



HAL
open science

The Maximum Zero-Sum Partition Problem

Guillaume Fertin, Oscar Fontaine, Géraldine Jean, Stéphane Vialette

► **To cite this version:**

Guillaume Fertin, Oscar Fontaine, Géraldine Jean, Stéphane Vialette. The Maximum Zero-Sum Partition Problem. 25th International Computer Symposium, ICS 2022, Dec 2022, Taoyuan, Taiwan. pp.73-85, 10.1007/978-981-19-9582-8_7. hal-04293802

HAL Id: hal-04293802

<https://hal.science/hal-04293802>

Submitted on 11 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The MAXIMUM ZERO-SUM PARTITION Problem

Guillaume Fertin¹[0000-0002-8251-2012], Oscar Fontaine², Géraldine Jean¹[0000-0002-1534-2682], and Stéphane Vialette³[0000-0003-2308-6970]

¹ Nantes Université, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

² Ecole Normale Supérieure PSL, DMA, F-75230 Paris, France

³ LIGM, CNRS, Université Gustave Eiffel, F-77454 Marne-la-Vallée, France

{guillaume.fertin,geraldine.jean}@univ-nantes.fr,

oscar.fontaine@ens.psl.eu, stephane.vialette@univ-eiffel.fr

Abstract. We study the MAXIMUM ZERO-SUM PARTITION problem (or MZSP), defined as follows: given a multiset $\mathcal{S} = \{a_1, a_2, \dots, a_n\}$ of integers $a_i \in \mathbb{Z}^*$ such that $\sum_{i=1}^n a_i = 0$, find a maximum cardinality partition $\{S_1, S_2, \dots, S_k\}$ of \mathcal{S} such that, for every $1 \leq i \leq k$, $\sum_{a_j \in S_i} a_j = 0$. Solving MZSP is useful in genomics for computing evolutionary distances between pairs of species. Our contributions are a series of algorithmic results concerning MZSP, in terms of complexity, (in)approximability, with a particular focus on the fixed-parameter tractability of MZSP with respect to either (i) the size k of the solution, (ii) the number of negative (resp. positive) values in \mathcal{S} and (iii) the largest integer in \mathcal{S} .

1 Introduction

In this paper, we study the MAXIMUM ZERO-SUM PARTITION (or MZSP).

MAXIMUM ZERO-SUM PARTITION (MZSP)

Instance : A multiset $\mathcal{S} = \{a_1, a_2, \dots, a_n\}$ of numbers $a_i \in \mathbb{Z}^*$ s.t. $\sum_{i=1}^n a_i = 0$.

Output : A maximum cardinality partition $\mathcal{S} = \{S_1, S_2, \dots, S_k\}$ of \mathcal{S} such that, for every $1 \leq i \leq k$, $\sum_{a_j \in S_i} a_j = 0$.

This problem emerged in bioinformatics, where determining the distance between two genomes was studied [1]. The distance formula in that context depends on a parameter to be optimized, which directly relates to answering MZSP. However, the MZSP problem in itself was not central to the results in [1], and thus has not been studied *per se* in that paper. Hence, the goal of the present paper is to extensively study the MZSP problem from an algorithmic point of view.

Definitions and Notations. For any integer n , $\llbracket 1, n \rrbracket$ denotes the set of integers from 1 to n . Given a (multi)set S of integers and an integer p , we say that S *sums to p* when the sum of the elements of S is equal to p . When $p = 0$, we say that S is a *zero-sum (multi)set*. For any instance \mathcal{S} of MZSP, we let **neg** (resp. **pos**) denote the number of negative (resp. positive) integers in \mathcal{S} and $m = \min\{\mathbf{neg}, \mathbf{pos}\}$. We denote by n^* the number of distinct values in \mathcal{S} , by $b = \sum_{i=1}^n \lceil \log_2(|a_i|) \rceil$ the

number of bits needed to encode \mathcal{S} (e.g. when \mathcal{S} is stored as a list, in which each element is binary encoded), and by $B = \max_{i \in \llbracket 1, n \rrbracket} \{|a_i|\}$. The cardinality of an optimal partition of \mathcal{S} , i.e. the size of the solution, is denoted by k . For example, if $\mathcal{S} = \{-7, -7, -7, -1, -1, -1, 2, 3, 3, 4, 4, 4, 4\}$, then $n = 13$, $\text{neg} = 6$, $\text{pos} = 7$, $m = 6$, $n^* = 5$, $B = 7$ and it can be seen that the optimal solution is $k = 4$: for instance, $S_1 = S_2 = \{-7, 4, 3\}$, $S_3 = \{-7, -1, 4, 4\}$ and $S_4 = \{-1, -1, 2\}$ form a solution. UNARY MZSP denotes MZSP for which unary encoding of the input instance is used. For any positive integer p , p -MZSP denotes the decision version of MZSP, in which, given p and a zero-sum integer multiset \mathcal{S} , we ask whether there exists a zero-sum partition S of \mathcal{S} such that $|S| \geq p$. We will also often use the O^* notation, as frequently done in parameterized complexity: for a given problem whose size of the input is n and parameter is k , $O^*(f(k))$ stands for $O(f(k) \cdot \text{poly}(n))$. In other words, O^* only describes the exponential part of the running time (in k) and discards the polynomial factor (in n).

First observations. Note that if we denote by $-\mathcal{S}$ the multiset \mathcal{S} to which all signs have been switched, then $-\mathcal{S}$ is a valid instance for MZSP, and both \mathcal{S} and $-\mathcal{S}$ have the same optimum k . Consequently, any result that applies to neg (resp. pos) applies to pos (resp. neg), and thus to m . Note also that an m -size zero-sum partition of \mathcal{S} is necessarily optimal, since at least one positive (resp. negative) element of \mathcal{S} needs to be present in any S_i from the partition. In other words, we always have $k \leq m$. For any given $p \in \mathbb{N}^*$, a YES-instance for p -MZSP is also a YES-instance for p' -MZSP as long as $p' \in \llbracket 1, p \rrbracket$. Indeed, merging any two sets in a size- p zero-sum partition of \mathcal{S} yields a size- $(p-1)$ zero-sum partition of \mathcal{S} . Finally, observe that if an integer a and its opposite $-a$ both belong to \mathcal{S} , then there always exists an optimal solution $S = \{S_1, S_2, \dots, S_k\}$ in which $S_i = \{-a, a\}$ for some $i \in \llbracket 1, n \rrbracket$. Indeed, suppose $-a$ and a both belong to \mathcal{S} , and observe an optimal solution $S = \{S_1, S_2, \dots, S_k\}$ in which $S_i \neq \{-a, a\}$ for every $i \in \llbracket 1, k \rrbracket$. Clearly, no S_i is such that $\{-a, a\} \subset S_i$, otherwise we could partition S_i into $\{-a, a\}$ and $S_i \setminus \{-a, a\}$, both summing to zero, contradicting the optimality of S . Thus $-a \in S_x$ and $a \in S_y$ for some $1 \leq x \neq y \leq k$. Now consider the following partition $S' = \{S'_1, S'_2, \dots, S'_k\}$ of \mathcal{S} : (i) $S'_i = S_i$ for every $i \in \llbracket 1, k \rrbracket$ such that $i \neq x$ and $i \neq y$, (ii) $S'_x = \{-a, a\}$ and (iii) $S'_y = (S_x \cup S_y) \setminus \{-a, a\}$. Every S'_i , $i \in \llbracket 1, k \rrbracket$, sums to zero, and $|S| = |S'| = k$.

In this paper, we study the MZSP problem under an algorithmic viewpoint, and, in particular, discuss its computational complexity, approximability and fixed-parameter tractability with respect to n , n^* , m , B and k (see Table 1).

2 Computational Complexity of MZSP

Theorem 1. *MZSP is strongly NP-complete, even if each S_i in the solution S contains at most four elements.*

Proof. The proof is by reduction from 3-PARTITION, which has been proved to be strongly NP-complete [3], and whose definition is as follows.

Parameter	Results
n	Strongly NP-complete (Thm 1) No $2^{o(n)}b^{O(1)}$ algorithm unless ETH fails (Thm 4) FPT (Thm 5) No approximation within ratio $O(n^{1-\epsilon})$ (Thm 7)
m	NP-complete, even if bounded (Thm 2)
k	NP-complete, even if bounded (Thm 2)
m	W[1]-hard (Thm. 8)
Unary encoded instance	XP(Cor. 2)
k	W[1]-hard (Cor. 1)
Unary encoded instance	XP(Thm 9)
n^*	No $2^{o(n^*)}b^{O(1)}$ algorithm unless ETH fails (Thm 10) XP(Thm 11)
B	FPT(Thm 13)
n^*+k	FPT(Thm 12)

Table 1. Summary of our main results, in relation to parameters n, m, k, n^*, B and b .

3-PARTITION

Instance : An integer C , a multiset $X = \{x_1, x_2, \dots, x_{3p}\}$ of integers such that (i) $\sum_{i=1}^{3p} x_i = C \cdot p$ and (ii) $\forall x_i \in X, \frac{C}{4} < x_i < \frac{C}{2}$.

Question : Does there exist a partition $\{X_1, X_2, \dots, X_p\}$ of X such that, for every $i \in \llbracket 1, p \rrbracket$, $\sum_{x_j \in X_i} x_j = C$?

Let $I = (C, X)$ be an instance of 3-PARTITION, and let \mathcal{S} be the multiset such that $\mathcal{S} = \{x_1, \dots, x_{3p}, -C, \dots, -C\}$, where $-C$ appears p times in \mathcal{S} . Note that, by definition of 3-PARTITION, the sum of all elements in \mathcal{S} is equal to zero, hence \mathcal{S} is an instance of MZSP. We now show that $I = (C, X)$ is a YES-instance for 3-PARTITION iff MZSP (with instance \mathcal{S}) has a solution of cardinality p .

(\Rightarrow) If I is a YES-instance for 3-PARTITION, there exists $t_j = (x_{i_{j,1}}, x_{i_{j,2}}, x_{i_{j,3}})$, $j \in \llbracket 1, p \rrbracket$, such that $x_{i_{j,1}} + x_{i_{j,2}} + x_{i_{j,3}} = C$. In particular, for every $j \in \llbracket 1, p \rrbracket$, $S_j = \{x_{i_{j,1}}, x_{i_{j,2}}, x_{i_{j,3}}, -C\}$ is a size- p partition of \mathcal{S} in which every S_j sums to zero. Moreover, such partition is optimal: since $\text{neg} = p$, no zero-sum partition of \mathcal{S} can contain strictly more than p sets.

(\Leftarrow) Suppose there exists a solution of MZSP of cardinality p , say $S = \{S_1, S_2, \dots, S_p\}$. Since any zero-sum subset in S contains at least one negative element, every S_j , $j \in \llbracket 1, p \rrbracket$, contains *exactly* one negative element, namely $-C$. Since, in 3-PARTITION, every x_i satisfies $\frac{C}{4} < x_i < \frac{C}{2}$, exactly 3 such elements are required to sum to C . Thus, any S_j , $j \in \llbracket 1, p \rrbracket$, contains 3 elements of the form x_i , together with $-C$. Since each S_j sums to zero, $\{x_1, \dots, x_{3p}\}$ can be partitioned in triplets, each summing to C , i.e. $I = (C, X)$ is a YES-instance for 3-PARTITION.

In unary, 3-PARTITION and MZSP are both encoded in $\Theta(p \cdot C)$ space. As 3-PARTITION is strongly NP-complete, MZSP is also strongly NP-complete. \square

As discussed above, finding a solution S to MZSP where every zero-sum set S_i in S satisfies $|S_i| = 2$ (if such a solution exists) is easy: look for two opposite values, combine them, and iterate. Theorem 1 proves that solving MZSP when every S_i contains 4 elements is (strongly) NP-complete, which rules out parameter “maximum size of an S_i ” for FPT considerations. Note that the case $|S_i| = 3$ (or equivalently $|S_i| \leq 3$) has been shown to be strongly NP-complete in [1].

Theorem 2. *MZSP is NP-complete, even when k and m are bounded.*

Proof. We show NP-completeness of MZSP in the specific case $m = k = 2$, by reduction from PARTITION which is known to be NP-complete [8].

PARTITION

Instance : A multiset $X = \{x_1, x_2, \dots, x_n\}$ of integers from \mathbb{N}^* .

Question : Does there exist a partition $\{X_1, X_2\}$ of X s.t. $\sum_{x_i \in X_1} x_i = \sum_{x_j \in X_2} x_j$?

Let X be an instance of PARTITION. We can always assume $\sum_{i=1}^n x_i$ to be even, otherwise we have a NO-instance. Thus assume $\sum_{i=1}^n x_i = 2N$. The MZSP instance we build from X is $\mathcal{S} = X \cup \{-N, -N\}$. We show that X is a YES-instance for PARTITION iff MZSP yields a size-2 zero-sum partition for \mathcal{S} .

(\Rightarrow) Suppose there exists a partition $\{X_1, X_2\}$ of X such that $\sum_{x_i \in X_1} x_i = \sum_{x_j \in X_2} x_j$. We thus have $\sum_{x_i \in X_1} x_i = \sum_{x_j \in X_2} x_j = N$, and $\{X_1 \cup \{-N\}, X_2 \cup \{-N\}\}$ is a zero-sum partition of \mathcal{S} . Moreover, this partition is optimal since $m = \text{neg} = 2$ and $k \leq m$.

(\Leftarrow) Suppose there exists a zero-sum partition of cardinality $k = 2$ of \mathcal{S} , say $\mathcal{S} = \{S_1, S_2\}$. Because $\text{neg} = 2$, we know that S_1 (resp. S_2) contains exactly one negative integer; thus, both in S_1 and S_2 , this integer is $-N$. Assume $S_1 = X_1 \cup \{-N\}$ and $S_2 = X_2 \cup \{-N\}$. In that case, $\{X_1, X_2\}$ is a partition of X , and because both S_1 and S_2 sum to zero, we have $\sum_{x_i \in X_1} x_i = \sum_{x_j \in X_2} x_j = N$. Thus X is a YES-instance for PARTITION. \square

Although we just showed that MZSP is strongly NP-complete in general, and remains NP-complete when k is bounded, we show there exists a pseudopolynomial algorithm that solves MZSP in the case $k = 2$.

Theorem 3. *2-MZSP can be solved in pseudopolynomial time.*

The following result gives two lower bounds on the time to solve MZSP, both based on the Exponential-Time Hypothesis (ETH, see e.g. [4] for a definition). Recall that b is the size of the input \mathcal{S} , assuming it is binary encoded.

Theorem 4. *Unless ETH fails, MZSP cannot be solved (i) in $2^{o(n)} \cdot b^{O(1)}$ or (ii) in $2^{o(\sqrt{b})}$.*

We now show that the above ETH bound based on n is essentially tight.

Theorem 5. *MZSP is solvable in $O^*(2^n)$.*

Proof. We solve MZSP by dynamic programming. Given an instance \mathcal{S} of MZSP, we create a dynamic programming 1-dimensional table T indexed by the subsets of \mathcal{S} . We set $T[\emptyset]$ to 0. Then, for increasing $i \in \llbracket 1, n \rrbracket$, and for every size- i subset P_i of \mathcal{S} , we fill $T[P_i]$ using the following rule:

$$T[P_i] = \begin{cases} \max_{a \in P_i} \{T[P_i \setminus \{a\}]\} & \text{if } P_i \text{ does not sum to } 0 \\ \max_{a \in P_i} \{T[P_i \setminus \{a\}]\} + 1 & \text{otherwise} \end{cases}$$

The optimal value k for MZSP is then found in $T[\mathcal{S}]$, and an optimal zero-sum partition of \mathcal{S} can be found by backtracking from that value. The space and time complexity of the above algorithm is $O^*(2^n)$, since it takes polynomial time to fill any of the 2^n elements in T . It remains to show correctness. For this, for any subset P of \mathcal{S} , we denote by k_P the cardinality of a maximum zero-sum subpartition of P , where the term *subpartition* describes a partition of a subset of P . Our goal is to show that for any P , $T[P] = k_P$. This is done by induction on $i = |P|$. When $i = 0$, this trivially holds as $T[\emptyset]$ is set to 0. Suppose now that for some $i \in \llbracket 0, n-1 \rrbracket$, any P such that $|P| = i$ satisfies $T[P] = k_P$. Let us now observe a set P of cardinality $i+1$. If P does not sum to zero, let us consider a maximum cardinality zero-sum subpartition of P , say (A_1, \dots, A_{k_P}) . Since P does not sum to zero, there exists $a \in P$ such that $a \notin \bigcup_{i=1}^{k_P} A_i$. Thus (A_1, \dots, A_{k_P}) is a maximum cardinality zero-sum subpartition of $P \setminus \{a\}$, otherwise this would contradict the cardinality maximality of (A_1, \dots, A_{k_P}) . Thus $k_{P \setminus \{a\}} = k_P$. Since $T[P \setminus \{a\}] = k_{P \setminus \{a\}}$, we have $T[P] \geq k_P$ by definition of $T[P]$. Conversely, if $a \in P$ and $(A_1, \dots, A_{k_{P \setminus \{a\}}})$ is a zero-sum subpartition of $P \setminus \{a\}$, then $(A_1, \dots, A_{k_{P \setminus \{a\}}})$ is a zero-sum subpartition of P . Thus $k_{P \setminus \{a\}} \leq k_P$, which means $T[P \setminus \{a\}] \leq k_P$ and thus implies $T[P] \leq k_P$. Altogether, we have $T[P] = k_P$. Now if P sums to zero, let (A_1, \dots, A_{k_P}) be a maximum cardinality zero-sum partition of P . We thus have $P = \bigcup_{j=1}^{k_P} A_j$. Let $a \in A_{k_P}$ (note that a exists, since A_{k_P} is non-empty). Thus (A_1, \dots, A_{k_P-1}) is a maximum zero-sum subpartition of $P \setminus \{a\}$, and hence $k_P = k_{P \setminus \{a\}} + 1$, which implies $T[P] \geq k_P$. Conversely, if $a \in P$, then $a \in A_j$ for a given $j \in \llbracket 1, k_P \rrbracket$. Using similar arguments as previously, we can prove that $k_P = k_{P \setminus \{a\}} + 1$, and thus $T[P] \leq k_P$. Altogether, we have $T[P] = k_P$. We conclude that $T[P] = k_P$ for any $P \subseteq \mathcal{S}$. In particular, $T[\mathcal{S}]$ contains a maximum cardinality zero-sum partition of \mathcal{S} . By backtracking in T , the sought partition can be found in polynomial time, which solves MZSP. \square

The previous theorem is based on the fact that the number of distinct subsets in \mathcal{S} is upper bounded by $O(2^n)$. It is also possible to upper bound this number by a function of b , the number of bits needed to binary encode \mathcal{S} .

Theorem 6. MZSP is solvable in $2^{O(\frac{b}{\log b})}$.

Proof. Let us partition \mathcal{S} into $S_p = \{a_i \in \mathcal{S} \text{ s.t. } |a_i| \leq \sqrt{b}\}$ and $S_q = \{a_i \in \mathcal{S} \text{ s.t. } |a_i| > \sqrt{b}\}$. Let us also denote, for any multiset E , by $\mathcal{P}(E)$ the set of subsets of E (e.g. $\mathcal{P}(\{2, 3, 3\}) = \{\emptyset, \{2\}, \{3\}, \{2, 3\}, \{3, 3\}, \{2, 3, 3\}\}$). In that case, we have $|\mathcal{P}(S_p)| \leq (b+1)^{2\sqrt{b}}$: indeed, by definition any element $a \in S_p$

satisfies $|a| \leq \sqrt{b}$. Moreover, a appears at most b times in S_p , since any a needs at least one bit to be encoded, while b bits are enough to encode \mathcal{S} . Thus S_p contains at most $2\sqrt{b}$ different numbers, each of them appearing at most b times in S_p . Hence $|\mathcal{P}(S_p)| \leq (b+1)^{2\sqrt{b}} \leq 2^{2\sqrt{b}\log(b+1)} \leq 2^{2\frac{b}{\log b}}$. On the other hand, S_q contains elements of size at least \sqrt{b} . Then S_q cannot be of cardinality greater than $2\frac{b}{\log b}$, otherwise encoding S_q would require more than b bits; thus $|\mathcal{P}(S_q)| \leq 2^{2\frac{b}{\log b}}$. Since S_p and S_q form a partition of \mathcal{S} , we have that $|\mathcal{P}(\mathcal{S})| = |\mathcal{P}(S_p)| \cdot |\mathcal{P}(S_q)| \leq 2^{4\frac{b}{\log b}}$. Since the dynamic programming algorithm from proof of Theorem 5 solves MZSP and has running time in $O^*(|\mathcal{P}(\mathcal{S})|)$, we conclude that MZSP can be solved in $2^{O(\frac{b}{\log b})}$, which proves the theorem. \square

We now end this section by turning our attention to the inapproximability of MZSP, in Theorem 7 below.

Theorem 7. *Unless $P=NP$, MZSP cannot be approximated within ratio $O(n^{1-\epsilon})$ for any $\epsilon > 0$.*

Proof. As for Theorem 2, we prove the result by reduction from PARTITION, which is known to be NP-complete [8]. Let $X = \{x_1, x_2, \dots, x_\ell\}$ be an instance of PARTITION, and let $\sum_{i=1}^{\ell} x_i = 2N$ with $N \geq 1$. We can indeed assume $\sum_{i=1}^{\ell} x_i$ to be non-zero and even, otherwise the problem is trivially answered. The reduction is as follows: let $q \geq 1$ be any integer, and let us recursively build a set $\{p_0, p_1, \dots, p_q\}$ of integers. More precisely, we set $p_0 = 1$, and $p_i = (2iN + 1)p_{i-1}$ for any $i \in \llbracket 1, q \rrbracket$. We note that for any $0 \leq j < i \leq q$, p_j divides p_i . Based on X and on the values p_0, p_1, \dots, p_q , we now construct the multiset $\mathcal{S} = \bigcup_{i=0}^q \{p_i X, -N p_i, -N \sum_{i=0}^q p_i\}$ where, for any $i \in \llbracket 0, q \rrbracket$, $p_i X$ denotes the values obtained by multiplying each element of X by p_i . It can be seen that \mathcal{S} sums to zero, and is thus a valid instance of MZSP. It can also be seen that the above reduction takes polynomial-time, as long as q remains polynomial in the input size of PARTITION. Intuitively, the above reduction consists in ‘‘copying’’, a certain amount of times, an (expanded) instance X of PARTITION, so that the solution size of MZSP on \mathcal{S} increases, while maintaining the property that the different ‘‘expanded copies’’ of X in \mathcal{S} do not mutually interact.

Let us now prove correctness of our reduction, by showing the following: (i) X is a YES-instance for PARTITION iff (ii) MZSP for \mathcal{S} yields a partition of cardinality $q+2$ iff (iii) MZSP for \mathcal{S} yields a partition of cardinality 2.

$((i) \Rightarrow (ii))$ Suppose X is a YES-instance for PARTITION. Then there exists $P \subset X$ such that $\sum_{a \in P} a = N$. By construction, for every $i \in \llbracket 0, q \rrbracket$, $\sum_{a \in P} p_i a = N p_i$. Hence, for every negative number $-s$ in \mathcal{S} (s being either $-N p_i$ for some $i \in \llbracket 0, q \rrbracket$, or $-N \sum_{i=0}^q p_i$), it is possible to find a subset of \mathcal{S} summing to $N p_i$, and moreover any pair of such sets is mutually disjoint. Hence \mathcal{S} can be partitioned into zero-sum subsets, and the cardinality of such a partition is $q+2$.

$((ii) \Rightarrow (iii))$ If MZSP for \mathcal{S} yields a partition of cardinality $q+2$, and since $q+2 \geq 2$ then, by merging any $q+1$ sets in this partition, we obtain a size-2 zero-sum partition.

((iii) \Rightarrow (i)) Suppose there exists a zero-sum partition of MZSP for \mathcal{S} , of cardinality 2. In that case, there exists a non-empty zero-sum (multi)set $P \subset \mathcal{S}$ that does not contain the negative integer $-N \sum_{i=0}^q p_i$. Let us denote i_0 the smallest index $i \in \llbracket 0, q \rrbracket$ such that $-Np_i$ belongs to P . Note that for all $i \in \llbracket 0, i_0 - 1 \rrbracket$, we have $p_i X \cap P = \emptyset$: indeed, suppose by contradiction that this is not the case, and let A be the sum of the elements of $\left(\bigcup_{i=0}^{i_0-1} p_i X\right) \cap P$. Then we have $A \leq (1 + p_1 + \dots + p_{i_0-1}) \cdot 2N$, hence $A \leq i_0 \cdot 2N \cdot p_{i_0-1}$, which yields $A < p_{i_0}$. In particular, p_{i_0} does not divide A , since $A \neq 0$. As p_{i_0} divides every other element of P , we conclude that $A = 0$, which is the sought contradiction.

Now, let us consider $P' = P \bmod p_{i_0+1}$. From the above, the only elements from P that induce non zero elements in P' are the elements of $p_{i_0} X \cap P$, together with $-Np_{i_0}$. We thus conclude there exists a (multi)set $K \subset X$ such that $\sum_{a \in K} p_{i_0} a \equiv Np_{i_0} \pmod{p_{i_0+1}}$. Since $p_{i_0+1} > 2Np_{i_0}$, $\sum_{a \in K} p_{i_0} a = Np_{i_0}$, and thus $\sum_{a \in K} a = N$. In other words, we have a YES-instance for PARTITION.

Now we have proved correctness of our reduction, let us turn to proving our inapproximability result. Let ϵ be any strictly positive value, and suppose that there exists an approximation algorithm \mathcal{A} for MZSP, of ratio $\rho = O(n^{1-\epsilon})$ with $n = |\mathcal{S}|$. Take now an instance X of PARTITION, and recall that $\ell = |X|$. Let C be a constant such that $\rho \leq Cn^{1-\epsilon}$ for sufficiently large n . We let $q = \max(C^{\frac{1}{1-\epsilon}} - 1, (C^{\frac{1}{1-\epsilon}}(\ell+1)+1)^{\frac{1}{\epsilon}-1} - 1)$, and we proceed with the above mentioned reduction by building the MZSP instance \mathcal{S} based on X and on parameter q . We have that $n = |\mathcal{S}| = (\ell + 1)(q + 1) + 1$.

Then, $\frac{q+1}{C^{\frac{1}{1-\epsilon}}} \geq 1$ and $(q+1)^{\frac{1}{1-\epsilon}-1} - C^{\frac{1}{1-\epsilon}}(\ell+1) \geq 1$ which yields $\frac{q+1}{C^{\frac{1}{1-\epsilon}}} \cdot ((q+1)^{\frac{1}{1-\epsilon}-1} - C^{\frac{1}{1-\epsilon}}(\ell+1)) \geq 1$ and $C((\ell+1)(q+1)+1)^{1-\epsilon} \leq q+1$. We thus conclude that $Cn^{1-\epsilon} \leq q+1$. We now apply \mathcal{A} on \mathcal{S} , and solve it polynomially within factor $\rho \leq Cn^{1-\epsilon}$. Thus we obtain $\rho \leq q+1$, hence $\rho < q+2$. However, we know from the above that if we have a YES-instance for PARTITION, then there exists a zero-sum partition of \mathcal{S} of cardinality $q+2$. In that case, the solution provided by the approximation algorithm \mathcal{A} is a zero-sum partition of \mathcal{S} of cardinality $c \geq \frac{q+2}{\rho} > 1$. Conversely, if \mathcal{A} provides a zero-sum partition of cardinality $c > 1$, then such zero-sum partition shows that X is a YES-instance for PARTITION.

Altogether, if there exists an approximation algorithm \mathcal{A} for MZSP of ratio $\rho = O(n^{1-\epsilon})$, it is possible to polynomially solve PARTITION: a contradiction, unless P=NP, to the fact that PARTITION is NP-complete [8]. \square

3 Parameterized Complexity of MZSP

Parameters k and m . We first consider fixed-parameterized complexity of MZSP with respect to the size k of the solution. On the way, we will also discuss parameter $m = \min\{\text{neg}, \text{pos}\}$, as we always have $m \geq k$. By Theorem 2, we know that, unless P=NP, MZSP is not FPT with respect to parameter k (resp. m), since MZSP is NP-complete even in the case where both these values are constant. The following theorem and corollary show W[1]-hardness of UNARY MZSP with respect to the same parameters.

Theorem 8. UNARY MZSP parameterized by m is $W[1]$ -hard.

Theorem 8 implies the following corollary, as we always have $m \geq k$.

Corollary 1. UNARY MZSP parameterized by k is $W[1]$ -hard.

Proof. (of Theorem 8) Let us assume that integers are encoded in unary. We reduce from UNARY BIN-PACKING, which is known to be $W[1]$ -hard with respect to parameter “size of the solution” [6], to UNARY MZSP. We first recall the definition of BIN-PACKING (presented here in its decision version):

BIN-PACKING

Instance : a multiset of strictly positive integers $P = \{w_1, \dots, w_n\}$, an integer W , an integer t .

Question : Does there exist a partition $\{J_1, \dots, J_t\}$ of P such that $\sum_{w_j \in J_i} w_j \leq W$ for every $i \in \llbracket 1, t \rrbracket$?

As mentioned above, UNARY BIN-PACKING is BIN-PACKING in which all integers are assumed to be encoded in unary; besides, UNARY BIN-PACKING, parameterized by the number t of bins, is known to be $W[1]$ -hard [6]. Let $I = (P, W, t)$ be an instance of UNARY BIN-PACKING. Moreover, assume $\sum_{i=1}^n w_i = tW$, since UNARY BIN-PACKING remains $W[1]$ -hard parameterized by the number t of bins under this condition [6]. Let us now construct the following instance of MZSP: $\mathcal{S} = \{w_1, \dots, w_n, \underbrace{-W, \dots, -W}_t\}$. Note that we have $m = t$. We now show that

MZSP admits a t -size zero-sum partition iff $I = (P, W, t)$ is a YES-instance for UNARY BIN-PACKING.

(\Rightarrow) Suppose (P, W, t) is a YES-instance for UNARY BIN-PACKING. Thus there exists a partition $\{J_1, \dots, J_t\}$ of P such that $\sum_{w_j \in J_i} w_j \leq W$ for every $i \in \llbracket 1, t \rrbracket$. However, since we assume $\sum_{i=1}^n w_i = tW$, we conclude that every J_i , $i \in \llbracket 1, t \rrbracket$, is such that $\sum_{w_j \in J_i} w_j = W$. Hence, $\{A_1, A_2, \dots, A_t\}$, where $A_i = J_i \cup \{-W\}$ for every $i \in \llbracket 1, t \rrbracket$, is a t -size zero-sum partition of \mathcal{S} .

(\Leftarrow) Conversely, suppose there exists a t -size zero-sum partition $\{A_1, A_2, \dots, A_t\}$ of \mathcal{S} . Since \mathcal{S} contains exactly t negative numbers, this implies that every A_i , $i \in \llbracket 1, t \rrbracket$, contains exactly one occurrence of $-W$. Thus $\{J_1, J_2, \dots, J_t\}$, where $J_i = A_i \setminus \{-W\}$ for every $i \in \llbracket 1, t \rrbracket$, is a partition of P . Moreover, since each A_i sums to zero, we know that each J_i sums to W , which ensures that (P, W, t) is a YES-instance for UNARY BIN-PACKING.

The above reduction is a valid parameterized reduction, since parameter t for UNARY BIN-PACKING is strictly equal to parameter m for UNARY MZSP. Moreover, the instance \mathcal{S} of UNARY MZSP that we built satisfies $k = m$. \square

Theorem 8 suggests that, even encoded in unary, MZSP admits no FPT algorithm parameterized by k . What we can show, in the following theorem, is that MZSP encoded in unary is in XP when parameterized by k .

Theorem 9. UNARY MZSP is in XP parameterized by k .

Since we always have $m \geq k$, Theorem 9 implies the following corollary.

Corollary 2. UNARY MZSP is in XP when parameterized by m .

Parameter number of distinct values n^ .* We now discuss parameter n^* , for which we first provide a complexity lower bound based on ETH. Recall that b is the (binary encoded) size of the input instance \mathcal{S} .

Theorem 10. *Unless ETH fails, MZSP cannot be solved in $2^{o(n^*)}b^{O(1)}$.*

Proof. The proof is based on a combination of (i) a reduction from 3-SAT to SUBSET-SUM presented in [7] and inspired from [10] and (ii) the reduction from SUBSET-SUM to MZSP from proof of Theorem 4. More precisely, starting from any instance of 3-SAT with n variables and m clauses, an instance of SUBSET-SUM containing $n' = 2n + 2m$ integers, among which $2n + m$ are pairwise distinct, is constructed. Moreover, in proof of Theorem 4, the instance \mathcal{S} of MZSP built from SUBSET-SUM contains $n'' = n' + 2 = 2n + 2m + 2$ integers, among which $n^* = 2n + m + 2$ are pairwise distinct. Under ETH, 3-SAT cannot be solved in $2^{o(n)}$. If we combine this information with the sparsification method [5] (which allows to consider only 3-SAT instances for which $m = O(n)$) and the above argument, we conclude that, under ETH, MZSP cannot be solved in $2^{o(n^*)}b^{O(1)}$. \square

Concerning parameterized complexity with respect to n^* , we suspect MZSP to be W[1]-hard parameterized by n^* , but the question remains open. Meanwhile, we are able to prove (see Theorem 11) that the problem is in XP with respect to n^* . The rationale for this result is that the multisets that constitute any maximum zero-sum partition of \mathcal{S} are few, and that we can efficiently compute them. In order to prove Theorem 11, we need to introduce several definitions, and first prove two propositions (Propositions 1 and 2).

Let us suppose that \mathcal{S} is a multiset containing n^* distinct values, denoted a_1, \dots, a_{n^*} . We introduce several notions: given any multiset \mathcal{M} built from a_1, \dots, a_{n^*} , we call *multiplicity multiset* of \mathcal{M} the multiset $\{u_1, u_2, \dots, u_{n^*}\}$ representing the multiplicities of each a_i in \mathcal{M} : more precisely, for any $i \in \llbracket 1, n^* \rrbracket$, $u_i \in \mathbb{N}$ is the number of times a_i appears in \mathcal{M} . With this notation, we can define a partial order \leq on multiplicity multisets as follows: let $u = \{u_1, u_2, \dots, u_{n^*}\}$ and $v = \{v_1, v_2, \dots, v_{n^*}\}$ be two multiplicity multisets; we write $u \leq v$ whenever $u_i \leq v_i$ for every $i \in \llbracket 1, n^* \rrbracket$. Now let $s = \{s_1, s_2, \dots, s_{n^*}\}$ be the multiplicity multiset of \mathcal{S} . We first define two sets, respectively named K and D : K is the set of irreducible multiplicity multisets of \mathcal{S} leading to zero-sum subsets of \mathcal{S} , and D is the set of all zero-sum subsets of \mathcal{S} . In the following, for simplicity we write $u = 0$ for any vector u whenever all its coordinates are equal to 0. Formally, K and D are defined as follows:

$$K = \left\{ u \in \mathbb{N}^{n^*} \mid \sum_{i=1}^{n^*} u_i a_i = 0, u \neq 0 \text{ and} \right. \\ \left. \forall v \in \mathbb{N}^{n^*}, v \leq u \text{ and } \sum_{i=1}^{n^*} v_i a_i = 0 \Rightarrow v = 0 \text{ or } v = u \right\} \text{ and}$$

$$D = \left\{ u \in \mathbb{N}^{n^*} \mid u \leq s, u \neq 0 \text{ and } \sum_{i=1}^{n^*} u_i a_i = 0 \right\}.$$

Any maximum zero-sum partition of \mathcal{S} is induced by elements of $K \cap D$ only. We define a third set Z as follows: $Z = \left\{ u \in \mathbb{N}^{n^*} \mid u \neq 0 \text{ and } \sum_{i=1}^{n^*} u_i a_i = 0 \right\}$. Note that both K and D are included in Z . We are now interested in two properties, related to the above definitions. We begin with Proposition 1.

Algorithm 1 XP algorithm for solving MZSP, parameterized by n^*

```

1: Compute  $D$ 
2: Initialize  $T$ 
3: for every  $i \in \llbracket 1, n \rrbracket$  do
4:   for every  $u \in D$  do
5:     for every  $v \in D$  do
6:       if  $v + u \in D$  then
7:          $T[v + u] = \max(T[v] + 1, T[v + u])$ 
8:       end if
9:     end for
10:  end for
11: end for
12: return  $T[s]$ 

```

Proposition 1. *Let U be a subset of Z such that $K \cap D \subseteq U$. There exists a size- p zero-sum partition of \mathcal{S} iff there exists $\ell \geq p$ elements of U , say u^1, u^2, \dots, u^ℓ such that $\sum_{i=1}^{\ell} u^i = s$.*

For solving p -MZSP, it thus suffices to compute $K \cap D$, and to test whether it is possible to reach s , using p or more elements of $K \cap D$. Before that, we discuss the maximum cardinalities of K and D (see Proposition 2), which will be useful to evaluate the time complexity to generate these sets.

Proposition 2. $D \subseteq \llbracket 0, n \rrbracket^{n^*}$ and $K \subseteq \llbracket 0, n^*B - 1 \rrbracket^{n^*}$.

In order to compute D (resp. K), it thus suffices to generate each element of $\llbracket 0, n \rrbracket^{n^*}$ (resp. $\llbracket 0, n^*B - 1 \rrbracket^{n^*}$), and check for each of them whether it belongs to D (resp. K). For each element of $\llbracket 0, n \rrbracket^{n^*}$, checking its membership to D can be achieved in $O(n^2 \log B)$, thus D can be computed in $O(n^{n^*+2} \log(B))$. Concerning K , testing if an element of $\llbracket 0, n^*B - 1 \rrbracket^{n^*}$ sums to zero can be done in $O(n^* \log(n^*B))$, and by dynamic programming, we can check if it is irreducible in $O((n^*B)^{n^*})$; thus K can be computed in $O((n^*B)^{2n^*})$.

The set K (and its computation) will be useful later for proving Theorem 13. In the following, we first focus on set D , whose cardinality is denoted c_D . Indeed, starting from D , Algorithm 1 shows that MZSP can be solved in $O(nc_D^2 n^* \log n + n^{n^*+2} \log B)$. Since, by Proposition 2 above, c_D is in $O(n^{n^*})$, this shows that MZSP is XP relatively to parameter n^* , as stated by the following theorem.

Theorem 11. *MZSP is in XP when parameterized by n^* .*

Proof. We provide an algorithm that runs in time $O(nc_D^2 n^* \log n + n^{n^*+2} \log B)$. The proof derives from Algorithm 1, in which T is an array indexed by the elements of D and which is initialized as follows: $T[0] = 0$, and for every other vector $v \in D$, $T[v] = -\infty$. We also recall that s is the multiplicity multiset of \mathcal{S} .

Clearly, Algorithm 1 runs in $O(nc_D^2 n^* \log n)$ where $c_D = |D|$, since nc_D^2 additions on vectors are realized, each taking $O(n^* \log n)$ time. To this complexity, $O(n^{n^*+2} \log B)$ should be added for the precomputation of D . We now show

that Algorithm 1 is correct. At the end of the algorithm, for any $v \in D$, $T[v]$ represents the largest number of elements of D that we can sum to obtain v . Let us denote w_v this value (thus $T[v] = w_v$). For any $i \in \llbracket 1, n \rrbracket$, let $\mathcal{P}(i)$ be the following property: for every $v \in D$, $T[v] = w_v$ if $w_v \in \llbracket 0, i \rrbracket$. Our goal is to prove, by induction on i , that $\mathcal{P}(i)$ holds for any $i \in \llbracket 1, n \rrbracket$. First, $\mathcal{P}(0)$ is true since $w_v = 0$ implies $v = 0$. Now let $i \geq 0$, and let us assume $\mathcal{P}(i)$ holds. Let $v \in D$. If $w_v \leq i$, then by induction hypothesis, we have $T[v] = w_v$. If not, then there exists $u \in D$ such that $w_{v-u} = i$, and hence $T[v-u] = i$. Hence, by construction of T , $T[v] = i + 1 = w_v$, which consequently proves that $\mathcal{P}(i+1)$ holds. By induction, $\mathcal{P}(i)$ holds for any $i \in \llbracket 1, n \rrbracket$. In particular, $T[s]$ represents the largest number of elements from D that can be summed in order to obtain s . By Proposition 1, there exists ℓ elements of D whose sum is s iff there exists a zero-sum partition of \mathcal{S} , of cardinality ℓ . Thus the cardinality of a maximum zero-sum partition of \mathcal{S} is $T[s]$, which shows correctness of Algorithm 1. \square

As mentioned before, we conjecture MZSP to be W[1]-hard parameterized by n^* . In contrast, we have the following result.

Theorem 12. *MZSP is FPT when parameterized by $n^* + k$.*

Parameter maximum absolute value B . Recall that B is the greatest integer (in absolute value) in an instance \mathcal{S} of MZSP.

Theorem 13. *MZSP is FPT when parameterized by B .*

Proof. In order to prove the result, we will provide an ILP model for our problem. We will then show that the number of variables of our ILP is a function of n^* and B only, which, combined with the fact that ILP is FPT parameterized by its number of variables [2, 9] and the fact that $n^* \leq 2B$, allows us to conclude. Given an integer k , we are interested in solving k -MZSP, which asks whether a size- k zero-sum partition of \mathcal{S} exists. Let (\mathcal{S}, k) be an instance of k -MZSP. Let us number the n^* distinct values in \mathcal{S} a_1, \dots, a_{n^*} and let s_1, \dots, s_{n^*} be their respective multiplicities in \mathcal{S} . Let $c_K = |K|$, where K is the set defined previously, and let us compute K – recall that c_K , by Proposition 2 and the discussion that follows, satisfies $c_K = O((n^*B)^{n^*})$, and that K can be computed in $O((n^*B)^{2n^*})$. Our ILP based on the following c_K variables x_u , $u \in K$, where x_u represents the number of times element u appears in a zero-sum partition of \mathcal{S} of cardinality at least k . The ILP formulation of the problem is as follows.

<p>k-MZSP (ILP model):</p> <div style="text-align: center; margin-top: 10px;"> <p>c.1 $\forall u \in K \quad x_u \geq 0$</p> <p>c.2 $\forall i \in \llbracket 1, n^* \rrbracket \quad \sum_{u \in K} u_i x_u = s_i$</p> <p>c.3 $\sum_{u \in K} x_u \geq k$</p> </div>
--

We now show correctness of our ILP model, by proving that there exists a zero-sum partition of MZSP of cardinality k iff the above ILP formulation admits a solution.

(\Rightarrow) Suppose \mathcal{S} admits a size- k zero-sum partition. Then, by Proposition 1, we know there exist $\ell \leq k$ elements of K which sum to s , that we will call u^1, \dots, u^ℓ . For $u \in K$, let x_u denote the number of times u appears in (u_1, \dots, u_ℓ) . Then, by definition, $\forall u \in K, x_u \geq 0$, $\sum_{u \in K} x_u u = s$ and $\sum_{u \in K} x_u = \ell \geq k$. Thus our ILP formulation admits a solution.

(\Leftarrow) Conversely, suppose there exists x_u for $u \in K$, which is a solution to the above ILP formulation. Let us build $(u^i)_{i \in [1, \ell]}$, where element u appears exactly x_u times. Then, from C.2, $\sum_{i=1}^{\ell} u^i = s$. Moreover, from C.3, $\ell \geq k$. Thus, from Proposition 1, there exists a size- k zero-sum partition of \mathcal{S} .

Since $n^* \leq 2B$, and since ILP, parameterized by the number x of variables, is FPT and can be solved in $O^*(x^{2.5x+o(x)})$ [2, 9], and since here $x = c_K = O((n^*B)^{n^*})$, the result follows. \square

4 Conclusion

We provided diverse algorithmic results concerning the MZSP problem: hardness, (in)approximability and fixed-parameterized complexity considerations with respect to parameters n, m, k, n^* and B . Some questions about MZSP remain unanswered. In particular, we conjecture MZSP to be W[1]-hard parameterized by the number n^* of distinct values in \mathcal{S} ; (dis)proving it remains open.

Acknowledgments. We thank Ton van der Zanden for his suggestions, in particular those concerning parameter b .

References

1. Fertin, G., Jean, G., Tannier, E.: Algorithms for computing the double cut and join distance on both gene order and intergenic sizes. *Algorithms for Molecular Biology* **12**(1), 16:1–16:11 (2017)
2. Frank, A., Tardos, É.: An application of simultaneous diophantine approximation in combinatorial optimization. *Comb.* **7**(1), 49–65 (1987)
3. Garey, M.R., Johnson, D.S.: *Computers and intractability : a guide to the theory of NP-completeness*. Bell Telephone Laboratories (1979)
4. Impagliazzo, R., Paturi, R.: On the complexity of k -SAT. *Journal of Computer and System Sciences* **62**(2), 367–375 (2001)
5. Impagliazzo, R., Paturi, R., Zane, F.: Which problems have strongly exponential complexity? *Journal of Computer and System Sciences* **63**(4), 512–530 (2001)
6. Jansen, K., Kratsch, S., Marx, D., Schlotter, I.: Bin packing with fixed number of bins revisited. *Journal of Computer and System Sciences* **79**(1), 39–49 (2013)
7. Jansen, K., Land, F., Land, K.: Bounding the running time of algorithms for scheduling and packing problems. In: Dehne, F., Solis-Oba, R., Sack, J.R. (eds.) *Algorithms and Data Structures*. pp. 439–450. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
8. Karp, R.M.: *Reducibility among Combinatorial Problems*, pp. 85–103. Springer US, Boston, MA (1972)
9. Lenstra, H.W.: Integer programming with a fixed number of variables. *Mathematics of Operations Research* **8**(4), 538–548 (1983)
10. Wegener, I.: *Complexity Theory*. Springer (2005)