$\mathcal{NP}$-completeness of generalized multi Skolem sequences

Gustav Nordh [1]

Department of Computer and Information Science, Linköpings Universitet, S-581 83 Linköping, Sweden

Abstract

A Skolem sequence is a sequence $a_1, a_2, \ldots, a_{2n}$ (where $a_i \in A = \{1, \ldots, n\}$), each a_i occurs exactly twice in the sequence and the two occurrences are exactly a_i positions apart. A set A that can be used to construct Skolem sequences is called a Skolem set. The existence question of deciding which sets of the form $A = \{1, \ldots, n\}$ are Skolem sets was solved by Thoralf Skolem [6] in 1957. Many generalizations of Skolem sequences have been studied. In this paper we prove that the existence question for generalized multi Skolem sequences is $\mathcal{NP}$-complete. This can be seen as an upper bound on how far the generalizations of Skolem sequences can be taken while still hoping to resolve the existence question.

Key words: Skolem sequence, design theory, NP-completeness

1 Introduction

Skolem sequences were introduced by Thoralf Skolem [6] in 1957, for the construction of Steiner triple systems. He considered sets of the form $A = \{1, 2, \ldots, n\}$ and asked whether one can always form a sequence with two copies of every element k in the set so that the two copies of k are placed k places apart in the sequence. Such sequences are called **Skolem sequences**. For example, the set $\{1, 2, 3, 4\}$ can be used to construct the sequence 42324311, but the set $\{1, 2, 3\}$ cannot be used to form such a sequence. A set that can be used to construct a Skolem sequence is called a **Skolem set**.

Email address: gusno@ida.liu.se (Gustav Nordh).

[1] Supported by the *National Graduate School in Computer Science* (CUGS), Sweden.

Many different aspects and generalizations of Skolem sequences have been studied. One reason for them being so well studied is that they have important applications in several branches of mathematics; Shalaby [5] describes applications in design theory and graph labelings.

Baker [1] introduced generalized Skolem sequences and used them to construct k-extended Skolem sequences. They have also been used in the construction of extended Langford sequences with small defects [3]. A **generalized Skolem sequence** is a sequence of positive integers and null symbols such that an integer appears exactly twice or not at all, and the two appearances of an integer j are j positions apart. If the integers in A can be used to construct a generalized Skolem sequence using only the positions in P, we say that (P, A) is a **generalized Skolem pair**. For example, $(\{1, 2, 4, 5, 7, 8\}, \{1, 5, 7\})$ is a generalized Skolem pair. The corresponding generalized Skolem sequence is 75011057 (0 occupies positions that are not in P). Note that a pair (P, A) is a generalized Skolem pair if and only if the positions in P can be partitioned into the differences in A, e.g., the example $(\{1, 2, 4, 5, 7, 8\}, \{1, 5, 7\})$ above is a generalized Skolem pair since $\{5-4, 7-2, 8-1\} = \{1, 5, 7\}$. Hence, we will refer to the elements in A as the differences in A.

We generalize the notion of generalized Skolem sequences slightly and allow the set of differences A to be a multiset. We call these sequences **generalized multi Skolem sequences**, and the corresponding pair (P, A) **generalized multi Skolem pair**. For example, $(\{1, 2, 4, 5, 7, 8\}, \{1, 6, 6\})$ is a generalized multi Skolem pair. The corresponding generalized multi Skolem sequence is 66011066 (0, occupies positions that are not in P). Linek and Shalaby [4] give some necessary conditions for the existence of generalized multi Skolem pairs. Moreover, they state that a basic question is to decide which pairs (P, A) are generalized multi Skolem pairs. We prove that the problem of deciding which pairs (P, A) are generalized multi Skolem pairs is $\mathcal{NP}$-complete. The proof is a reduction from the $\mathcal{NP}$-complete problem Multiple Choice Matching [2]. We refer the reader to Garey and Johnson [2] for an in-depth treatment of the theory of $\mathcal{NP}$-completeness.

2 $\mathcal{NP}$-completeness of Generalized multi Skolem sequences

We prove that Generalized Multi Skolem Sequences is $\mathcal{NP}$-complete by giving a reduction from the $\mathcal{NP}$-complete problem Multiple Choice Matching. Before presenting the reduction we define the two problems formally and prove some additional properties of Multiple Choice Matching that we will use in the reduction.

GENERALIZED MULTI SKOLEM SEQUENCES

INSTANCE: A multiset A of positive integers, $|A| = m$, a set P of positive integers, $|P| = 2m$.
QUESTION: Is (P, A) a generalized multi Skolem pair? That is, can the positions in P be partitioned into the differences in A?

MULTIPLE CHOICE MATCHING
INSTANCE: A graph $G = (V, E)$, a partition of the edges E into disjoint sets $E_1, E_2, \ldots, E_m$, and a positive integer K.
QUESTION: Is there a subset $M \subseteq E$ with $|M| \geq K$ such that no two edges in M share a common vertex and such that M contains at most one edge from each E_i, $1 \leq i \leq m$?

The definition of MULTIPLE CHOICE MATCHING is taken from Garey and Johnson [2], where it is also stated that the problem remains $\mathcal{NP}$-complete even if each E_i contains at most 2 edges, and $K = |V|/2$. We will make use of these properties in the reduction. A set M, as defined above is called a multiple choice matching. We think of the edges in E_i as being labeled with the label i. Another property of MULTIPLE CHOICE MATCHING that we need is $\mathcal{NP}$-completeness even in the restricted case where none of the edges with the same edge label share a common vertex. This is obviously true since two edges that share a common vertex can never be part of the same matching. Thus we can simply assign one of these edges a new edge label not previously used in G, see Figure 1. In the resulting graph none of the edges with the same edge label share a common vertex, and it is easy to see that the resulting graph has a multiple choice matching if and only if the original graph has one.

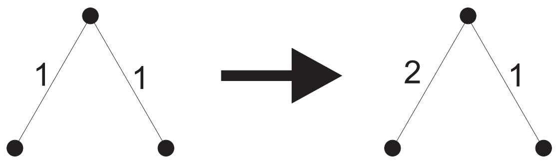

Fig. 1. The label 2 is a new one not previously used in the graph.

To simplify the reduction we prove one final property of MULTIPLE CHOICE MATCHING. MULTIPLE CHOICE MATCHING is $\mathcal{NP}$-complete even if the number of different edge labels in the graph is greater than half the number of vertices in the graph. If m equals the number of different edge labels in the graph and the number of vertices is $2n$, (assuming that $m \leq n$) we add $(n - m) + 1$ copies of the component in Figure 2 to the graph. The resulting graph have the property $m > n$, and again it is easy to see that the resulting graph has a multiple choice matching if and only if the original graph has one.

Theorem 1 GENERALIZED MULTI SKOLEM SEQUENCES *is $\mathcal{NP}$-complete.*

First note that it is easy to see that GENERALIZED MULTI SKOLEM SEQUENCES $\in \mathcal{NP}$. The proof will be a reduction from the $\mathcal{NP}$-complete prob-

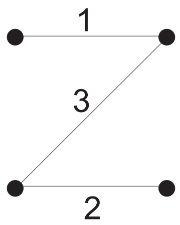

Fig. 2. All edge labels (in each of the $(n-m)+1$ copies of the component above) are new ones not previously used in the graph.

lem Multiple Choice Matching. Because of the discussion above we need only to consider instances of Multiple Choice Matching that have the following properties:

(1) No more than two edges have the same edge label.
(2) No two edges with the same edge label share a common vertex.
(3) The number of different edge labels in the graph is greater than half of the number of vertices in the graph.
(4) $K = |V|/2$

Overall structure: Given an instance of Multiple Choice Matching, $G = (V, E)$ we will construct a pair (P, A) which is a generalized multi Skolem pair if and only if G has a multiple choice matching. The construction will be divided into two main parts, the construction of a *matching component* and the construction of a *garbage component*. The *matching component* guarantees that if G does not have a multiple choice matching, then (P, A) will not be a generalized multi Skolem pair. The *garbage component* guarantees that if G has a multiple choice matching, then (P, A) will be a generalized multi Skolem pair. P can be represented as a sequence of zeros and ones where $p_i \in P$ if and only if position p_i in the sequence is occupied by a one. The polynomial under each component is the number of positions in that component, $n = |V|/2$, and m is the number of different edge labels in G. The overall structure of P is:

$$\underbrace{\textit{matching component}\ldots}_{\substack{6nm^2+14nm+\\6m^2+2n+2m}}\underbrace{\textit{barrier}\ldots}_{\substack{6nm^2+14nm+\\6m^2+2n+2m}}\underbrace{\textit{garbage component}\ldots}_{\substack{18nm^3+36nm^2-6n^2m^2+\\8nm-14n^2m+18m^3+4m^2-4n^2}}$$

Each barrier consists of all zeros, and the barrier above is longer than any of the differences we give later. The overall structure of the *matching component* is:

$$\underbrace{\textit{start pos}\ldots}_{\substack{2nm^2+2nm+\\2m^2}}\underbrace{\textit{barrier}\ldots}_{\substack{2nm^2+8nm+\\2m^2}}\underbrace{\textit{match pos}\ldots}_{2n}\underbrace{\textit{barrier}\ldots}_{\substack{2nm^2+2nm+\\2m^2}}\underbrace{\textit{stop pos}\ldots}_{2nm+2m}$$

Here *match pos* consists of all ones, and *start pos* consists of a sequence of m blocks of the form

$$\underbrace{\textit{start}\ldots}_{2n}\underbrace{\textit{barrier}\ldots}_{2nm+2m}$$

where *start* contains at most two ones. *Stop pos* consists of a sequence of m blocks of the form

$$\underbrace{\textit{stop}\ldots}_{2n}\underbrace{\textit{barrier}}_{2}$$

where *stop* contains at most two ones. The overall structure of the *garbage component* is:

$$\underbrace{\textit{garbage I}\ldots}_{\substack{12n^2m^2+28n^2m+\\12nm^2+4nm+4n^2}}\ \underbrace{\textit{garbage II}\ldots}_{\substack{(m-n)(10nm^2+24nm+\\10m^2+2m+4n)}}\ \underbrace{\textit{garbage III}\ldots}_{\substack{(m-n)(8nm^2+18nm+\\8m^2+2m+4n)}}$$

In what follows *start* and *absorb* consist of all ones. *Garbage I* consists of a sequence of n copies of the following component:

$$\underbrace{\textit{start}}_{1}\underbrace{\textit{barrier}\ldots}_{\substack{4nm^2+10nm+\\4m^2+2n}}\underbrace{\textit{absorb}\ldots}_{\substack{2nm^2+4nm+\\2m^2+2m-1}}\underbrace{\textit{barrier}\ldots}_{\substack{6nm^2+14nm+\\6m^2+2m+2n}}$$

Garbage II consists of a sequence of $m-n$ copies of the following component:

$$\underbrace{\textit{start}}_{1}\underbrace{\textit{barrier}\ldots}_{\substack{2nm^2+\\8nm+2m^2}}\underbrace{\textit{absorb}\ldots}_{\substack{2nm^2+2nm+\\2m^2+2n-1}}\underbrace{\textit{barrier}\ldots}_{\substack{6nm^2+14nm+\\6m^2+2m+2n}}$$

Garbage III consists of a sequence of $m-n$ copies of the following component:

$$\underbrace{\textit{start}}_{1}\underbrace{\textit{barrier}\ldots}_{\substack{2nm^2+\\2nm+2m^2}}\underbrace{\textit{absorb}\ldots}_{\substack{2nm+\\2m+2n-1}}\underbrace{\textit{barrier}\ldots}_{\substack{6nm^2+14nm+\\6m^2+2m+2n}}$$

Assume that the vertices in the graph G have been labeled with the numbers $1,\ldots,2n$, and that the edges have been labeled with the numbers $0,\ldots,m-1$. Also assume that if (i,j) is an edge in G, then $i<j$. P and A will be constructed in a step by step fashion, starting with the *matching component.* In each step we label the set which is constructed there in the most obvious way. For example if we want to add a set of differences to A in Step M2.1, we denote this set $A_{M2.1}$. Of course, we then let A be the union of all these sets, that is $A=A_x\cup A_y\cup\cdots\cup A_z$. P is handled in exactly the same way. We begin by giving the precise construction of P and A. A more easily understood explanation of what is actually achieved in every step is then given.

Matching component

Step M1: This step deals with the positions in *match pos.*
Step M1.1: Add all positions in *match pos* to P, that is $P_{M1.1}=\{4nm^2+10nm+4m^2+1,\ldots,4nm^2+10nm+4m^2+2n\}$.

Step M2: This step deals with the positions in *start pos* and *stop pos.*

Step M2.1: For all $k \in \{0, \ldots, m-1\}$, let $A_{M2.1.k} = \{4nm^2 + 10nm + 4m^2 - k(2nm + 2n + 2m)\}$
Step M2.2: For all $k \in \{0, \ldots, m-1\}$, let $A_{M2.2.k} = \{2nm^2 + 4nm + 2m^2 + 2m + 2n - (k+1)(2n+2)\}$.
Step M2.3: For each edge (i,j) in G labeled k, $k \in \{0, \ldots, m-1\}$, let $P_{M2.3.k.(i,j)} = \{k(2nm + 2n + 2m) + i, 6nm^2 + 14nm + 6m^2 + 2m + 2n - (k+1)(2n+2) + j\}$.
Step M2.4: For each edge (i,j) in G labeled k, $k \in \{0, \ldots, m-1\}$, let $A_{M2.4.k.(i,j)} = \{(6nm^2 + 14nm + 6m^2 + 2m + 2n - (k+1)(2n+2) + j) - (k(2nm + 2n + 2m) + i)\}$.

Garbage component

Step G1: This step deals with the positions in *garbage I*.
Step G1.1: Remember that the *garbage I* component consists of a sequence of n identical components. In each of these n components, add the position in *start* and all positions in *absorb* to P. That is for all $k \in \{0, \ldots, n-1\}$, let $P_{G1.1.k} = \{(k+1)(12nm^2 + 28nm + 12m^2 + 4n + 4m) + 1\} \cup \{(k+1)(12nm^2 + 28nm + 12m^2 + 4n + 4m) + 4nm^2 + 10nm + 4m^2 + 2n + 2, \ldots, (k+1)(12nm^2 + 28nm + 12m^2 + 4n + 4m) + 6nm^2 + 14nm + 6m^2 + 2n + 2m\}$
Step G1.2: Let $A_{G1.2} = \{1^{(n^2m^2 + 2n^2m + m^2 + m - 3n)}, 2^{2n}\}$, where a^b denotes b copies of a.

Step G2: This step deals with the positions in *garbage II*.
Step G2.1: Remember that the *garbage II* component consists of a sequence of $m-n$ identical components. In each of these $m-n$ components, add the position in *start* and all positions in *absorb* to P. That is for all $k \in \{0, \ldots, (m-n)-1\}$, let $P_{G2.1.k} = \{12n^2m^2 + 28n^2m + 24nm^2 + 32nm + 4n^2 + 12m^2 + 4n + 4m + k(10nm^2 + 24nm + 10m^2 + 2m + 4n) + 1\} \cup \{12n^2m^2 + 28n^2m + 26nm^2 + 40nm + 4n^2 + 14m^2 + 4n + 4m + k(10nm^2 + 24nm + 10m^2 + 2m + 4n) + 2, \ldots, 12n^2m^2 + 28n^2m + 28nm^2 + 42nm + 4n^2 + 16m^2 + 6n + 4m + k(10nm^2 + 24nm + 10m^2 + 2m + 4n)\}$
Step G2.2: Let $A_{G2.2} = \{1^{(m-n)(nm^2 + nm + m^2 + n - 2)}, 2^{(m-n)}\}$.

Step G3: This step deals with the positions in *garbage III*.
Step G3.1: Remember that the *garbage III* component consists of a sequence of $m-n$ identical components. In each of these $m-n$ components, add the position in *start* and all positions in *absorb* to P. That is for all $k \in \{0, \ldots, (m-n)-1\}$, let $P_{G3.1.k} = \{2n^2m^2 + 4n^2m + 10nm^3 + 38nm^2 + 34nm + 10m^3 + 14m^2 + 4n + 4m + k(8nm^2 + 18nm + 8m^2 + 2m + 4n) + 1\} \cup \{2n^2m^2 + 4n^2m + 10nm^3 + 40nm^2 + 36nm + 10m^3 + 16m^2 + 4n + 4m + k(8nm^2 + 18nm + 8m^2 + 2m + 4n) + 2, \ldots, 2n^2m^2 + 4n^2m + 10nm^3 + 40nm^2 + 38nm + 10m^3 + 16m^2 + 6n + 6m + k(8nm^2 + 18nm + 8m^2 + 2m + 4n)\}$

Step G3.2: Let $A_{G3.2} = \{1^{(m-n)(nm+n+m-2)}, 2^{(m-n)}\}$.

Explanation

Step M1: Here (in Step M1.1) we add all $2n$ positions in *match pos* to P. *Match pos* is in some sense the heart of the whole construction. The $2n$ positions in *match pos* represent the $2n$ vertices in G. The first position in *match pos* represents the vertex labeled 1, the second position represents the vertex labeled 2, and so on.

Step M2: Here we add some differences to A as well as certain positions in *start pos* and *stop pos* to P. Remember that *start pos* and *stop pos* each consists of m blocks, one for each of the different edge labels in G. The k-th leftmost block in *start pos* and the k-th rightmost block in *stop pos* is dealing with the edges in G labeled k.

In Step M2.1 we add one difference to A for each of the m different edge labels in G. These differences are used to represent the different edge labels in G and transfer information from the *start pos* component to the *match pos* component. More specifically, the difference in $A_{M2.1.k}$ is used to transfer information from the k-th leftmost block in *start pos* to the *match pos* component. Remember that if $a \in A$, a must be placed at positions p_i and p_j in the sequence, such that $|p_i - p_j| = a$ and $\{p_i, p_j\} \subseteq P$. The structure of P ensures that the difference in $A_{M2.1.k}$ can only be fitted in the *matching component* if one of the copies is placed in *match pos* and the other one in the k-th leftmost block in *start pos*. The difference in $A_{M2.1.k}$ has the additional property that if one copy is placed at the i-th leftmost position in its corresponding *start pos* block, the other copy must be placed at the i-th leftmost position in *match pos*. This is how we can transfer information from the *start pos* component to the *match pos* component. Note that if a difference in A is not fitted in the *matching component* it must be fitted in the *garbage component*.

The differences added to A in Step M2.2 have the same function and properties as the differences added to A in Step M2.1, with the difference that they transfer information from the *stop pos* component to the *match pos* component. Also note that the difference in $A_{M2.2.k}$ is used to transfer information from the k-th rightmost block (as opposed to from the k-th leftmost block, as is the case for the difference in $A_{M2.1.k}$) in *stop pos* to the *match pos* component.

In Step M2.3 we do the following: for each edge (i, j) in G labeled k, $k \in \{0, \ldots, m-1\}$, we add the i-th leftmost position in the k-th leftmost block in *start pos*, and the j-th leftmost position in the k-th rightmost block in *stop pos* to P. All other positions in *start pos* and *stop pos* are occupied by zeros (are not in P). Thus we have made sure that a difference from A can be fitted with one copy at the i-th leftmost position in the k-th leftmost block in *start* *pos* and the other copy at the i-th leftmost position in *match pos* if and only if there is an edge in G labeled k starting at vertex i. Similarly, a difference from A can be fitted with one copy at the j-th leftmost position in the k-th rightmost component in *stop pos* and the other copy at the j-th leftmost position in *match pos* if and only if there is an edge in G labeled k ending at vertex j.

Now let us take a moment and think of what remains to be done. We must make sure that either both of the differences in $A_{M2.1.k} \cup A_{M2.2.k}$ are fitted in the *garbage component*, or both of them are fitted in the *matching component*. We also must make sure that if the difference in $A_{M2.1.k}$ is fitted with one copy at the i-th leftmost position in *match pos* and the difference in $A_{M2.2.k}$ is fitted with one copy at the j-th leftmost position in *match pos*, then (i, j) is an edge (labeled k) in G. This is exactly what is achieved in Step M2.4. We formulate this as a lemma.

Lemma 2 *In a completed multi Skolem sequence with positions P and differences A, either both of the differences in $A_{M2.1.k} \cup A_{M2.2.k}$ are fitted in the* garbage component*, or both of them are fitted in the* matching component*. Moreover, if in the completed sequence the difference in $A_{M2.1.k}$ is fitted with one copy at the i-th leftmost position in* match pos *and the difference in $A_{M2.2.k}$ is fitted with one copy at the j-th leftmost position in* match pos*, then (i, j) is an edge (labeled k) in G.*

PROOF. It can be verified that if $A_{M2.4.k.(i,j)} = A_{M2.4.l.(a,b)}$, then $k = l$ and $(a, b) = (i + w, j + w)$ for some w. This is due to the fact that at most two edges in G have the same edge label. In particular, a difference in the multiset $\bigcup_{k.(i,j)} A_{M2.4.k.(i,j)}$ can occur at most twice, and if $A_{M2.4.k.(i,j)}$ and $A_{M2.4.k.(i+w,j+w)}$ are both present then the difference they contain may be fitted in two places in the *matching component*, i.e., the difference may occupy the positions in $P_{M2.3.k.(i,j)}$ or it may occupy the positions in $P_{M2.3.k.(i+w,j+w)}$.

Now assume that the difference in $A_{M2.1.k}$ is fitted with one copy at the i-th leftmost position in *match pos* and the difference in $A_{M2.2.k}$ is fitted with one copy at the j-th leftmost position in *match pos*, but that (i, j) is not an edge (labeled k) in G. This could only happen if there were two edges, (i, x) and (y, j) labeled k in G. But now consider the y-th leftmost position in the k-th leftmost block in *start pos*. This position is empty and must be filled by a difference in A, but no such difference exists. The only two candidates are the difference already used at the i-th leftmost position in *match pos* and the (possibly repeated) difference in $A_{M2.4.k.(y,j)}$, which cannot be used either because the j-th leftmost position in the k-th rightmost block is already taken.

Now assume that just one of the two differences in $A_{M2.1.k}$ and $A_{M2.2.k}$ is placed in the *matching component.* But following the same line of reasoning as above there would be no difference in A to fit in the empty position (in the k-th leftmost/rightmost block in *start pos/stop pos*) where the missing difference should have been placed. □

Now we are in a position to prove that if the pair (P, A) that we have constructed is a generalized multi Skolem pair, then G has a multiple choice matching.

Lemma 3 *If a multi Skolem sequence exists with positions P and differences A, then G has a multiple choice matching.*

PROOF. Consider a generalized multi Skolem sequence which is generated from (P, A). It is easy to see that exactly $(m - n)$ of the differences added to A in Step M2.1 must be fitted in the *garbage component*, more specifically in the *garbage II* component (because no other differences in A can be put in the start position of the *garbage II* component). The remaining n differences added to A in Step M2.1 must be fitted in the *matching component*, because they simply do not fit anywhere else.

Lemma 2 says that if the difference in $A_{M2.1.k}$ is fitted with one copy at the i-th leftmost position in *match pos* and the difference in $A_{M2.2.k}$ is fitted with one copy at the j-th leftmost position in *match pos*, then (i, j) is an edge (labeled k) in G. So for all such elements in the sequence, let the edge $(i, j) \in M$. Then M will clearly be a multiple choice matching on G. □

Now what remains to be proved is that if G has a multiple choice matching M, then (P, A) is a generalized multi Skolem pair.

Lemma 4 *If G has a multiple choice matching M, then a generalized multi Skolem sequence exists with positions P and differences A.*

PROOF. Remember that every difference added to A in Step M2.4 corresponds to a specific edge (i, j) in G. If this edge $(i, j) \notin M$, then fit the difference in $A_{M2.4.k.(i,j)}$ in the *matching component* in the positions of $P_{M2.3.k.(i,j)}$. Note that this is where we make use of the fact that no two edges in G of the same edge label share a common vertex. If this property did not hold we could not guarantee that this difference could be fitted in the *matching component* at all. For all k such that M contains an edge (i, j) labeled k, fit the difference in $A_{M2.1.k}$ with one copy at the i-th leftmost position in *match pos*, and fit the difference in $A_{M2.2.k}$ with one copy at the j-th leftmost position in *match pos.* Now we have filled all positions in P in the *matching component.* But some of the differences added to A in Step M2.1, Step M2.2 and Step M2.4 have not been fitted yet. More specifically the differences that remain to be fitted are $m-n$ of the differences added to A in Step M2.1, $m-n$ of the differences added to A in Step M2.2 and n of the differences added to A in Step M2.4. These superfluous differences must be fitted in the *garbage component*. How this is done is explained in Step G1-3 below.

Step G1: Here we make sure that the n superfluous differences added to A in Step M2.4 can be fitted in the *garbage component*. We know that there will be exactly n of them because the differences added to A in Step M2.4 are fitted in the *matching component* if and only if the corresponding edge in G is not part of the multiple choice matching. It can also be deduced that an even number of these n differences are even. We formulate this as a claim.

Claim: *An even number of the n superfluous differences added to A in Step M2.4 are even.*

The difference in $A_{M2.4.k.(i,j)}$ is of the form $(2x+j)-(2y+i)$ which has the same parity as $j-i$. The n superfluous differences added to A in Step M2.4 are exactly those that were added to A for the n edges in the multiple choice matching. The vertices in G are labeled with $1, \ldots, 2n$, i.e., an equal number of even and odd vertices. An edge (i,j) where $i-j$ is odd connects two vertices of different parity, an edge (i,j) where $i-j$ is even connects two vertices of the same parity. Thus for these n edges to be a multiple choice matching it is necessary that the number of edges (i,j) with $i-j$ even, is even. It follows that an even number of the n superfluous differences added to A in Step M2.4 are even.

In Step G1.1 we add all positions in *start* and *absorb* in each of the n identical blocks in *garbage I* to P. In Step G1.2 we add $n^2m^2+2n^2m+m^2+m-3n$ ones and $2n$ twos to A. What we want to show is that the n superfluous differences added to A in Step M2.4 and the differences added to A in Step G1.2 can always be fitted in the *garbage I* component. First note that the number of positions in P in the *garbage I* component is exactly twice the number of differences from A that we want to fit here, thus all positions in P will be filled. Place one difference of the n superfluous differences added to A in Step M2.4 in each of the n blocks in *garbage I*, such that the first copy is placed at the *start* position and the second copy is placed somewhere in *absorb*. If the difference that was placed is even, place a two at the two positions that are neighbors to the position in *absorb* where the second copy was placed. This guarantees that the positions in P that remain to be filled, are divided into blocks of consecutive positions of even length. These positions can clearly be filled by placing ones at these, but we have too few ones, and some twos also remain to be placed. Remember that at the beginning we had $2n$ twos to place and we have placed an even number of these, thus an even number of twos remain to be placed. This fact and the observation that 1111 can be replaced by 2222 makes sure that we can fill these remaining positions in P by using the remaining differences that were added to A in Step G1.2. Thus we can conclude that the n superfluous differences added to A in Step M2.4 and the differences added to A in Step G1.2 can always be fitted in the *garbage I* component, while leaving no positions in P in the *garbage I* component unfilled.

Step G2: The $(m - n)$ superfluous differences added to A in Step M2.1 are placed in *garbage II* with one copy at *start* position and the other copy in *absorb*. Place a two on each side of the second copy in *absorb*. Since each difference in $A_{M2.2.k}$ is even, there are an even number of free positions in *absorb* on each side of the two two's that were placed there. Fill these positions with ones. A simple count shows that all of the differences added in step G.2.2 have now been used.

Step G3: Clearly the $m - n$ superfluous differences added to A in Step M2.2 can be packed into *garbage III* in the same way that *garbage II* was filled, and this completes the construction of the generalized multi Skolem sequence. □

Lemma 3 and Lemma 4 prove that the constructed (P, A) is a generalized multi Skolem pair if and only if G has a multiple choice matching. It is easy to see that the reduction can be done in polynomial time. Thus, we have proved that Generalized Multi Skolem Sequences is $\mathcal{NP}$-complete. An actual example of the reduction in action might have been nice to enhance the understanding of the reduction. But unfortunately even a small graph with only six vertices and four edges is reduced to a sequence of 6224 positions.

3 Conclusions

The $\mathcal{NP}$-completeness result of Generalized Multi Skolem Sequences (Theorem 1) can be seen as an upper bound on how far the generalizations of Skolem sequences can be taken while still hoping to resolve the existence question. Future research could be aimed at trying to lower this upper bound by proving the $\mathcal{NP}$-completeness of generalized Skolem sequences.

4 Acknowledgements

The author would like to thank Svante Linusson for proposing this interesting problem and Daniel Hedberg, Peter Jonsson, Václav Linek and Nabil Shalaby for valuable remarks and careful reading of earlier versions of the manuscript.

A special thank goes to the referee for valuable comments that improved the presentation in many ways.

References

[1] C.A. Baker. Extended Skolem sequences. *J. Combin. Des*, 3(5):363–379, 1995.

[2] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.

[3] V. Linek and Z. Jiang. Extended Langford sequences with small defects. *J. Combin. Theory Ser. A*, 84(1):38–54, 1998.

[4] V. Linek and N. Shalaby. Rosa sequences. Manuscript.

[5] N. Shalaby. *The CRC Handbook of Combinatorial Designs*, chapter Skolem Sequences. CRC Press, 1995.

[6] T. Skolem. On certain distributions of integers in pairs with given differences. *Math. Scand*, 5:57–68, 1957.