A NOTE ON THE HARDNESS OF SKOLEM-TYPE SEQUENCES

GUSTAV NORDH

Abstract. The purpose of this note is to give upper bounds (assuming **P** different from **NP**) on how far the generalizations of Skolem sequences can be taken while still hoping to resolve the existence question. We prove that the existence questions for both multi Skolem sequences and generalized Skolem sequences are strongly **NP**-complete. These results are significant strengthenings and simplifications of the recent **NP**-completeness result for generalized multi Skolem sequences.

1. Introduction

Skolem sequences originates from the work by Thoralf Skolem [11] on the construction of Steiner triple systems. He asked when you could partition the set $P = \{1, \ldots, 2n\}$ in n pairs (p_i, p'_i) such that the set of differences $A = \{p_i - p'_i \mid i = 1, \ldots, n\} = \{1, \ldots, n\}$. Later, this problem was reformulated into the (equivalent) problem of deciding which sets $A = \{1, \ldots, n\}$ that can generate a sequence of length $2n$ with two copies of every element k in A so that the two copies of k are placed k places apart in the sequence. For example, the set $\{1, 2, 3, 4\}$ can generate the sequence 42324311, but the set $\{1, 2, 3\}$ cannot generate such a sequence. Skolem [11] solved the existence question for Skolem sequences by proving that $A = \{1, \ldots, n\}$ can generate a Skolem sequence if and only if $n \equiv 0, 1 \pmod 4$.

Skolem sequences have been generalized in mainly two directions. The first is to let P, the set of positions that elements in the sequence are allowed to occupy be a set of $2n$ positive integers $\{p_1, \ldots, p_{2n}\}$, cf. [1, 6, 7]. Such sequences are called generalized Skolem sequences and the existence question is the problem of deciding which sets $P = \{p_1, \ldots, p_{2n}\}$ that can be partitioned into the differences in $A = \{1, \ldots, n\}$. The second line of generalization is to let the set of differences A be a multiset of positive integers. Such sequences are called multi Skolem sequences, cf. [2, 9], and the corresponding existence question is the problem of deciding for which multisets $A = \{a_1, \ldots, a_n\}$ there is a partition of $\{1, \ldots, 2n\}$ into the differences in A.

If these two generalizations are allowed simultaneously, we get the concept of generalized multi Skolem sequences. The existence question for generalized multi Skolem sequences is then the problem of deciding whether a set $P = \{p_1, \ldots, p_{2n}\}$ of $2n$ positive integers (the positions) can be partitioned into the multiset of differences $A = \{a_1, \ldots, a_n\}$. For example, the answer for $P = \{1, 2, 4, 5, 7, 8\}$ and $A = \{1, 6, 6\}$ is positive since $\{5 - 4, 7 - 1, 8 - 2\} = \{1, 6, 6\}$, and the corresponding generalized multi Skolem sequence is 66_11_66. The existence question for generalized multi Skolem sequences was proved to be **NP**-complete in [8] by a rather complicated reduction.

In this note we strengthen the hardness result in [8] by showing that the existence questions for both generalized Skolem sequences and multi Skolem sequences are **NP**-complete in the strong sense. Although these results are easy consequences of the recent **NP**-completeness results in [5] and [12], respectively, they are interesting since they give strong upper bounds (assuming $\mathbf{P} \neq \mathbf{NP}$) on how far the generalizations of Skolem sequences can be taken while still being able to resolve the existence question in a "nice" way (where the precise meaning of "nice" is "by conditions verifiable in polynomial-time"). In particular, any polynomial-time verifiable conditions for the existence of generalized Skolem sequences and multi Skolem sequences (e.g., those presented in [7, 9]) must fail on an infinite number of instances (assuming $\mathbf{P} \neq \mathbf{NP}$). Moreover, the **NP**-completeness result for multi Skolem sequences is surprising in light of the following conjecture.

Conjecture 1 ([9]). *$P = \{1, \ldots, 2n\}$ can be partitioned into the differences in $A = \{a_1, a_2, \ldots, a_n\}$ with $a_1 < a_2 < \cdots < a_n$ if and only if the number of even a_i's is even, and $\sum_{i=m}^{n} a_i \leq n^2 - (m-1)^2$ for each $1 \leq m \leq n$.*

For further information on Skolem-type sequences and their applications we refer the reader to [10], and for background on the theory of **NP**-completeness we point the reader to [3].

2. **NP**-completeness of generalized Skolem sequences

The existence question for generalized Skolem sequences can be formulated as follows.

Generalized Skolem Sequences:
Instance: A set P of $2n$ positive integers.
Question: Can P be partitioned into the differences in $A = \{1, \ldots, n\}$?

We show that this problem is **NP**-complete in the strong sense by giving a simple reduction from a special case of Numerical Matching with Target Sums where all integers are distinct (denoted DNMTS) that was recently proved to be strongly **NP**-complete [5].

Distinct Numerical Matching with Target Sums (DNMTS):
Instance: Three sets $A = \{a_1, \ldots, a_n\}, B = \{b_1, \ldots, b_n\}, C = \{c_1, \ldots, c_n\}$ of pairwise distinct positive integers such that $\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} c_i$.
Question: Can the set $A \cup B \cup C$ be partitioned into n triples (a_i, b_i, c_i), $i = 1, \ldots, n$ such that $a_i + b_i = c_i$?

By analysing the proof from [5] (more specifically, the proof of Lemma 5 and the comment after Corollary 8) it is easy to verify that the problem is still **NP**-complete when restricted to instances satisfying $\max(B) < \min(C)$. Given an instance $A = \{a_1, \ldots, a_n\}, B = \{b_1, \ldots, b_n\}, C = \{c_1, \ldots, c_n\}$ of DNMTS we reduce it to the question of deciding whether the set $P = B \cup C$ can be partitioned into the differences in A. We observe that the conditions $\max(B) < \min(C)$ and $\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} c_i$ implies that any such partition $p_i - p'_i = a_i$ $(i = 1, \ldots, n)$ of $P = B \cup C$ into the differences in A must have $p_i \in C$ and $p'_i \in B$. This is because otherwise $\sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p'_i < \sum_{i=1}^{n} a_i$, which contradicts $p_i - p'_i = a_i$ $(i = 1, \ldots, n)$.

If the answer to the DNMTS instance is yes, then the partition into triples (a_i, b_i, c_i) with $a_i + b_i = c_i$ $(i = 1, \ldots, n)$ gives a partition of P into the differences in A by $c_i - b_i = a_i$ $(i = 1, \ldots, n)$. Similarly, if $P = B \cup C$ can be partitioned into the differences in A by a partition $p_i - p'_i = a_i$ $(i = 1, \ldots, n)$, then as observed above $p_i \in C$ and $p'_i \in B$, and thus $a_i + p'_i = p_i$ $(i = 1, \ldots, n)$ is a solution to the instance $A = \{a_1, \ldots, a_n\}, B = \{b_1, \ldots, b_n\}, C = \{c_1, \ldots, c_n\}$ of DNMTS.

To complete the reduction to Generalized Skolem Sequences we construct a set P' that can be partitioned into the differences in $A' = \{1, \ldots, \max(A)\}$ if and only if P can be partitioned into the differences in A. Let $k = \max(A)$ and define $Q = \{q_1, \ldots, q_{k-n}\} = \{\max(P) + 2i \cdot k \mid 1 \leq i \leq k-n\}$, $\{r_1, \ldots, r_{k-n}\} = A' \setminus A$, and finally $P' = P \cup Q \cup \{q_i + r_i \mid 1 \leq i \leq k-n\}$.

3. **NP**-completeness of multi Skolem sequences

The existence question for multi Skolem sequences can be formulated as follows.

Multi Skolem Sequences:
Instance: A multiset A of n positive integers.
Question: Can $P = \{1, \ldots, 2n\}$ be partitioned into the differences in A?

The fact that this problem is strongly **NP**-complete follows immediately from a rather sophisticated **NP**-hardness proof, due to Yu et al. [12], for the problem of minimizing the makespan for two machine (flow shop) coupled task scheduling with two operations per task, unit processing times, and exact delays between operations (in Graham et al.'s [4] three field notation $F2|l_j, p_{ij} = 1|C_{max}$). In this scheduling problem there are two machines, M_1 and M_2 available from time 0 and onwards and each machine is capable of performing at most one operation at a time. Each task consists of two operations, the first must be processed by M_1, the second by M_2, and the processing times of all operations are 1. For each task t_i there is a (non-negative) delay of exactly a_i time units between the completion of the first operation on M_1 and the start of the second operation on M_2. A schedule σ is a specification of the completion time $\sigma(t_i)$ of each task t_i on machine M_2 such that all conditions above are met.

$F2|l_j, p_{ij} = 1|C_{max}$ (denoted F2UD):
Instance: A set of n tasks represented by the multiset $L = \{l_1, \ldots, l_n\}$ where $l_i \in \mathbb{N}$ is the delay between the completion time of the first operation of task t_i on M_1 and the start time of the second operation of task t_i on M_2.
Objective: Minimize the makespan (i.e., find a schedule such that the completion time of the last job on M_2 is minimal over all schedules).

Yu et al. [12] prove that the decision variant of F2UD is strongly **NP**-complete, that is, they prove that the problem of deciding whether there exists a schedule with makespan less than or equal to y is strongly **NP**-complete. Moreover, their proof shows that the problem is strongly **NP**-complete if $y = 2n$ (where n is the number of tasks) and the schedules are further restricted so that the completion time of the last operation on M_1 must be n and the start time of the first operation on M_2 must be n. We denote this restricted decision variant of F2UD by R-F2UD. Now, the reduction to Multi Skolem Sequences is trivial. Given an instance $L = \{l_1, \ldots, l_n\}$ of R-F2UD, then $A = \{l_1 + 1, \ldots, l_n + 1\}$ is the corresponding instance of Multi Skolem Sequences. If there is a schedule σ such that the answer to the R-F2UD instance is "yes", then $(\sigma(t_i) - (l_i + 1), \sigma(t_i))$ $(i = 1, \ldots, n)$ is a partition of $\{1, \ldots, 2n\}$ into the differences in $A = \{l_1 + 1, \ldots, l_n + 1\}$ by $\sigma(t_i) - (\sigma(t_i) - (l_i + 1)) = l_i + 1$ $(i = 1, \ldots, n)$. Similarly, if $\{1, \ldots, 2n\}$ can be partitioned into the differences in $A = \{l_1 + 1, \ldots, l_n + 1\}$ by $p_i - p'_i = l_i + 1$ $(i = 1, \ldots, n)$, then the schedule $\sigma(t_i) = p'_i$ $(i = 1, \ldots, n)$ shows that $L = \{l_1, \ldots, l_n\}$ is a "yes" instance of R-F2UD.

References

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

[2] C.A. Baker, R.J. Nowakowski, N. Shalaby, and A. Sharary. m-fold and extended m-fold Skolem sequences. *Utilitas Math.*, 45:153–167, 1994.

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

[4] R.L. Graham, E.L. Lawler, J.K. Lenstra, and A.H.G. Rinnooy Kan. Optimization and approximation in deterministic sequencing and scheduling: A survey. *Ann. Discrete Math*, 5:297–326, 1979.

[5] H. Hulett, T.G. Will, and G.J. Woeginger. Multigraph realizations of degree sequences: Maximization is easy, minimization is hard. *Operations Research Letters*, 36(5):594–596, 2008.

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

[7] V. Linek and N. Shalaby. The existence of (p,q)-extended Rosa sequences. *Discrete Math.*, 308(9):1583–1602, 2008.

[8] G. Nordh. NP-completeness of generalized multi-Skolem sequences. *Discrete Applied Math.*, 155(16):2061–2068, 2007.

[9] G. Nordh. Perfect Skolem sets. *Discrete Math.*, 308(9):1653–1664, 2008.

[10] N. Shalaby. *Handbook of Combinatorial Designs, Second Edition*, chapter Skolem and Langford Sequences, pages 612–617. CRC Press, 2006.

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

[12] W. Yu, H. Hoogeveen, and J.K. Lenstra. Minimizing makespan in a two-machine flow shop with delays and unit-time operations is NP-hard. *J. Scheduling*, 7(5):333–348, 2004.

Department of Computer Science, Linköpings Universitet, S-581 83 Linköping, Sweden
E-mail address: gusno@ida.liu.se