets where we add additional requirements that the solution must fulfill. In Section 5, we show NP-completeness for the following three variations by proposing reductions from Alternating Partition: Equal Sum Subsets with Enforced Element (where a specific element, say a_n, must belong to one of the two subsets), Equal Sum Subsets of Different Cardinality (where the two subsets must be of different cardinality), and Equal Sum Subsets of Different by One Cardinality (where the cardinalities of the two subsets must differ by exactly one). We also show that Alternating Partial Partition is NP-complete by reduction from Equal Sum Subsets. As a last result of this section, we show that Equal Sum Subsets of Equal Cardinality (i.e., the problem in which the two equal sum subsets must also be of equal cardinality) is NP-complete by reduction from Alternating Partial Partition.

In order to determine the realm of NP-completeness around Alternating Partial Partition, we study closely related problems, namely the Equal Sum Subsets from Two Sets problem and some of its variations; in this problem, we are given two sets of positive numbers A and B and the question is if there are subsets $X \subseteq A$ and $Y \subseteq B$ of equal sum. In Section 6, we show that Equal Sum Subsets from Two Sets is NP-complete, even if we require the two sets to be of equal cardinality, or to have disjoint indices sets, or disjoint covering indices sets, or identical indices sets.

We conclude in Section 7 with a brief discussion and some ideas for further research.

2 Number Representation

In many of our proofs, we use numbers which are expressed in the number system of some base B. We denote by $\langle a_1, \ldots, a_n \rangle_B$ the number $\sum_{1 \leq i \leq n} a_i B^{n-i}$; we say that a_i is the i-th digit of this number. Usually, we choose base B large enough such that adding up numbers will not lead to carry-digits from one digit to the next. Therefore, we can add numbers digit by digit. The same holds for scalar products. For example, having base $B = 27$ and numbers $\alpha = \langle 3, 5, 1 \rangle, \beta = \langle 2, 1, 0 \rangle$, then $\alpha + \beta = \langle 5, 6, 1 \rangle$ and $3 \cdot \alpha = \langle 9, 15, 3 \rangle$.

We will generally make liberal use of the notation such as allowing different bases for each digit or dropping the base B from our notation if this is clear from the context. We define the concatenation of two numbers by $\langle a_1, \ldots, a_n \rangle \lhd \langle b_1, \ldots, b_m \rangle := \langle a_1, \ldots, a_n, b_1, \ldots, b_m \rangle$, i.e. $\alpha \lhd \beta = \alpha B^m + \beta$, where m is the number of digits in β. We will use $\Delta_n(i) := \langle 0, \ldots, 0, 1, 0, \ldots, 0 \rangle$ for the number that has n digits, all 0's except for the i-th position where the digit is 1. Furthermore, $\mathbb{1}_n := \langle 1, \ldots, 1 \rangle$ is the number that has n digits, all 1's, and $\mathbb{0}_n := \langle 0, \ldots 0 \rangle$ has n zeros. Notice that $\mathbb{1}_n = B^n - 1$.

3 NP-completeness of Factor-r Sum Subsets

In this section, we study a natural generalization of Equal Sum Subsets that is important *per se*, namely the Factor-r Sum Subsets problem, where we want to find two subsets whose sums have a specific ratio r. This is closely related to the minimization version of Equal Sum Subsets studied in [1].

Definition (Factor-r Sum Subsets). *Given a set of n numbers $A = \{a_1, \ldots, a_n\}$, are there two disjoint nonempty subsets $X, Y \subseteq A$ such that $sum(X) = r \cdot sum(Y)$?*

For $r = 1$ the problem is Equal Sum Subsets and therefore NP-complete [9]. We show that Factor-r Sum Subsets is actually NP-complete for all $r \in \mathbb{Q}^+$. The proof consists of two different reductions from One-in-Three 3-Satisfiability, where the second reduction is just for the case $r = 2$ and involves an argument about the connectivity graph of a Boolean formula. One-in-Three 3-Satisfiability is NP-complete [3] and defined as follows: Given a CNF Boolean formula consisting of clauses with three positive literals each, is there a (satisfying) assignment that satisfies exactly one literal per clause?

Lemma 1. One-in-Three 3-Satisfiability $\leq_p$ Factor-r Sum Subsets *for any* $r \in \mathbb{Q}^+,\ r \notin \{1, 2, \frac{1}{2}\}$.

Proof. Let $r = p/q$, where p, q are positive integers with no common divisor except 1 (coprimes) and $p < q$ (the case $p > q$ is equivalent by interchanging sets X and Y in problem definition). We distinguish several cases, depending on the values of p and q. We only give a detailed proof for the first case; for the other cases the proof is quite similar, so we just mention the construction of the necessary numbers.

CASE 1: $p > 3$. Consider an instance of ONE-IN-THREE 3-SATISFIABILITY with a set of n variables $V = \{v_1, \ldots, v_n\}$ and a set of m clauses $C = \{c_1, \ldots, c_m\}$. An instance of FACTOR-r SUM SUBSETS is constructed as follows. For each variable v_i, a number $a_i = \sum_{v_i \in c_j} \Delta_m(j)$ is created (i.e. a_i has m digits and its non-zero digits correspond to clauses where x_i appears). Two additional numbers a_{n+1} and a_{n+2} are constructed which are multiples of $\mathbb{1}_m$: $a_{n+1} = (p-1) \cdot \mathbb{1}_m$ and $a_{n+2} = q \cdot \mathbb{1}_m$. For all numbers we assume base $B = q(p+q+2)+1$ (this way we avoid carry-digits when adding a_i's). Let $A = \{a_1, \ldots, a_{n+2}\}$. We show below that there is an 1-in-3 satisfying assignment for the variables in V satisfying exactly one literal in each clause in C if and only if there are two disjoint nonempty subsets $X, Y \subseteq A$ such that $\text{sum}(X) = r \cdot \text{sum}(Y)$.

"only if": The existence of an 1-in-3 satisfying assignment implies that there exists a subset $R \subseteq \{a_1, \ldots, a_n\}$ such that $\text{sum}(R) = \mathbb{1}_m$: for each clause c_j, there is exactly one of the three variables in c_j set to TRUE, say x_k, and the corresponding a_k has a one in the j-th digit. By setting $X = R \cup \{a_{n+1}\}$ and $Y = \{a_{n+2}\}$ we have

$$\text{sum}(X) = p \cdot \mathbb{1}_m = r \cdot q \cdot \mathbb{1}_m = r \cdot \text{sum}(Y)$$

"if": Assume that X and Y exist such that $\text{sum}(X) = r \cdot \text{sum}(Y)$; equivalently,

$$q \cdot \text{sum}(X) = p \cdot \text{sum}(Y)$$

Since the base of our numbers is sufficiently large ($B = q(p+q+2)+1$), we have that the sum of all numbers in A consists of m digits that are all equal to $p+q+2$ ($\text{sum}(A) = (p+q+2) \cdot \mathbb{1}_m$) and therefore $\text{sum}(X) + \text{sum}(Y)$ also consists of m digits of value at most $p+q+2$. Notice that for each $i \leq m$ the i-th digit of $\text{sum}(X) + \text{sum}(Y)$ can be the sum of at most five numbers: $1, 1, 1, p-1$, and q. We will argue that the only way to have $\text{sum}(X)/\text{sum}(Y) = p/q$ is if each digit of $\text{sum}(X)$ is equal to p and each digit of $\text{sum}(Y)$ is equal to q.

Let $Z_X = q \cdot \text{sum}(X)$ and $Z_Y = p \cdot \text{sum}(Y)$. We will make use of the equality $Z_X = Z_Y$. Notice that, again due to the sufficiently large base B, even if we add all numbers in A q times no carry-digits will occur; hence the same happens if we add numbers in X q times or numbers in Y p times. This means that the i-th bit of Z_X is equal to qx_i, where x_i is the i-th bit of $\text{sum}(X)$, and the i-th bit of Z_Y is equal to py_i, where y_i is the i-th bit of $\text{sum}(Y)$. Therefore, for all $1 \leq i \leq m$ we have $qx_i = py_i$ which implies that either $x_i = y_i = 0$ or q divides y_i and p divides x_i; since $x_i + y_i \leq p+q+2$ and $q > p > 3$, we get $x_i = p$ and $y_i = q$ for some i (there must be non-zero digits since we assumed non-empty X and Y).

It is not difficult to see that this can only be achieved if $Y = \{a_{n+2}\}$ and $X = \{a_{n+1}\} \cup R$, where $R \subseteq A$ and $\text{sum}(R) = \mathbb{1}_m$. The variables corresponding to numbers in R form an 1-in-3 satisfying assignment for the given clauses.

CASE 2: $p = 3, q > 4$. $a_1, \ldots, a_n$ as in Case 1, $a_{n+1} = 3 \cdot \mathbb{1}_m$, $a_{n+2} = (q-1) \cdot \mathbb{1}_m$.

CASE 3: $p = 3, q = 4$. $a_1, \ldots, a_n$ as in Case 1, $a_{n+1} = 3 \cdot \mathbb{1}_m$, $a_{n+2} = 2 \cdot \mathbb{1}_m$.

CASE 4: $p = 2, q > 3$. $a_1, \ldots, a_n$ as in Case 1, and only one additional number $a_{n+1} = (q-1) \cdot \mathbb{1}_m$ is constructed.

CASE 5: $p = 2, q = 3$. For each variable v_i, $a_i = \sum_{v_i \in c_j} 3 \cdot \Delta_m(j)$, i.e. a_i has a digit 3 in each position that corresponds to a clause that contains v_i. We also set $a_{n+1} = \mathbb{1}_m$. Note that $\text{sum}(A) = 10 \cdot \mathbb{1}_m$.

Again, "only if" is easy: the satisfying assignment corresponds to numbers that add up to $3 \cdot \mathbb{1}_m$ which together with a_{n+1} consitute X. For the "if" direction we observe that the only way to have the required ratio is by having two sets X, Y such that $\text{sum}(X) = 4 \cdot \mathbb{1}_m, \text{sum}(Y) = 6 \cdot \mathbb{1}_m$; this implies $a_{n+1} \in X$ and hence the variables corresponding to $X - \{a_{n+1}\}$ constitute an 1-in-3 satisfying assignment.

CASE 6: $p = 1, q > 2$. In this case $a_1, \ldots, a_n$ are constructed as in Case 1. There is only one additional number $a_{n+1} = q \cdot \mathbb{1}_m$. □

Lemma 2. ONE-IN-THREE 3-SATISFIABILITY $\leq_p$ FACTOR-2 SUM SUBSETS.

Proof. We use a restricted, but still NP-hard version of ONE-IN-THREE 3-SATISFIABILITY for our reduction to FACTOR-2 SUM SUBSETS. Given a ONE-IN-THREE 3-SATISFIABILITY instance with variables $x_1, \ldots, x_n$ and clauses $c_1, \ldots, c_m$ with only positive literals, let $G = (V, E)$ be the graph with vertices $V = \{x_1, \ldots, x_n\}$ (i.e., each variable corresponds to a vertex) and, for $i, j = 1, \ldots, n$, edges $(x_i, x_j) \in E$ if and only if x_i and x_j both occur in a clause c_k, for some $k \in \{1, \ldots, n\}$. The ONE-IN-THREE 3-SATISFIABILITY variation in which the corresponding graph G is connected is still NP-hard, because we could use a polynomial algorithm for this variation to solve the unrestricted ONE-IN-THREE 3-SATISFIABILITY problem by applying the algorithm for each component of the corresponding graph.

We reduce ONE-IN-THREE 3-SATISFIABILITY with a connected graph to FACTOR-2 SUM SUBSETS as follows: Assume the satisfiability instance has n variables $x_1, \ldots, x_n$ and m clauses $c_1, \ldots, c_m$. We construct an instance of FACTOR-2 SUM SUBSETS by creating exactly one number a_i for each variable x_i with $a_i = \sum_{x_i \in c_j} \Delta_n(j)$, where we set the $j-$th digit to 1, if x_i appears as a literal in clause c_j. We let the base B of these numbers be 7.

Assume that we have an 1-in-3 satisfying assignment for the variables of the ONE-IN-THREE 3-SATISFIABILITY instance. We then construct a solution X, Y of the FACTOR-2 SUM SUBSETS instance, where Y contains all numbers a_i for which the corresponding variable x_i has been set to TRUE, and X contains all remaining numbers. Thus, $sum(Y) = \langle 1, 1, \ldots, 1 \rangle$ and $sum(X) = \langle 2, 2, \ldots, 2 \rangle$, and therefore $sum(X) = 2 \cdot sum(Y)$.

Now assume that we are given a solution X, Y of the FACTOR-2 SUM SUBSETS instance with $sum(X) = 2 \cdot sum(Y)$. Since each digit is set to one in exactly three of the numbers a_i, and since no carry-digits can occur when summing up the a_i's because base B is sufficiently large, $sum(Y)$ must contain only ones (and zeros) in its digits and $sum(X)$ contains only twos (and zeros). Since the sets cannot be empty, at least one digit must be set to one. We assign the value TRUE to a variable x_i with corresponding number a_i if $a_i \in Y$, and we assign the value FALSE, if $a_i \in X$. Thus, if a clause $c_j = (x_f, x_g, x_h)$ exists, then either one of the three numbers a_f, a_g, or a_h is in Y and the other to numbers are in X, or neither X nor Y contain a_f, a_g, or a_h. In the latter case, we know that $sum(X)$ and $sum(Y)$ contain a zero at position j.

However, the numbers $sum(X)$ and $sum(Y)$ cannot contain any zero digits because of the connectedness of graph G. In order to see this, assume for the sake of contradiction that $sum(Y)$ contains zero digits. Then, $sum(X)$ must contain zero digits at the same positions. Let digit j be such a zero and let $c_j = (x_f, x_g, x_h)$ be the corresponding clause. Consider the set S of all variables that occur in clauses which represent zero digits. Then the subgraph of G with only the vertices corresponding to variables from set S must be a component in the graph G without any edges to other vertices, because, if such an edge would exist, it would imply that the corresponding digit is not set to zero in either $sum(X)$ or $sum(Y)$. To see this, consider an edge $e = (x_f, x_g)$ arising from clause $c_j = (x_f, x_g, x_h)$ with $x_f \in S$ and $x_g \notin S$. Then $a_g \in X \cup Y$, but a_f (and a_h) must be in $X \cup Y$ as well, in order to achieve the factor 2 in the j-th digit. □

Since FACTOR-r SUM SUBSETS is obviously in NP and since ONE-IN-THREE 3-SATISFIABILITY is NP-hard, Lemmas 1 and 2 and the NP-completeness of EQUAL SUM SUBSETS imply:

Theorem 3. FACTOR-r SUM SUBSETS *is NP-complete for all $r \in \mathbb{Q}^+$.*

4 NP-completeness of k EQUAL SUM SUBSETS

The second variation of EQUAL SUM SUBSETS that we study is called k EQUAL SUM SUBSETS. For an integer $k \geq 2$ it is defined as follows:

Definition (k EQUAL SUM SUBSETS). *Given a multi-set[2] of n numbers $\{a_1, \ldots, a_n\}$, are there $k \geq 2$ non-identical subsets $X_1, \ldots, X_k \subseteq \{a_1, \ldots, a_n\}$ with $sum(X_1) = \ldots = sum(X_k)$?*

k EQUAL SUM SUBSETS is a very natural generalization of EQUAL SUM SUBSETS (which is the case $k = 2$) and it is an interesting problem for its own sake. We present a reduction from ALTERNATING PARTITION which is the following NP-complete [3] variation of PARTITION: Given n pairs of numbers $(u_1, v_1), \ldots, (u_n, v_n)$, are there two disjoint sets of indices I and J with $I \cup J = \{1, \ldots, n\}$ such that $\sum_{i \in I} u_i + \sum_{j \in J} v_j = \sum_{i \in I} v_i + \sum_{j \in J} u_j$ (equivalently, $\sum_{i \in I} u_i + \sum_{j \in J} v_j = \sum_{i \notin I} u_i + \sum_{j \notin J} v_j$)?

Theorem 4. k EQUAL SUM SUBSETS *is NP-complete.*

Proof. The problem is obviously in NP. To show NP-hardness, we reduce ALTERNATING PARTITION to it. We transform a given ALTERNATING PARTITION instance with pairs $(u_1, v_1), \ldots, (u_n, v_n)$ into a k EQUAL SUM SUBSETS instance as follows: For each pair (u_i, v_i), we create two numbers $u'_i = \langle u_i \rangle \lhd \Delta_n(i)$ and $v'_i = \langle v_i \rangle \lhd \Delta_n(i)$. In addition, we create $k-2$ (equal) numbers $c_1, \ldots, c_{k-2}$ with $c_i = \langle \frac{1}{2} \sum_i (u_i + v_i) \rangle \lhd \mathbf{1}(n)$. While we let the base of the first digit be $k \cdot \sum_i (u_i + v_i)$, all other digits have base $n + 1$ in order to ensure that no carry-digits can occur in any additions.

To see how this reduction works, assume first that we are given a solution of the ALTERNATING PARTITION instance, i.e., two indices sets G and H. We create k equal sum subsets $S_1, \ldots, S_k$: for $k = 1, \ldots, k-2$ we have $S_i = \{c_i\}$; for the remaining two subsets, we let $u'_i \in S_{k-1}$, if $i \in G$, and $v'_i \in S_{k-1}$, if $i \in H$, and we let $u'_i \in S_k$, if $i \in H$, and $v'_i \in S_k$, if $v_i \in G$.

Now assume we are given a solution of the k EQUAL SUM SUBSETS instance, i.e., k equal sum subsets $S_1, \ldots, S_k$. Since each of the n right-most digits (i.e., the base $n+1$ digits) is set to one in exactly k numbers, we can assume w.l.o.g. that $S_i = \{c_i\}$ for $i = 1, \ldots, k-2$. The remaining two subsets naturally form an alternating partition as u'_i and v'_i can never be in the same subset for any $i = 1, \ldots, n$. All numbers u'_i and v'_i must occur in one of the remaining two subsets in order to match the ones in the base $n + 1$ digits of the other subsets. Matching the first digit gives us the equal sum subsets. □

Note that this proof works as well, if we require the subsets of k EQUAL SUM SUBSETS to be disjoint and non-empty (rather than non-identical).

[2]We allow multi-sets for this problem. The NP-completeness proof for this problem without allowing multi-sets is very similar to the one given in Theorem 4. However, it is more technical and therefore omitted.

5 NP-completeness of Equal Sum Subsets Variations with Additional Requirements

As a further class of NP-complete variations of Equal Sum Subsets, we study problems where we add specific requirements that a solution must fulfill. This approach allows us to explore the sphere of NP-completeness that forms around Equal Sum Subsets. We focused on quite natural additional requirements. The problems are defined as follows:

Definition (Equal Sum Subsets with Enforced Element). *Given a set of n numbers $A = \{a_1, \dots, a_n\}$, are there two disjoint subsets $X, Y \subseteq A$ with $a_n \in X$ such that $sum(X) = sum(Y)$?*

Definition (Equal Sum Subsets of Different Cardinality). *Given a set of n numbers $A = \{a_1, \dots, a_n\}$, are there two disjoint nonempty subsets $X, Y \subseteq A$ with $|X| \neq |Y|$ such that $sum(X) = sum(Y)$?*

Definition (Equal Sum Subsets of Different by One Cardinality). *Given a set of n numbers $A = \{a_1, \dots, a_n\}$, are there two disjoint subsets $X, Y \subseteq A$ with $|X| = |Y| + 1$ such that $sum(X) = sum(Y)$?*

Definition (Equal Sum Subsets of Equal Cardinality). *Given a set of n numbers $A = \{a_1, \dots, a_n\}$, are there two disjoint nonempty subsets $X, Y \subseteq A$ with $|X| = |Y|$ such that $sum(X) = sum(Y)$?*

Definition (Alternating Partial Partition). *Given n pairs of numbers $(u_1, v_1), \dots, (u_n, v_n)$, are there two disjoint nonempty sets of indices I and J such that $\sum_{i \in I} u_i + \sum_{j \in J} v_j = \sum_{i \in I} v_i + \sum_{j \in J} u_j$?*

The NP-completeness of the first three problems is shown by giving reductions from Alternating Partition. After that we reduce Equal Sum Subsetsto Alternating Partial Partition, and then the latter to Equal Sum Subsets of Equal Cardinality to establish the NP-hardness of these two problems.

Lemma 5. Alternating Partition $\leq_p$ Equal Sum Subsets with Enforced Element.

Proof. Let $(u_1, v_1), \dots, (u_n, v_n)$ be the input pairs for Alternating Partition. Let $S = \sum_{i=1}^{n} (u_i + v_i)$, $a_i = \langle u_i \rangle \lhd\ \Delta_n(i)$ and $b_i := \langle v_i \rangle \lhd\ \Delta_n(i)$ for all $1 \leq i \leq n$, and $c = \langle \frac{S}{2} \rangle \lhd\ \mathbb{1}_n$. As usual, we use a base large enough such that no carry digits occur.

Let $\{a_i \mid 1 \leq i \leq n\} \cup \{b_i \mid 1 \leq i \leq n\} \cup \{c\}$ be the input for Equal Sum Subsets with Enforced Element. Then c is the enforced element. There exists a solution for the Alternating Partition instance if and only if there exists a solution for the Equal Sum Subsets with Enforced Element instance.

"only if": Let I and J be a solution for Alternating Partition. Then $\sum_{i \in I} u_i + \sum_{j \in J} v_j = \frac{S}{2}$. We define $X := \{c\}$ and $Y := \{a_i \mid i \in I\} \cup \{b_j \mid j \in J\}$. Then

$$\text{sum}(Y) = \sum_{i \in I} a_i + \sum_{j \in J} b_j$$

$$
\begin{aligned}
&= \sum_{i\in I}(\langle u_i\rangle \lhd\ \Delta_n(i)) + \sum_{j\in J}(\langle v_j\rangle \lhd\ \Delta_n(j)) \\
&= \langle \sum_{i\in I} u_i + \sum_{j\in J} v_j\rangle \lhd\ (\sum_{i\in I}\Delta_n(i) + \sum_{j\in J}\Delta_n(j)) \\
&= \langle \frac{S}{2}\rangle \lhd \sum_{i=1}^{n}\Delta_n(i) \\
&= \langle \frac{S}{2}\rangle \lhd \mathbb{1}_n \\
&= \mathrm{sum}(X).
\end{aligned}
$$

"if": Let X, Y be a solution for the Equal Sum Subsets with Enforced Element instance. Assume w.l.o.g. $c \in X$. All numbers in the input have $n+1$ digits. For each index $i \in \{2, \ldots, n+1\}$, only three numbers, namely c, a_i and b_i, have a one in the i'th digit, all other numbers in the input have a zero in the i'th digit. For each digit the sum over all elements in X and in Y yields the same result. Therefore, since $c \in X$, exactly one of a_i or b_i will be in Y for each $1 \le i \le n$, and $X = \{c\}$, since any other element would add a second one in some digit i, which then could not be equalized by elements in Y. Summing up the first digit of all elements in Y yields exactly the first digit of c, which is $\frac{S}{2}$. Thus, $I = \{i \in \{1, \ldots, n\} \mid a_i \in Y\}$ and $J = \{j \in \{1, \ldots, n\} \mid b_j \in Y\}$ yields a solution for the Alternating Partition instance. □

Lemma 6. Alternating Partition $\le_p$ Equal Sum Subsets of Different Cardinality.

Proof (sketch) The proof follows along the lines of the previous reduction. Each number a_i in one set enforces b_i to be in the other set, and vice versa. Thus, they yield sets X and Y of equal cardinalities. Therefore, element c has to be in either X or Y.

Lemma 7. Alternating Partition $\le_p$ Equal Sum Subsets of Different by One Cardinality.

Proof. This proof is similar to the previous proofs, except that we add n dummy elements to blow up the cardinality of subset which contains c.

Let $(u_1, v_1), \ldots, (u_n, v_n)$ be the input pairs for Alternating Partition. Let $S := \sum_{i=1}^{n}(u_i + v_i)$ and $M := n \cdot 2^{n+2}$. Define $a_i := \langle u_i\rangle \lhd\ \Delta_n(i) \lhd\ \langle \frac{M}{n}\rangle$ and $b_i := \langle v_i\rangle \lhd\ \Delta_n(i) \lhd\ \langle \frac{M}{n}\rangle$ for all $1 \le i \le n$. Define $c := \langle \frac{S}{2}\rangle \lhd\ \mathbb{1}_n \lhd\ \langle M - (2^n - 1)\rangle$. For $1 \le k \le n$, we define dummy elements $d_k = \langle 0\rangle \lhd\ \mathbb{0}_n \lhd\ \langle 2^{k-1}\rangle$.

As before, any partial partition with only a_i's and b_i's will have equal cardinality. Thus, c will be in one of the sets, say X, and n of the a_i's and b_i's will be in the other set Y to achieve equal sums in the first $n+1$ digits of the elements in X and Y. To achieve an equal sum in the last digit as well, d_k must bebe in set X for all $1 \le k \le n$. □

Lemma 8. Equal Sum Subsets $\le_p$ Alternating Partial Partition.

Proof. Given an instance of Equal Sum Subsets, i.e. a set of numbers $A = \{a_1, \ldots, a_n\}$, we reduce it to an instance of Alternating Partial Partition by mapping each number a_i to a pair (u_i, v_i) with $u_i = a_i$ and $v_i = 0$. Clearly, if there are disjoint sets $X, Y \subseteq A$ such that $\mathrm{sum}(X) = \mathrm{sum}(Y)$ then there are disjoint sets of indices $I = \{i \mid a_i \in X\}$ and $J = \{j \mid a_j \in Y\}$

such that $\sum_{i \in I} u_i + \sum_{j \in J} v_j = \sum_{i \in I} v_i + \sum_{j \in J} u_j$. Conversely, if there is an alternating partial partition of the resulting pairs, i.e. appropriate sets of indices I, J, then the sets $X = \{a_i \mid i \in I\}$ and $X = \{a_j \mid j \in J\}$ form a partial partition of the original set A. □

Lemma 9. ALTERNATING PARTIAL PARTITION $\leq_p$ EQUAL SUM SUBSETS OF EQUAL CARDINALITY.

Proof. Given an instance of ALTERNATING PARTIAL PARTITION we map each pair (u_i, v_i) to two numbers $a_i = \langle u_i \rangle \lhd\ \Delta_n(i)$, $a'_i = \langle v_i \rangle \lhd\ \Delta_n(i)$; i.e., we make new numbers with u_i (resp. v_i) as the most significant digits, followed by n digits, the i-th of which is a 1 and all the rest are 0's. We then set A to consist of all a_i's and all a'_i's. For the numbers we use base $B = \sum_{i=1}^{n}(u_i + v_i) + 1$, which is sufficiently large such that adding any subset of numbers in A never gives carry digits. Thus, we can have an equal cardinality partial partition in A if and only if for each a_i on one side there is a'_i on the other side, and vice versa, since only a_i and a'_i have a one in the $i+1$'th digit. This is equivalent to having an alternating partial partition in the original instance. □

From the previous lemmas and the fact that the problems are obviously in NP, we get the following:

Theorem 10. *The problems*
ALTERNATING PARTIAL PARTITION,
EQUAL SUM SUBSETS OF EQUAL CARDINALITY,
EQUAL SUM SUBSETS OF DIFFERENT CARDINALITY,
EQUAL SUM SUBSETS OF DIFFERENT BY ONE CARDINALITY, *and*
EQUAL SUM SUBSETS WITH ENFORCED ELEMENT
are NP-complete.

6 NP-completeness of Finding Equal Sum Subsets from Two Sets (and its Variations)

As a last set of NP-complete problems, we investigate EQUAL SUM SUBSETS FROM TWO SETS and its variations.

Definition (EQUAL SUM SUBSETS FROM TWO SETS). *Given two sets of numbers $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$, are there two nonempty subsets $U \subseteq A$ and $V \subseteq B$ such that sum(U) = sum(V)?*

We study this problem in order to explore the limits of the sphere of NP-completeness around ALTERNATING PARTIAL PARTITION to which the EQUAL SUM SUBSETS FROM TWO SETS is closely related. ALTERNATING PARTIAL PARTITION is the "partial" equivalent of ALTERNATING PARTITION, which is an important, well-known variation of PARTITION (see [3]). We show NP-completeness of EQUAL SUM SUBSETS FROM TWO SETS by proposing a reduction from SUBSET SUM, which is defined as follows: Given a set of n numbers $P = \{p_1, \ldots, p_n\}$ and a number S, is there a subset $X \subseteq P$ such that $\text{sum}(X) = S$?

Lemma 11. SUBSET SUM $\leq_p$ EQUAL SUM SUBSETS FROM TWO SETS.

Proof. Let $\{p_1, \ldots, p_n\}$ and S be an instance of Subset Sum. Let $A := \{p_1, \ldots, p_n\}$ and $B := \{S\}$ be an instance of Equal Sum Subsets from Two Sets. If X is a solution for the Subset Sum instance, then $X \subseteq A$ and $\text{sum}(X) = S$. Any solution $U \subseteq A$ and $V \subseteq B$ for the Equal Sum Subsets from Two Sets instance will have $V = B = \{S\}$, and therefore $\text{sum}(U) = S$. Thus, a solution for the Subset Sum instance transforms easily in a solution for the Equal Sum Subsets from Two Sets instance, and vice versa. □

In an approach similar to the one followed in Section 5, we define restricted variations of Equal Sum Subsets from Two Sets. We present NP-completeness results that give insight as to how far NP-completeness goes.

Definition (Equal Sum Subsets of Equal Cardinality from Two Sets). *Given two sets of numbers $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$, are there two nonempty subsets $U \subseteq A$ and $V \subseteq B$ with $|U| = |V|$ such that sum(U) = sum(V)?*

Lemma 12. Subset Sum $\leq_p$ Equal Sum Subsets of Equal Cardinality from Two Sets.

Proof. Given an instance $\{p_1, \ldots, p_n\}$ and S of Subset Sum we construct an instance of Equal Sum Subsets of Equal Cardinality from Two Sets, i.e. sets A and B as follows:

set A	set B
$a_1 := \langle p_1, 1, 0\rangle$	$b_1 := \langle 0, 1, 0\rangle$
$\vdots$	$\vdots$
$a_i := \langle p_i, i, 0\rangle$	$b_i := \langle 0, i, 0\rangle$
$\vdots$	$\vdots$
$a_n := \langle p_n, n, 0\rangle$	$b_n := \langle 0, n, 0\rangle$
$a_{n+1} := \langle 0, 0, 1\rangle$	$b_{n+1} := \langle S, 0, 1\rangle$

We will show that there is a set $X \subseteq \{p_1, \ldots, p_n\}$ such that $\text{sum}(X) = S$ if and only if there are nonempty sets $U \subseteq A$ and $V \subseteq B$ such that $|U| = |V|$ and $\text{sum}(U) = \text{sum}(V)$.
"only if": If there is a set $X \subseteq \{p_1, \ldots, p_n\}$ such that $\text{sum}(X) = S$ then, by defining $U = \{a_i \mid x_i \in X\} \cup \{a_{n+1}\}$ and $V = \{b_i \mid x_i \in X\} \cup \{b_{n+1}\}$ we have that $\text{sum}(U) = \text{sum}(V) = \langle S, k, 1\rangle$, where $k = \sum_{x_i \in X} i$.
"if": Assume that nonempty sets $U \subseteq A$ and $V \subseteq B$ exist such that $|U| = |V|$ and $\text{sum}(U) = \text{sum}(V)$. Then $b_{n+1} \in V$ is necessary to have equal sums in the first digit. This implies that there are a_i's in U such that $\text{sum}(\{p_i \mid a_i \in U\}) = S$, i.e., the corresponding p_i's form a solution for the original Subset Sum instance. □

The following variation asks for two equal sum subsets that have disjoint indices:

Definition (Equal Sum Subsets with Disjoint Indices from Two Sets). *Given two sets of n numbers $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$, are there two nonempty sets of indices $I, J \subseteq \{1, \ldots, n\}$ with $I \cap J = \emptyset$ such that $\sum_{i \in I} a_i = \sum_{j \in J} b_j$?*

Lemma 13. Equal Sum Subsets from Two Sets $\leq_p$ Equal Sum Subsets with Disjoint Indices from Two Sets.

Proof. *(Sketch)* Given an instance $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$ of EQUAL SUM SUBSETS FROM TWO SETS, we can construct an instance of EQUAL SUM SUBSETS WITH DISJOINT INDICES FROM TWO SETS (A', B') as follows:

	set A'	set B'
$\forall 1 \le i \le n:$	$a'_i := \langle a_i \rangle \lhd \mathbf{0}_n \lhd \mathbf{0}_n$	$b'_i := \langle 0 \rangle \lhd \Delta_n(i) \lhd \mathbf{0}_n$
$\forall 1 \le i \le n:$	$a'_{n+i} := \langle 0 \rangle \lhd \mathbf{0}_n \lhd \Delta_n(i)$	$b'_{n+i} := \langle b_i \rangle \lhd \mathbf{0}_n \lhd \mathbf{0}_n$

It is easy to see that there are two equal sum subsets of A and B if and only if there equal sum subsets of A' and B' with disjoint indices, since only subsets of the first n numbers in A' and the last n numbers in B' can yield equal sums. □

An even more restricted variation asks for subsets with disjoint indices that cover the whole set of indices.

Definition (EQUAL SUM SUBSETS WITH DISJOINT COVERING INDICES FROM TWO SETS). *Given two sets of n numbers $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$, are there two sets of indices $I, J \subseteq \{1, \dots, n\}$ with $I \cap J = \emptyset$ and $I \cup J = \{1, \dots, n\}$ such that $\sum_{i \in I} a_i = \sum_{j \in J} b_j$?*

Lemma 14. PARTITION $\le_p$ EQUAL SUM SUBSETS WITH DISJOINT COVERING INDICES FROM TWO SETS.

Proof. Given an instance of PARTITION $\{a_1, \dots, a_n\}$ we construct an instance of EQUAL SUM SUBSETS WITH DISJOINT COVERING INDICES FROM TWO SETS by setting $A' = B' = A$. Now, if A can be partitioned into X and Y, then choosing the corresponding elements in A' and B' respectively gives us a solution for the EQUAL SUM SUBSETS WITH DISJOINT COVERING INDICES FROM TWO SETS instance, and vice versa. □

We finally examine the variation where we want the sets of indices to be identical.

Definition (EQUAL SUM SUBSETS WITH IDENTICAL INDICES FROM TWO SETS). *Given two sets of n numbers $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$, is there a nonempty set of indices $I \subseteq \{1, \dots, n\}$ such that $\sum_{i \in I} a_i = \sum_{i \in I} b_i$?*

Lemma 15. SUBSET SUM $\le_p$ EQUAL SUM SUBSETS WITH IDENTICAL INDICES FROM TWO SETS.

Proof. We use the same reduction as in Lemma 12. It suffices to observe that any two equal sum subsets $U \subseteq A$ and $V \subseteq B$ either have identical indices or there is always $V' \subseteq B$ such that $\mathrm{sum}(V) = \mathrm{sum}(V') = \mathrm{sum}(U)$ and V' has identical indices with U. □

From the previous lemmas and the fact that the problems are obviously in NP, we get the following:

Theorem 16. *The problems*

EQUAL SUM SUBSETS FROM TWO SETS,
EQUAL SUM SUBSETS OF EQUAL CARDINALITY FROM TWO SETS,
EQUAL SUM SUBSETS WITH DISJOINT INDICES FROM TWO SETS,
EQUAL SUM SUBSETS WITH DISJOINT COVERING INDICES FROM TWO SETS, *and*
EQUAL SUM SUBSETS WITH IDENTICAL INDICES FROM TWO SETS

are NP-complete.

7 Conclusions

We have presented NP-completeness results for many natural and interesting variations of Equal Sum Subsets.

The results in this paper are only a first step in investigating variations of Equal Sum Subsets. A line of future research is to further explore the brink of NP-completeness in terms of Equal Sum Subsets variations. Potential examples of such variations are: a variation with additive factor (instead of multiplicative) or a variation in which the required cardinalities of the two subsets are given as part of the input. Moreover, cases where the input contains negative numbers deserve consideration. It would also be interesting to study some of our variations in their full partition counterparts.

Acknowledgments

We would like to thank Christoph Ambühl, Zsuzsanna Lipták, Paolo Penna, and Peter Widmayer for their help in this work.

References

[1] C. Bazgan, M. Santha and Zs. Tuza; *Efficient approximation algorithms for the Subset-Sum Equality problem*; Proc. ICALP'98, pp. 387–396.

[2] Mark Cieliebak, Stephan Eidenbenz, Aris Pagourtzis, Paolo Penna; *Variations of the Partial Digest Problem*; Manuscript, to be published 2002.

[3] M.R. Garey and D.S. Johnson; *Computers and Intractability: A Guide to the Theory of NP-completeness*; Freeman, San Fransisco, 1979.

[4] R.M. Karp; *Reducibility among combinatorial problems*; in R.E. Miller and J.W. Thatcher (eds.), Complexity of Computer Computations, Plenum Press, New York, pp. 85 – 103, 1972.

[5] S. Martello and P. Toth; *Knapsack Problems*; John Wiley & Sons, Chichester, 1990.

[6] C.H. Papadimitriou; *On the complexity of the parity argument and other inefficient proofs of existence*; Journal of Computer and System Sciences, 48, pp. 498 – 532, 1994.

[7] C.H. Papadimitriou; *Computational Complexity*; Addison-Wesley Publishing Company, 1994.

[8] P. Pevzner; *Computational Molecular Biology*; MIT Press, 2000.

[9] G.J. Woeginger and Z.L. Yu; *On the equal-subset-sum problem*; Information Processing Letters, 42(6), pp. 299–302, 1992.