

On Finding Small 2-Generating Sets

Isabelle Fagnot¹, Guillaume Fertin², and Stéphane Vialette³

¹ IGM-LabInfo, CNRS UMR 8049, Université Paris-Est,
5 Bd Descartes 77454 Marne-la-Vallée, France
and Université Paris Diderot-Paris 7, France
`fagnot@univ-mlv.fr`

² Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR CNRS 6241
Université de Nantes, 2 rue de la Houssinière, 44322 Nantes Cedex 3 - France
`fertin@lina.univ-nantes.fr`

³ IGM-LabInfo, CNRS UMR 8049, Université Paris-Est,
5 Bd Descartes 77454 Marne-la-Vallée, France
`viallette@univ-mlv.fr`

Abstract. Given a set of positive integers S , we consider the problem of finding a minimum cardinality set of positive integers X (called a *minimum 2-generating set of S*) such that every element of S is an element of X or is the sum of two (non-necessarily distinct) elements of X . We give some elementary properties of 2-generating sets and prove that finding a minimum cardinality 2-generating set is hard to approximate within ratio $1 + \varepsilon$ for any $\varepsilon > 0$. We then prove our main result, which consists in a standard representation lemma for minimum cardinality 2-generating sets.

Keywords: Algorithmic number theory, Set covering, Parameterized complexity.

1 Introduction

In this paper, we consider the problem of 2-generating a set of positive integers S with a minimum cardinality set of integers X , where X is said to *2-generate* S if every element of S is an element of X or is the sum of two (non-necessarily distinct) elements of X . We note that, in this context, X does not have to be a subset of S . We refer to this problem as MINIMUM 2-GENERATING SET.

MINIMUM 2-GENERATING SET is a simple restriction of MINIMUM GENERATING SET (a natural problem in number theory) [4]. The MINIMUM GENERATING SET problem is defined as follows: Given a set of positive integers S , find a minimum cardinality set of integers X such that every element of S is the sum of a subset of X . MINIMUM GENERATING SET has been shown to be **NP**-hard [4], and is related, among other things, to planning radiation therapy: elements of S represent radiation dosages required at various points, while an element of X represents a dose delivered simultaneously to multiple points. Other variants, namely the cases in which the elements of X can be negative or fractional, are considered in [5,14].

Strongly related to our work are minimum sum covers of finite Abelian groups as investigated in [10,8]. A subset X of a finite Abelian group G is said to be a *sum cover* of G if $\{x + x' : x, x' \in X\} = G$, a *strict sum cover* of G if $\{x + x' : x, x' \in X \wedge x \neq x'\} = G$, and a *difference cover* of G if $\{x - x' : x, x' \in X\} = G$. Swanson [19] gives some constructions and computational results for maximum difference packings of cyclic groups. Haanpää, Huima, and Östergård compute maximum sum and strict sum packings of cyclic groups [11]. Fitch and Jamison [8] give minimum sum and strict sum covers of small cyclic groups, and Wiedemann [20] determines minimum difference covers for cyclic groups of order at most 133.

Another area of research related to our work (and this was in fact our initial motivation for addressing the present problem) is the problem of covering a set of strings S with a set X of substrings in S , where X is said to *cover* S if every string in S can be written as a concatenation of the substrings in X [13,2] (see also [15] and [3] for a more general treatment of

the combinatorial rank). Covering a set of strings S with a set X of substrings in S is indeed the MINIMUM GENERATING SET problem for unary alphabet under the unary encoding scheme. To narrow the context, notice that, given a set of binary strings S , finding a minimum cardinality set X of substrings in S such that every string in S can be written as a concatenation of *at most* two substrings in X is **NP**-complete (the proof being an easy binary alphabet encoding of the result of Néraud [15]). Finally, Hajiaghayi *et al.* [12] considered the MINIMUM MULTICOLORED SUBGRAPH problem, which can be seen as a generalization of MINIMUM 2-GENERATING SET when every integer in the input set is bounded by a polynomial in the length of the input.

This paper is organized as follows: we first recall basic definitions in Section 2, and we then formally introduce the considered problem. In Section 3, we give some elementary properties of 2-generating sets. Section 4 is devoted to prove hardness of MINIMUM 2-GENERATING SET and we prove in Section 5 a standard representation lemma.

2 Preliminaries

We use \mathbb{N}^* to refer to the set of all natural numbers excluding zero, *i.e.*, $\mathbb{N}^* = \{1, 2, \dots\}$. Let $S = \{s_1, s_2, \dots, s_n\} \subset \mathbb{N}^*$. For any $k \in \mathbb{N}^*$, we write kS for the set of all integers that can be expressed as the sum of *exactly* k *non necessarily distinct* integers of S *i.e.*, $kS = \{s_{i_1} + s_{i_2} + \dots + s_{i_k} : s_{i_1}, s_{i_2}, \dots, s_{i_k} \in S\}$. According to this definition, for any set S , $S = 1S$. A set $X \subset \mathbb{N}^*$ is a *k-generating set* of S (or *k-generates* S) if $S \subseteq \bigcup_{i=1}^k iX$. (Notice here that we do not require the additional constraint $\bigcup_{i=1}^k iX \subseteq S$.) It is called a *minimum k-generating set* of S if X is a *k-generating set* of S of minimum cardinality. The *k-rank* of S , denoted $\text{rk}_k(S)$, is the cardinality of a minimum *k-generating set* of S . A set $S \subset \mathbb{N}^*$ is *k-elementary* if $\text{rk}_k(S) = |S|$. Let $\min(S)$ and $\max(S)$ stand for $\min\{s_i : s_i \in S\}$ and $\max\{s_i : s_i \in S\}$, respectively. The *length* of S , denoted $\text{len}(S)$, is defined to be $\text{len}(S) = \max(S) - \min(S)$.

MINIMUM k -GENERATING SET

- **Input** : A set $S \subset \mathbb{N}^*$.
- **Solution** : A k -generating set X of S .
- **Measure** : The cardinality of X .

Our main interest here is in finding minimum 2-generating sets, and hence we shall be concerned in this paper with MINIMUM 2-GENERATING SET only.

Of particular importance, we assume hereafter any reasonable (e.g. binary) encoding of any instance of MINIMUM 2-GENERATING SET so that the input is in $O(n \log m)$ space, where $n = |S|$ and $m = \max(S)$.

We assume readers have basic knowledge about graph theory [6] and we shall only recall basic notations. We write $G = (V, E)$ for a graph with *vertex set* V and *edge set* E . For a graph G and a vertex $u \in V$, we write $d_G(u)$ for the *degree* of u in G . A graph is *bipartite* if it does not contain an odd cycle. For a graph G , we write CC_G for the set of all *connected components* of G . If u is a vertex of G , we denote by $\text{CC}_G(u)$ the connected component of G u is part of, and, by abuse of notation, we denote the number of vertices in $\text{CC}_G(u)$ by $|\text{CC}_G(u)|$.

3 Elementary properties

3.1 Generalities

To fix the context, we begin by giving easy bounds for $\text{rk}_2(S)$.

Lemma 1. *For any $S \subset \mathbb{N}^*$ of cardinality n , $\lceil \frac{1}{2}(\sqrt{8n+9} - 3) \rceil \leq \text{rk}_2(S) \leq n$.*

Proof. The upper bound is trivial. To prove the lower bound, let $X \subset \mathbb{N}^*$ be a 2-generating set of S , and let k stand for $|X|$. For one, $|X \cup 2X| \leq \binom{k}{2} + 2k$. For another, $|X \cup 2X| \geq n$ since X 2-generates S . Combining the two inequalities yields the claimed lemma. \square

Combinatorial properties of intervals [9] will prove to be a simple but powerful tool for 2-generating sets. We write $[i : i+k]$ for the set of consecutive integers (*i.e.*, interval) $\{i, i+1, \dots, i+k\}$. For any interval system \mathcal{I} , the *matching number* of \mathcal{I} , denoted $\nu(\mathcal{I})$, is the maximum number of pairwise disjoint intervals of \mathcal{I} . Let $S = \{s_i : 1 \leq i \leq n\} \subset \mathbb{N}^*$. Define the 2-generating interval system of S , in symbols $\mathcal{I}_2(S)$, to be $\mathcal{I}_2(S) = \{[s_i/2] : s_i : s_i \in S\}$.

Lemma 2. *Let $S \subset \mathbb{N}^*$ and $X \subset \mathbb{N}^*$ be a 2-generating set of S . Then, for every $s \in S$, $|X \cap [s/2] : s| \neq \emptyset$.*

Proof. Suppose the lemma is false. Then some $s \in S$ is obtained by summing at most 2 integers of X , each upper-bounded by $\lceil s/2 \rceil - 1$. But $2(\lceil s/2 \rceil - 1) < 2((s_i/2 + 1) - 1) = s$ which yields the desired contradiction. \square

Corollary 1. *For any $S \subset \mathbb{N}^*$, $\nu(\mathcal{I}_2(S)) \leq \text{rk}_2(S)$.*

It follows from Lemma 1 that if $\nu(\mathcal{I}_2(S)) = |S|$ then S is 2-elementary. The converse is false as shown by $S = \{7, 8, 9\}$. The following application of Corollary 1 will prove useful in the sequel.

Lemma 3. *Let $A = \{a_i : 1 \leq i \leq n\} \subset \mathbb{N}^*$ be such that (i) $a_1 \geq 4$ and (ii) $a_{i+1} > 4a_i - 3$, $1 \leq i \leq n-1$. Then, the set $S = \{2a_i - 1 : 1 \leq i \leq n\} \cup \{4a_i - 3 : 1 \leq i \leq n\} \subset \mathbb{N}^*$ is 2-elementary.*

Proof. The lemma reduces to proving $\text{rk}_2(S) = |S| = 2n$. Let $\mathcal{I}_2(S)$ be the 2-generating interval system of S . We observe that all intervals of $\mathcal{I}_2(S)$ are disjoint except pairs of intervals $[a_i : 2a_i - 1]$ and $[2a_i - 1 : 4a_i - 3]$, $1 \leq i \leq n$. Then it follows that $\nu(\mathcal{I}_2(S)) = n$, and hence, according to Corollary 1, $\text{rk}_2(S) \geq n$. We claim that in fact $\text{rk}_2(S) = 2n$. Indeed, suppose, aiming at a contradiction, that $n \leq \text{rk}_2(S) < 2n$, and let X be a minimum cardinality 2-generating set of S . Then it follows that for some $1 \leq i \leq n$, $X \cap [a_i : 2a_i - 1] = X \cap [2a_i - 1 : 4a_i - 3] = \{2a_i - 1\}$. Therefore, $(4a_i - 3) \notin X$, and hence there exist $x, y \in X$ such that $x + y = 4a_i - 3$. Observe that, since $4a_i - 3$ is odd, we must have $x \neq y$. Furthermore, it must hold that $\max\{x, y\} \geq \lceil 2a_i - 3/2 \rceil = 2a_i - 1$. Combining this with $X \cap [2a_i - 1 : 4a_i - 3]$ yields $\max\{x, y\} = 2a_i - 1$, and hence $\min\{x, y\} = (4a_i - 3) - (2a_i - 1) = 2a_i - 2$. This is the desired contradiction since $(2a_i - 2) \in [a_i : 2a_i - 1]$ and $X \cap [a_i : 2a_i - 1] = \{2a_i - 1\}$. \square

3.2 Integer arithmetic sequences

An *integer arithmetic sequence* is a sequence of integers such that the difference of any two successive members of the sequence is a constant.

Lemma 4. *Let $S \subset \mathbb{N}^*$ be an integer arithmetic sequence of length n . Then $\text{rk}_2(S) = \Theta(\sqrt{n})$.*

Proof. Write $S = \{s_0 + ic : 0 \leq i \leq n-1\}$ for some $s_0 \in \mathbb{N}^*$ and $c \in \mathbb{N}^*$. Define $X = X_1 \cup X_2$, where $X_1 = \{s_0 + ic \lceil \sqrt{n} \rceil : 0 \leq i \leq \lceil \sqrt{n} \rceil - 1\}$, and $X_2 = \{ic : 1 \leq i \leq \lceil \sqrt{n} \rceil - 1\}$. An easy check shows that $S \subseteq X \cup 2X$, and hence X is a 2-generating set of S . Clearly, $|X_1| = \lceil \sqrt{n} \rceil$ and $|X_2| = \lceil \sqrt{n} \rceil - 1$. Therefore, $|X| = 2\lceil \sqrt{n} \rceil - 1 \leq 2(\sqrt{n} + 1) - 1 = 2\sqrt{n} + 1$. Combining this with Lemma 1 yields the claimed result. \square

In case S is an arithmetic sequence of length $n = k^2$, the above lemma reduces to $\text{rk}_2(S) \leq 2\sqrt{n} - 1$. We note in passing that this is a strict upper-bound for arithmetic sequences of length n . Indeed, for $S = \{1, 2, \dots, 9\}$, we have $2\sqrt{n} - 1 = 5$ whereas $X = \{1, 3, 4, 6\}$ is a 2-generating set of S of cardinality 4.

We finally observe that Lemma 4 could be an issue for dealing with dense sets. Define a set $S \subset \mathbb{N}^*$ to be ε -dense if $|S| = \varepsilon \text{len}(S)$ for some $\varepsilon > 0$. The following result is an immediate consequence of Lemma 4 (the easy proof can be turned into an approximation algorithm with performance ratio $O(\sqrt{\varepsilon})$ for ε -dense sets).

Corollary 2. *Let $S \subset \mathbb{N}^*$ be an ε -dense set of cardinality n . Then $\text{rk}_2(S) = O(\sqrt{n/\varepsilon})$.*

3.3 Integer geometric sequences

An *integer geometric sequence* is a sequence of numbers where each term after the first is found by multiplying the previous one by a fixed integer $r \geq 2$ called the *common ratio*. Results turn out to be more precise compared to arithmetic sequences.

Lemma 5. *Let $S \subset \mathbb{N}^*$ be an integer geometric sequence of length n with common ratio $r \geq 2$. Then, (i) $\text{rk}_2(S) = \lceil n/2 \rceil$ if $r = 2$ and (ii) $\text{rk}_2(S) = n$ if $r > 2$.*

Proof. A straightforward application of Corollary 1 proves (ii). To prove (i), write $S = \{s_i : 1 \leq i \leq n\}$ and $S_{\text{odd}} = \{s_i : s_i \in S \wedge i \equiv 1 \pmod{2}\}$. For one, $X = S_{\text{odd}}$ is a 2-generating set of S , and hence $\text{rk}_2(S) \leq |S_{\text{odd}}| = \lceil n/2 \rceil$. For another, $\nu(\mathcal{I}_2(S)) \geq |S_{\text{odd}}|$ since $S_{\text{odd}} \subseteq S$ and $\nu(\mathcal{I}_2(S_{\text{odd}})) = |S_{\text{odd}}|$ (the latter point follows from the fact that $s_{i+2}/2 = 2s_i > s_i$ for $1 \leq i \leq n-2$). Combining this with Corollary 1 yields $\text{rk}_2(S) \geq |S_{\text{odd}}| = \lceil n/2 \rceil$. \square

3.4 Expansion and contraction

Let $S \subset \mathbb{N}^*$. For any $c \in \mathbb{N}^*$, we write $S \times c$ for the set $\{s_i c : s_i \in S\}$ and we refer to $S \times c$ as the *c-expansion* of S . Similarly, for any $c \in \mathbb{N}^*$ common divisor of S , we write S/c for the set $\{s_i/c : s_i \in S\}$ and we refer to S/c as the *c-contraction* of S .

Lemma 6 (c-expansion). *Let $S \subset \mathbb{N}^*$ and $c \in \mathbb{N}^*$. Then $\text{rk}_2(S \times c) \leq \text{rk}_2(S)$.*

Proof. It is enough to notice that for any 2-generating set X of S , $X \times c$ is a 2-generating set of $S \times c$. \square

Replacing S by S/c in Lemma 6 yields a formulation well-suited for contraction considerations.

Corollary 3. *Let $S \subset \mathbb{N}^*$ and $c \in \mathbb{N}^*$ be a common divisor of S . Then $\text{rk}_2(S) \leq \text{rk}_2(S/c)$.*

Lemma 7 (c-contraction). *Let $S \subset \mathbb{N}^*$ and $c \in \mathbb{N}^*$ be a common divisor of S . Then, $\text{rk}_2(S/c) = \text{rk}_2(S)$ if c is odd and $\text{rk}_2(S) \leq \text{rk}_2(S/c) \leq 2 \text{rk}_2(S)$ if c is even.*

Proof. According to Corollary 3, it is enough to prove that $\text{rk}_2(S) \geq \text{rk}_2(S/c)$ if c is odd and $\text{rk}_2(S/c) \leq 2 \text{rk}_2(S)$ if c is even. Let X be a 2-generating set of S . For each $0 \leq a < c$, define

$$X_a = \{x_i \in X : x_i \equiv a \pmod{c}\}$$

For each $0 \leq a < c$, $a \in \mathbb{N}^*$, write $X + a$ (resp. $X - a$ provided that $a \geq \min(X)$) for the set $\{x + a : x \in X\}$ (resp. $\{x - a : x \in X\}$). Now, for each $0 \leq a < c$, $a \in \mathbb{N}^*$, define

$$X'_a = \begin{cases} X_a - a & \text{if } 0 \leq a < c/2, \\ (X_a - a) \cup (X_a + a) & \text{if } a = c/2, \\ X_a + (c - a) & \text{if } c/2 < a \leq c - 1. \end{cases}$$

Let $X' = X_0 \cup X'_1 \cup X'_2 \cup \dots \cup X'_{c-1}$. Clearly $|X'| = |X|$ if c is odd and $|X'| \leq 2|X|$ if c is even. We claim that X' is a 2-generating set of S . Indeed, let s_i be any integer of S . We need to consider

two cases. (i) If $s_i \in X_0 \cup 2X_0$ we are done since $X_0 \subseteq X'$. (ii) If $s_i \notin X_0 \cup 2X_0$ then there exists $x_j \in X_a$ and $x_k \in X_{c-a}$, $a \leq c-a$, such that $s_i = x_j + x_k$ (this follows from $s_i \equiv 0 \pmod{c}$). If $a \neq c-a$ then $s_i = x_j + x_k = (x_j - a) + (x_k + (c - (c-a)))$, and hence $s_i \in 2(X'_a \cup X'_{c-a})$. If $a = c-a$ (thus, c must be even), s_i can be written $s_i = x_j + x_k = (x_j - a) + (x_k + a)$, and hence $s_i \in 2X'_{c/2}$. Then it follows that $X'/c \subset \mathbb{N}^*$ is a 2-generating set of S/c , and hence $\text{rk}_2(S) \geq \text{rk}_2(S/c)$ if c is odd and $\text{rk}_2(S/c) \leq 2\text{rk}_c(S)$ if c is even. \square

To complement Lemma 7, we observe that we may have $\text{rk}_2(S/2c) < 2\text{rk}_2(S)$ for even c as shown in the following example.

Example 1. For any $c \in \mathbb{N}^*$, let $S = \{14c, 16c, 18c\}$. Clearly, $X = \{7c, 9c\}$ is a 2-generating set of S , and hence $\text{rk}_2(S) = 2$. But $S/2c = \{7, 8, 9\}$ has no smaller 2-generating set than itself, and hence $\text{rk}_2(S/2c) = \text{card}(S/2c) = 3$.

The upper-bound $\text{rk}_2(S/c) \leq 2\text{rk}_c(S)$ in Lemma 7 is, however, not over-estimated, as shown by the following lemma.

Lemma 8. *For any $n \in \mathbb{N}^*$, there exists a set $S \subseteq \mathbb{N}^*$ of cardinality n such that*

$$\frac{\text{rk}_2(S/2)}{\text{rk}_2(S)} = 2 - \frac{1}{n+1}.$$

Proof. Let $b > 8$ be some fixed even integer. For any $n \in \mathbb{N}^*$, let $S = \{2\} \cup \{b^i + 2 : 1 \leq i \leq n\} \cup \{2b^i + 2 : 1 \leq i \leq n\}$. Let us decompose our proof into two claims.

Claim 1. $\text{rk}_2(S) = n + 1$.

Proof. For one, $X = \{1\} \cup \{b^i + 1 : 1 \leq i \leq n\}$ is a 2-generating set of S , and hence $\text{rk}_2(S) \leq n + 1$. For another, for $S' = \{2\} \cup \{b^i + 2 : 1 \leq i \leq n \wedge i \equiv 1 \pmod{2}\} \cup \{2b^i + 2 : 1 \leq i \leq n \wedge i \equiv 0 \pmod{2}\} \subset S$, the 2-generating interval system $\mathcal{I}_2(S')$ is composed of pairwise disjoint intervals, and hence, according to Corollary 1, $\text{rk}_2(S') = |S'| = n + 1$. Therefore, $\text{rk}_2(S) \geq \text{rk}_2(S') = n + 1$. \square

Claim 2. $\text{rk}_2(S/2) = 2n + 1$.

Proof. The upper-bound $\text{rk}_2(S/2) \leq 2n + 1$ trivially follows from $|S/2| = |\{1\} \cup \{b^i/2 + 1 : 1 \leq i \leq n\} \cup \{b^i + 1 : 1 \leq i \leq n\}| = 2n + 1$. To prove $\text{rk}_2(S/2) \geq 2n + 1$, apply Lemma 3 with $a_i = b^i/4 + 1$, $1 \leq i \leq n$, to the subset $(S/2) \setminus \{2\} = \{b^i/2 + 1 : 1 \leq i \leq n\} \cup \{b^i + 1 : 1 \leq i \leq n\}$, and conclude by observing that $\text{rk}_2(S/2) = 1 + \text{rk}_2((S/2) \setminus \{2\})$ as soon as $b > 8$. \square

Combining the above two claims yields the lemma. \square

4 Hardness

MINIMUM GENERATING SET (*i.e.* given a set of positive integers S , find a minimum cardinality set of integers X such that every element of S is the sum of a subset of X) was proved to be **NP**-complete in [4]. We complement this result by showing that MINIMUM 2-GENERATING SET is **APX**-hard, *i.e.*, hard to approximate within ratio $1 + \varepsilon$ for any $\varepsilon > 0$.

Proposition 1. MINIMUM 2-GENERATING SET is **APX**-hard.

Proof. We propose an L-reduction [17] from VERTEX COVER for cubic graphs: Given a cubic graph $G = (V, E)$, find a minimum cardinality vertex cover of G , *i.e.*, a subset $V' \subseteq V$ such that, for each edge $\{u, v\} \in E$, at least one of u and v belongs to V' . MINIMUM VERTEX COVER for cubic graphs is **APX**-complete [1,18].

Assume, without loss of generality, that $V = \{1, 2, \dots, n\}$. Define the corresponding instance of MINIMUM k -GENERATING SET by defining $S \subset \mathbb{N}^*$ to be $S = \{b^0\} \cup \{b^i : 1 \leq i \leq n\} \cup \{2b^i : 1 \leq i \leq n\} \cup \{b^0 + b^i : 1 \leq i \leq n\} \cup \{b^0 + b^i + b^j : \{i, j\} \in E\}$ for some even constant b to be defined later. We claim that there exists a vertex cover of G of cardinality at most k if and only if there exists a 2-generating set for S of cardinality at most $n + k + 1$.

Suppose that there exists a vertex cover $V' \subseteq V$ of cardinality k of G . Define $X \subset \mathbb{N}^*$ (actually $X \subset S$) to be $X = \{b^0\} \cup \{b^i : 1 \leq i \leq n\} \cup \{b^0 + b^i : i \in V'\}$. We claim that X is a 2-generating set for S . Since $X \subset S$ and $b^0 \in X$, it is enough to prove that, for each $\{i, j\} \in E$, $b^0 + b^i + b^j$ is 2-generated by X . Indeed, since V' is a vertex cover of G , we have $i \in V'$ or $j \in V'$ (possibly both), and if we let $\ell = i$ if $i \in V'$ and $\ell = j$ if $i \notin V'$, we have $(b^0 + b^\ell) \in X$. Therefore $b^0 + b^i + b^j$ is 2-generated by X as $(b^0 + b^\ell) + b^{\ell'}$, where $\ell' = j$ if $\ell = i$ and $\ell' = i$ otherwise.

Conversely, Let X be a 2-generating set of S . We first note that, by integrality, $b^0 \in X$. Consider any integer $1 \leq i \leq n$, and let I_i be the interval $[b^i/2 : 2b^i]$. According to Lemma 2, $|X \cap [b^i/2 : b^i]| \geq 1$ and $|X \cap [b^i : 2b^i]| \geq 1$ since $b^i \in S$ and $2b^i \in S$. Then it follows that $|X \cap I_i| \geq 1$, and $b^i \in X$ if the inequality holds as equality. We now observe that for $b > 4$ we have $2b^i < b^{i+1}/2$, $1 \leq i < n$. Then it follows that the intervals I_i , $1 \leq i \leq n$, are pairwise disjoint, and hence $|X| \geq n + 1$. Now, let $k \in \mathbb{N}^*$ be such that $|X| = n + k + 1$, and let $V' \subseteq V$ be such that $|X \cap I_i| > 1$ for every $i \in V'$. According to the above, we have $|V'| \leq k$. We now claim that V' is a vertex cover of G . Indeed, assume, aiming at a contradiction, that there exists $\{i, j\} \in E$ such that $|X \cap I_i| = 1$ and $|X \cap I_j| = 1$, and, to shorten notation, set $s_i = b^0 + b^i + b^j$. Then it follows that $X \cap I_i = \{b^i\}$ and $X \cap I_j = \{b^j\}$. But $s_i \in S$, and hence $|X \cap [s_i/2 : s_i]| \geq 1$ (Lemma 2). Furthermore, if we assume $i > j$, we have $b^i/2 < s_i/2$ and $s_i < 2b^i$, and hence $[s_i/2 : s_i] \subset I_i$, *i.e.*, $[s_i/2 : s_i]$ is a subinterval of I_i . But $X \cap I_i = \{b^i\}$, and hence we must have $(b^0 + b^j) \in X$. This is the desired contradiction since $(b^0 + b^j) \in I_j$ and $X \cap I_j = \{b^j\}$.

We omit the easy proof that the described reduction is indeed an L-reduction with parameters $\alpha = 8$ and $\beta = 1$ (crucial is the fact that $|V| \leq 6|V'|$ for any vertex cover V' since G is a cubic graph). \square

It remains open whether MINIMUM 2-GENERATING SET is strongly **NP**-complete, *i.e.* whether MINIMUM 2-GENERATING SET is **NP**-complete if every integer in S is bounded by a polynomial in the length of the input. Indeed, neither Proposition 1 nor the **NP**-hardness result of [4] rule out the existence of a pseudo-polynomial algorithm for MINIMUM 2-GENERATING SET. Observe that this question reduces to 2-covering a set of strings S for an unary alphabet with a set X of substrings in S , where X is said to 2-cover S if every string in S can be written as a concatenation of at most two substrings in X [13].

Approximation issues of MINIMUM 2-GENERATING SET are completely unexplored yet. Notice, however, that, as long as every integer in S is not bounded by a polynomial in the length of the input, none of the approximation results of [12] and [13] applies.

5 Put the blame on $\text{rk}_2(S)$ only

Let S be any instance of MINIMUM 2-GENERATING SET. Write $n = |S|$, $m = \max(S)$ and $k = \text{rk}_2(S)$. This section is devoted to finding a minimum cardinality 2-generating set of S (from an effective computational point of view [7,16]).

As a first attempt, let us consider the brute-force approach: generate all k -subsets X of $\{1, 2, \dots, m\}$ and check for each of them whether it 2-generates S , *i.e.*, $S \subseteq X \cup 2X$. Correctness of this algorithm is of course immediate. There are $\binom{m}{k}$ such subsets and each subset X can be identified as a 2-generating set of S in $O(k^2 \log k)$ time (assuming a standard unit-cost RAM model with $\log m$ word size). Therefore, the brute-force algorithm is, as a whole, a $O(m^k k^2 \log k)$ time procedure. But m (and even $\log m$) can be arbitrarily large compared to $n = O(k^2)$ and this naturally leads us to the problem of trying to confine the seemingly inevitable combinatorial explosion of computational difficulty to a function of k only [7,16]. We prove here that such an

algorithm does exist for finding a minimum cardinality 2-generating set of S . Surprisingly enough, the time complexity of the proposed algorithm turns out to be even independent of $\max(S)$ (again assuming a standard unit-cost RAM model with $\log m$ word size). The main result of this paper can be stated as follows.

Lemma 9 (standard representation). *Let $S = \{s_i : 1 \leq i \leq n\} \subset \mathbb{N}^*$ and write k for $\text{rk}_2(S)$. Then, there exist rationals $\alpha_{i,j} \in \{-1, -2^{-1}, 0, 2^{-1}, 1\}$, $1 \leq i \leq k$ and $1 \leq j \leq n$, such that*

$$X = \left\{ \sum_{j=1}^n \alpha_{i,j} s_j : 1 \leq i \leq k \right\}$$

is a minimum cardinality 2-generating set of S .

Before proving Lemma 9, we need a new definition that translates the problem to elementary graph theory terms. Let $S = \{s_1, s_2, \dots, s_n\}$ be a set of positive integers and $X = \{x_1, x_2, \dots, x_k\}$ be a 2-generating set for S . Define an X -realization of S to be a bipartite graph $B = (S, X, E)$ such that $d_B(s) \in \{1, 2\}$ for all $s \in S$, and

- if $d_B(s) = 1$, say $\{s, x_i\} \in E$, then $s = x_i$ or $s = 2x_i$, and
- if $d_B(s) = 2$, say $\{s, x_i\} \in E$ and $\{s, x_j\} \in E$, $x_i \neq x_j$, then $s = x_i + x_j$.

We refer to Figure 1 for an illustration.

Note that, in the above definition of an X -realization, X (resp. S) is considered as a set of integers, *and* as a set of vertices in a graph. We chose not to correct this ambiguity in the rest of the paper, in order to avoid heavy notations. Besides, the context will always be clear about the fact that we are concerned with integers or vertices.

Coming back to X -realizations, it is clear that every simple cycle of B has length at least 6 (a simple cycle of length 4, say (x_1, s_1, x_2, s_2) , would result in the contradiction $s_1 = x_1 + x_2 = s_2$). An X -realization of S is said to be *minimum* if X is a minimum cardinality 2-generating set of S . Of course, an X -realization of a set S may not be unique ; for example, referring to Figure 1, replacing edge $\{4, 2\}$ by $\{4, 1\}$ and $\{4, 3\}$ results in another X -realization of S .

Structures in a minimum X -realization of a set S will prove extremely useful.

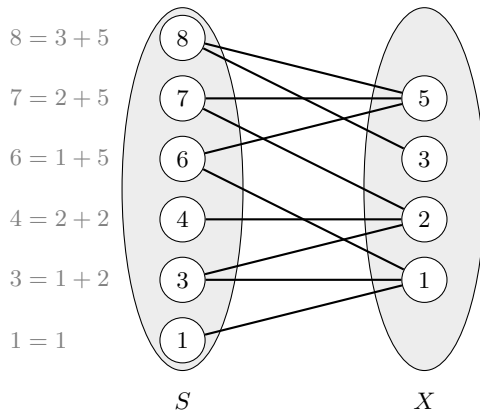


Fig. 1. An X -realization of $S = \{1, 3, 4, 6, 7, 8\}$ for $X = \{1, 2, 3, 5\}$.

Lemma 10. *Let $S \in \mathbb{N}^*$. Then there exists a minimum X -realization $B = (S, X, E)$ of S such that, for every $x \in X$, $d_B(x) > 1$ if $|\text{CC}_B(x)| > 2$.*

Proof. Let $\mathcal{B}(S)$ be the set of all X -realizations of S for all minimum cardinality 2-generating sets X of S . For each $B = (S, X, E) \in \mathcal{B}(S)$, define $f(B) = |\{x \in X : d_B(x) = 1 \wedge |\text{CC}_B(x)| > 2\}|$. Then, let $B_{\text{opt}} = (S, X_{\text{opt}}, E_{\text{opt}}) \in \mathcal{B}(S)$ be such that (i) $f(B_{\text{opt}}) \leq f(B)$ for all $B \in \mathcal{B}(S)$, and (ii) $|\text{CC}_{B_{\text{opt}}}| \geq |\text{CC}_B|$ for any $B \in \mathcal{B}(S)$ satisfying $f(B) = f(B_{\text{opt}})$. We claim that $f(B_{\text{opt}}) = 0$, thereby proving the lemma. Assume, aiming at a contradiction, that $f(B_{\text{opt}}) > 0$. Then, there exists $x \in X_{\text{opt}}$ such that $d_{B_{\text{opt}}}(x) = 1$ and $|\text{CC}_{B_{\text{opt}}}(x)| > 2$. Let s be the neighbor of x in B_{opt} . Since $|\text{CC}_{B_{\text{opt}}}(x)| > 2$, $d_{B_{\text{opt}}}(s) = 2$ and let $x' \in X_{\text{opt}}$ be the second neighbor of s , *i.e.*, $s = x + x'$. Clearly, $X = (X_{\text{opt}} \setminus \{x\}) \cup \{s\}$ is a minimum cardinality 2-generating set of S as well (notice that we must have $s \notin X_{\text{opt}}$ by minimality of X_{opt}). We now observe that an X -realization $B = (S, X, E)$ of S can be obtained from B_{opt} by simply deleting the edge $\{s, x'\}$. We now need to consider two cases: (1) If $d_B(x') > 1$ then $f(B) < f(B_{\text{opt}})$, and (2) If $d_B(x') = 1$ then $f(B) = f(B_{\text{opt}})$ and $|\text{CC}_{B_{\text{opt}}}| < |\text{CC}_B|$. Therefore both cases result in a contradiction, and hence $f(B_{\text{opt}}) = 0$. \square

Lemma 11. *Let $S \in \mathbb{N}^*$ and $B = (S, X, E)$ be a minimum X -realization of S such that B is connected. If every vertex $s \in S$ lies on a cycle, then there exists a simple cycle of B of length $4\ell + 2$ for some $\ell \geq 1$.*

Proof. Assume, aiming at a contradiction, that every vertex $s \in S$ lies on a cycle and that every simple cycle of B has length 4ℓ for some $\ell \geq 1$. Thus, by construction, $d_B(s) = 2$ for every $s \in S$. Define the S -contraction of B to be the graph $B' = (X, E')$, where $E' = \{\{x_i, x_j\} : \exists s \in S \text{ such that } \{x_i, s\} \in E \text{ and } \{x_j, s\} \in E\}$. In other words, B' is obtained from B by contracting every path (x_i, s, x_j) into a single edge $\{x_i, x_j\}$. Clearly, every cycle of B' has certainly even length since every cycle of B has length 4ℓ for some $\ell \geq 1$, and hence B' is bipartite. Write $X = X_1 \cup X_2$ the associated bipartition. Now, let $x_{\min} = \min\{x \in X\}$, and assume, without loss of generality, that $x_{\min} \in X_1$. Define $X' = X'_1 \cup X'_2$, where $X'_1 = \{x - x_{\min} : x \in X_1 \setminus \{x_{\min}\}\}$ and $X'_2 = \{x + x_{\min} : x \in X_2\}$. We claim that X' is a 2-generating set of S of cardinality $|X| - 1$, and hence get a contradiction of the minimality of X . Indeed, let s be any element of S and let $\{x_i, s\}$ and $\{x_j, s\}$ be the two associated edges of B , *i.e.*, $s = x_i + x_j$. According to the above, x_i and x_j belong to different partite sets in B' , say $x_i \in X_1$ and $x_j \in X_2$. We need to consider two cases. If $x_i \neq x_{\min}$ and $x_j \neq x_{\min}$, then $s = (x_i - x_{\min}) + (x_j + x_{\min})$ with $(x_i - x_{\min}) \in X'_1$ and $(x_j + x_{\min}) \in X'_2$. Otherwise, we must have $x_i = x_{\min}$, and hence $s = x_i + x_j = x_{\min} + x_j$ with $(x_j + x_{\min}) \in X'_2$. \square

We are now in position to prove Lemma 9.

Proof (of Lemma 9). Write $k = \text{rk}_2(S)$. Let $X = \{x_i : 1 \leq i \leq k\}$ be a minimum cardinality 2-generating set of S and $B = (S, X, E)$ be any X -realization of S . Let B_1, B_2, \dots, B_q be the connected components of B . We consider each connected component of B separately.

Consider thus any connected component $B_i = (S_i, X_i, E_i)$ of B with $S_i \subseteq S$ and $X_i \subseteq X$. Without loss of generality, write $S_i = \{s_1, s_2, \dots, s_{n_i}\}$. It is enough to show that for any $x \in X_i$, there exist rationals $\alpha_j \in \{-1, -2^{-1}, 0, 2^{-1}, 1\}$, $1 \leq j \leq n_i$, such that $x = \sum_{1 \leq j \leq n_i} \alpha_j s_j$, *i.e.*, x is the linear combination with coefficients taken from $\{-1, -2^{-1}, 0, 2^{-1}, 1\}$ of the vertices in S_i . We need to consider four cases: (1). B_i is a tree, or (2.1). B_i is not a tree and x lies on a simple cycle of length $4\ell + 2$ for some $\ell \geq 1$, or (2.2). B_i is not a tree, x does not lie on a simple cycle, and there exists a simple cycle of length $4\ell + 2$ for some $\ell \geq 1$ in B_i , or (2.3). B_i is not a tree, and every simple cycle C of B_i has length $4\ell_C$ for some $\ell_C \geq 1$ (regardless of whether or not x lies on any cycle).

(1). B_i is a tree. According to Lemma 10, we may assume that there exists a vertex of S_i , say s_1 , such that $d_{B_i}(s_1) = 1$, *i.e.*, s_1 is a leaf in B_i . Let P be the path from vertex s_1 to vertex x (such a path exists since B_i is connected and is unique since B_i is acyclic). Without loss of

generality, write $P = (s_1, x_1, s_2, x_2, \dots, x_{k-1}, s_k, x)$. Then it follows that

$$\begin{cases} s_1 = \delta x_1 \\ s_2 = x_1 + x_2 \\ \vdots \\ s_k = x_{k-1} + x \end{cases}$$

for some $\delta \in \{1, 2\}$, and hence

$$x = \frac{(-1)^k}{\delta} s_1 + \sum_{i=2}^k (-1)^{k-i} s_i. \quad (1)$$

Therefore there exist rationals $\alpha_i \in \{-1, -2^{-1}, 2^{-1}, 1\}, 1 \leq i \leq k$, such that

$$x = \sum_{i=1}^k \alpha_i s_i,$$

i.e., x is the linear combination with coefficients taken from $\{-1, -2^{-1}, 2^{-1}, 1\}$ of those vertices s_i that lie on the - unique - path from s_1 to x .

(2). B_i **is not a tree**. In that case, notice that since graph B_i is bipartite, any cycle that starts at a vertex in S_i must alternate between vertices in S_i and X_i , and hence must be of even length (on return to the start vertex again). We now need to distinguish three subcases.

– (2.1). **Vertex x lies on a cycle C of length $4\ell + 2$ for some $\ell \geq 1$** . Without loss of generality, write $C = (s_1, x_1, s_2, x_2, \dots, s_k, x)$ a cycle of length $k = 4\ell + 2$ for some $\ell \geq 1$ that contains our target vertex x . Then it follows that

$$\begin{cases} s_1 = x_1 + x \\ s_2 = x_1 + x_2 \\ \vdots \\ s_k = x_{k-1} + x \end{cases}$$

and hence

$$2x = \sum_{i=1}^k (-1)^{i+1} s_i \quad (2)$$

since C has length $4\ell + 2$ for some $\ell \geq 1$ Therefore there exist rationals $\alpha_i \in \{-2^{-1}, 2^{-1}\}, 1 \leq i \leq k$, such that

$$x = \sum_{i=1}^k \alpha_i s_i,$$

i.e., x is the linear combination with coefficients taken from $\{-2^{-1}, 2^{-1}\}$ of those vertices s_i that lie on a cycle vertex x is part of.

– (2.2). **Vertex x does not lie on a simple cycle, and there exists a simple cycle of length $4\ell + 2$ for some $\ell \geq 1$ in B_i** . Observe first that, since every vertex of S has degree at most 2 in B_i , every path yielding our target vertex x to any cycle C of B_i enters C at a vertex of X_i . Consider a shortest path yielding vertex x to a cycle C of length $4\ell + 2$ for some $\ell \geq 1$, say $P = (x, s_1, x_1, \dots, x_{k-1}, s_k, x_k)$ and $C = (x_k, s_{k+1}, x_{k+1}, \dots, x_{k+p-1}, s_{k+p})$, with

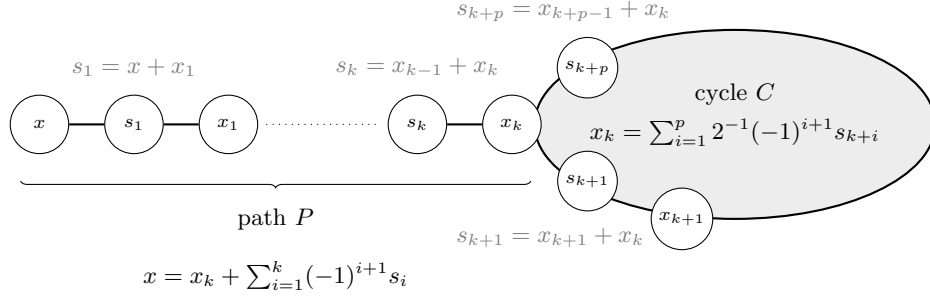


Fig. 2. Connected component B_i is not acyclic and vertex x does not lie on any cycle.

$p = 2\ell + 1$; see Figure 2 for an illustration. Clearly, since vertex x does not lie on a cycle and P is a shortest path, all vertices of P but vertex x_k do not lie on cycle C . For one, we have

$$\begin{cases} s_1 = x_1 + x \\ s_2 = x_1 + x_2 \\ \vdots \\ s_k = x_{k-1} + x_k \end{cases}$$

and hence

$$x = x_k + \sum_{i=1}^k (-1)^{i+1} s_i. \quad (3)$$

For another, according to case 2.1. above, we have

$$x_k = \sum_{i=1}^p 2^{-1}(-1)^{i+1} s_{k+i}. \quad (4)$$

Combining (3) and (4) yields

$$x = \sum_{i=1}^p 2^{-1}(-1)^{i+1} s_{k+i} + \sum_{i=1}^k (-1)^{i+1} s_i.$$

Then it follows that there exist rationals $\alpha_i \in \{-1, -2^{-1}, 0, 2^{-1}, 1\}, 1 \leq i \leq n$, such that

$$x = \sum_{i=1}^n \alpha_i s_i.$$

More precisely, x is the linear combination with coefficients taken from $\{-1, -2^{-1}, 2^{-1}, 1\}$ of those vertices s_i that lie on a shortest path yielding vertex x to a cycle C or lie on the cycle C .

- (2.3). **Every simple cycle C of B_i has length $4\ell_C$ for some $\ell_C \geq 1$ (regardless of whether or not x lies on any cycle).**

According to Lemma 11, not all vertices $s \in S_i$ have degree 2. Let $s_k \in S_i$ be any vertex such that $d_B(s_k) = 1$ and let $P = (x, s_1, x_1, \dots, x_{k-1}, s_k)$ be any simple path from x to s_k . Again,

we have

$$\begin{cases} s_1 = x_1 + x \\ s_2 = x_1 + x_2 \\ \vdots \\ s_k = \delta x_{k-1} \end{cases}$$

for some $\delta \in \{1, 2\}$, and hence

$$x = \frac{(-1)^{k+1}}{\delta} s_k + \sum_{i=1}^{k-1} (-1)^{i+1} s_i.$$

Then it follows that there exist rationals $\alpha_i \in \{-1, -2^{-1}, 0, 2^{-1}, 1\}, 1 \leq i \leq n$, such that

$$x = \sum_{i=1}^n \alpha_i s_i.$$

More precisely, x is the linear combination with coefficients taken from $\{-1, -2^{-1}, 2^{-1}, 1\}$ of those vertices s_i that lie on a path yielding vertex x to a vertex s_k of degree 1. \square

Thanks to Lemma 9, we prove that there exists an algorithm for MINIMUM 2-GENERATING SET that confines the combinatorial explosion of computational difficulty to a function of $k = \text{rk}_2(S)$ only.

Proposition 2. *Assuming a standard unit-cost RAM model with $\log m$ word size ($m = \max(S)$), there exists an algorithm for finding a minimum cardinality 2-generating set of S that runs in $O(5^{\frac{k^k(k+3)^k}{2^k}} k^2 \log k)$ time, where $k = \text{rk}_2(S)$.*

Proof. We propose a brute-force algorithm for finding a (standard representation of a) minimum cardinality 2-generating set of S . The basic idea is to consider the set $C(S)$ of all linear combinations $\alpha_1 s_1 + \alpha_2 s_2 + \dots + \alpha_n s_n$ with coefficients taken from $\{-1, -2^{-1}, 0, 2^{-1}, 1\}$. Clearly, there exist 5^n such combinations. The algorithm simply tries each k -subset X of $C(S)$ and checks whether $S \subseteq X \cup 2X$. Correctness of the algorithm follows from Lemma 9. We now turn to proving its time complexity. Let N be the number of k -subsets of $C(S)$. Clearly, $N = \binom{5^n}{k} = O(5^{n^k})$. But $n \leq \frac{k(k+3)}{2}$, and hence $N = O(5^{\frac{k^k(k+3)^k}{2^k}})$. Since a k -subset X of $C(S)$ can be identified as a 2-generating set of S in $O(k^2 \log k)$ time (assuming a standard unit-cost RAM model with $\log m$ word size), the total running time is $O(Nk^2 \log k) = O(5^{\frac{k^k(k+3)^k}{2^k}} k^2 \log k)$. \square

6 Conclusion

MINIMUM 2-GENERATING SET is a natural restriction of MINIMUM GENERATING SET with prospective applications (see [4]). Our standard representation (Lemma 9) provides a first positive algorithmic result for computing minimum 2-generating sets. We mention here some directions of interest for future works:

- Is MINIMUM 2-GENERATING SET pseudo-polynomial time solvable? Notice that this question is related to 2-covering a set of strings S for a unary alphabet with a set X of substrings in S , where X is said to 2-cover S if every string in S can be written as a concatenation of at most two substrings in X [13].

- For any $k > 1$, a set of integers S is said to be k -simplifiable if $\text{rk}_k(S) < |S|$ [15]. Is there a polynomial-time algorithm for deciding whether S is 2-simplifiable? An important related problem is the following: Given a set of integers S and a 2-generating set X of S of cardinality p , design an efficient algorithm that either returns a 2-generating set of S of cardinality $p - 1$ or returns that no such 2-generating set of S exists.
- Considering the general MINIMUM k -GENERATING SET problem, is there an analog of Lemma 9 for every fixed $k \geq 2$?

Acknowledgments

The authors are thankful to Olivier Serre for helpful discussions.

References

1. P. Alimonti and V. Kann, *Some APX-completeness results for cubic graphs*, Theoretical Computer Science **237** (2000), no. 1-2, 123–134.
2. H.L. Bodlaender, R.G. Downey, M.R. Fellows, M.T. Hallett, and H.T. Wareham, *Parameterized complexity analysis in computational biology*, Computer Applications in the Biosciences **11** (1995), 49–57.
3. C. Choffrut and J. Karhumäki, *Handbook of formal languages, Vol. 1: Word, language, grammar*, ch. Combinatorics of words, pp. 329–438, Springer-Verlag, 1997.
4. M.J. Collins, D. Kempe, J. Saia, and M. Young, *Nonnegative integral subset representations of integer sets*, Information Processing Letters **101** (2007), no. 3, 129–133.
5. M. Develin, *Optimal subset representations of integer sets.*, Journal of Number Theory **89** (2001), 212–221.
6. R. Diestel, *Graph theory*, second ed., Graduate texts in Mathematics, no. 173, Springer-Verlag, 2000.
7. R. Downey and M. Fellows, *Parameterized complexity*, Springer-Verlag, 1999.
8. M.A. Fitch and R.E. Jamison, *Minimum sum covers of small cyclic groups*, Congressus Numerantium **147** (2000), 65–81.
9. A. Gyárfás, *Combinatorics of intervals, preliminary version*, Institute for Mathematics and its Applications (IMA) Summer Workshop on Combinatorics and Its Applications, 2003, available online at <http://www.math.gatech.edu/news/events/ima/newag.pdf>.
10. H. Haanpää, *Minimum sum and difference covers of abelian groups*, Journal of Integer Sequences **7** (2004), no. 2, Article 04.2.6.
11. H. Haanpää, A. Huima, and P.R.J. Östergård, *Sets in \mathbb{Z}_n with distinct sums of pairs*, Discrete Applied Mathematics **138** (2004), no. 1-2, 99–106.
12. M. Hajiaghayi, K. Jain, L. Lau, A. Russell I. Mandoiu, and V. Vazirani, *Minimum multicolored subgraph problem in multiplex PCR primer set selection and population haplotyping*, Proc. 6th International Conference on Computational Science (ICCS), Part II, Reading, UK (V.N. Alexandrov, G. Dick van Albada, P.M.A. Sloot, and J. Dongarra, eds.), LNCS, vol. 3994, Springer, 2006, pp. 758–766.
13. D. Hermelin, D. Rawitz, R. Rizzi, and S. Vialette, *The minimum substring cover problem*, Proc. 5th International Workshop on Approximation and Online Algorithms (WAOA), Eilat, Israel (C. Kaklamani and M. Skutella, eds.), Lecture Notes in Computer Science, no. 4927, 2007, pp. 170–183.
14. D. Moulton and D. Petrie, *Representing powers of numbers as subset sums of small sets.*, Journal of Number Theory **89** (2001), 193–211.
15. J. Néraud, *Elementariness of a finite set of words is coNP-complete*, Theoretical Informatics and Applications **24** (1990), no. 5, 459–470.
16. R. Niedermeier, *Invitation to fixed parameter algorithms*, Lecture Series in Mathematics and Its Applications, Oxford University, Press, 2006.
17. C.H. Papadimitriou, *Computational complexity*, Addison-Wesley, 1994.
18. C.H. Papadimitriou and M. Yannakakis, *Optimization, approximation and complexity classes*, Journal of Computer and System Sciences **43** (1991), 425–440.
19. C.N. Swanson, *Planar cyclic difference packings*, Journal of Combinatorial Designs **8** (2000), 426–434.
20. D. Wiedemann, *Cyclic difference covers through 133*, Congressus Numerantium **90** (1992), 181–185.