Information Processing Letters 34 (1990) 293–298
North-Holland

28 May 1990

COVERING A SET WITH ARITHMETIC PROGRESSIONS IS NP-COMPLETE

Lenwood S. HEATH

Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Communicated by D.A. Plaisted
Received August 1989
Revised December 1989

Keywords: Computational complexity, NP-complete, set covering, arithmetic progressions

1. Introduction

An important class of combinatorial problems consists of set covering and partitioning problems. Many such problems are NP-complete. A typical problem is the MINIMUM COVER problem [2].

MINIMUM COVER

Instance: Collection C of subsets of a finite set S, positive integer $K \leqslant |C|$.

Question: Does C contain a cover for S of size K or less, i.e., a subcollection $C' \subset C$ with $|C'| \leqslant K$ such that every element of S belongs to at least one member of C'?

In an instance of MINIMUM COVER, the subsets of S are listed, and S itself is implicitly represented as the union of these subsets. For our problems, the representation of the instance is in some sense the reverse. The instance consists of a set of integers. The subsets are implicitly derived from the set; in particular, the subsets are the finite arithmetic progressions contained in the set.

We define two new problems for this class of instances. The first is the set covering problem that is the analog of MINIMUM COVER. The second is an exact covering problem where the cover must also be a partition of the set. Our result is that both problems are NP-complete.

Arithmetic progressions in finite sets of integers have been studied by mathematicians for many years (see, for example, [1,5,6]). The primary purpose of this study has been to estimate $r_k(n)$, the greatest integer such that some subset of $\{1, \ldots, n\}$ of cardinality $r_k(n)$ contains no arithmetic progression of length k.

One practical motivation for our problems appears in Grobman and Studwell [3] in a two-dimensional version. A programmed electron-beam machine places patterns at discrete locations (points) on a mask used in the manufacture of VLSI chips. The pattern points are given as a set of integral grid coordinates. An $m \times n$ rectilinear grid is a set of mn pattern points arranged regularly to form both m rows of n points each and n columns of m points each. The electron-beam machine is able to process the set significantly faster when its input is encoded as a set of rectilinear grids that cover the pattern set. The encoding problem is then to cover the pattern set with as few rectilinear grids as possible. Grobman and Studwell give an heuristic for this two-dimensional encoding problem. The corresponding one-dimensional problem is our EXACT COVER BY ARITHMETIC PROGRESSIONS. Our NP-completeness results imply the NP-completeness of the two-dimensional problem of Grobman and Studwell.

0020-0190/90/$3.50 © 1990 – Elsevier Science Publishers B.V. (North-Holland)

2. Definitions

Let X be a finite set of positive integers. Let $Y = \{Y_i \mid 1 \leqslant i \leqslant M\}$ be a collection of subsets of X such that $X = \bigcup_{i=1}^{M} Y_i$. A subcollection $\{Y_{i_j} \mid 1 \leqslant j \leqslant K\}$ chosen from Y is a *K-cover* for X (or simply a *cover*) if $X = \bigcup_{j=1}^{K} Y_{i_j}$. The subcollection is an *exact K-cover* (or simply an *exact cover*) for X if it is a *K-cover* and $Y_{i_j} \cap Y_{i_k} = \emptyset$ whenever $i_j \neq i_k$.

An *arithmetic progression (AP) of length k* is a set of integers of the form: $\{a + bi \mid 0 \leqslant i < k\}$ where $a \geqslant 1$ and $b \geqslant 1$ are constants; b is the *increment* of the AP. An arithmetic progression of length k is called a *k-AP*. If $x \leqslant y$, let $[x, y]$ denote the $(y - x + 1)$-AP $\{x, x+1, \ldots, y\}$.

The known NP-complete problem that will be used in the reduction of the next section is EXACT COVER BY 3-SETS [2].

EXACT COVER BY 3-SETS (X3C)

Instance: Set X with $|X| = 3q$ and a collection Y of 3-element subsets of X.

Question: Does Y contain an exact cover for X, i.e., a subcollection $Y' \subset Y$ such that every element of X occurs in exactly one member of Y'?

X3C was shown to be NP-complete by Karp [4]. Note that an exact cover Y' for X has cardinality $|Y'| = q$.

We now define the decision problem versions of our two set covering problems.

EXACT COVER BY ARITHMETIC PROGRESSIONS (XAP)

Instance: Set S of integers, positive integer $J \leqslant |S|$.

Question: Is there an exact J'-cover of S, $J' \leqslant J$, taken from the collection

$$Z = \{Z_i \subset S \mid Z_i \text{ is an AP}\}?$$

COVER BY ARITHMETIC PROGRESSIONS (CAP)

Instance: Same as XAP.

Question: Is there a J'-cover of S, $J' \leqslant J$, taken from the collection

$$Z = \{Z_i \subset S \mid Z_i \text{ is an AP}\}?$$

Note that the collection Z is not explicitly part of the instance but is defined implicitly as all arithmetic progression in S. It is easy to see that both XAP and CAP are in NP (for definitions of NP and NP-complete, see [2]). A nondeterministic Turing machine can guess a collection of subsets of S and check in polynomial time that each subset is a k-AP, that the size of the collection is no greater than J, and that the collection is an exact cover (or cover) for S. The reduction of an instance of X3C to an instance of XAP or CAP is given in the next section, where several properties of the reduction are presented as lemmas. The NP-completeness of XAP and CAP is proved in Section 4.

3. The reduction

Let X and Y constitute an instance of X3C such that $|X| = 3q$. We may assume that $X = \{1, 2, \ldots, 3q\}$ (otherwise, first map X to $\{1, 2, \ldots, 3q\}$). This instance will be reduced to an instance S, J of either XAP or CAP. Let $|Y| = M$. Let the bound on the number of APs in a cover be $J = q + 4M$. The S we construct contains no AP of length greater than 3, and $|S| = 3(q + 4M)$. Thus any cover of S must consist of exactly J 3-APs; in particular, we need not consider 2-APs. We construct inductively a sequence of sets $S_0, S_1, \ldots, S_M$ such that $S_0 \subset S_1 \subset \cdots \subset S_M = S$. S_0 represents X, and $S_j - S_{j-1}$ represents $Y_j \in Y$.

Along with each set S_j, we define an integer N_j which is always the largest element of S_j. Initially define

$$S_0 = \{e_i \mid e_i = 3 \times 4^i,\ 1 \leqslant i \leqslant 3q\},$$

$$N_0 = 3 \times 4^{3q}.$$

The integer $e_i \in S_0$ represents $i \in X$. Note that the elements of S_0 encode the elements of X so that no three elements of S_0 form a 3-AP.

Now assume that S_{j-1} and N_{j-1} have been defined, $0 < j \leqslant M$. Let $Y_j = \{r, s, t\}$, $r < s < t$, be the jth subset of X in Y. Let $a_j = e_r$, $b_j = e_s$ and $c_j = e_t$. Note that these are the three elements of S_0 that represent r, s, and t. Define twelve integers to represent Y_j in S:

$a'_j = a_j + 600N_{j-1}$, $a''_j = a_j + 1200N_{j-1}$,

$b'_j = b_j + 606N_{j-1}$, $b''_j = b_j + 1212N_{j-1}$,

$c'_j = c_j + 612N_{j-1}$, $c''_j = c_j + 1224N_{j-1}$,

$\alpha_j = a_j + 900N_{j-1}$,

$\beta_j = b_j + 909N_{j-1}$,

$\gamma_j = c_j + 918N_{j-1}$,

$\delta_j = 2\beta_j - \alpha_j = 2b_j - a_j + 918N_{j-1}$,

$\epsilon_j = \frac{4}{3}\delta_j - \frac{1}{3}\gamma_j = \frac{1}{3}(8b_j - 4a_j - c_j) + 918N_{j-1}$,

$\zeta_j = \frac{2}{3}\delta_j + \frac{1}{3}\gamma_j = \frac{1}{3}(4b_j - 2a_j + c_j) + 918N_{j-1}$.

Now define S_j and N_j:

$T_j = \{a'_j, b'_j, c'_j\}$.

$U_j = \{\alpha_j, \beta_j, \gamma_j, \delta_j, \epsilon_j, \zeta_j\}$,

$V_j = \{a''_j, b''_j, c''_j\}$,

$S_j = S_{j-1} \cup T_j \cup U_j \cup V_j$,

$N_j = c''_j$.

Finally, define $S = S_M$.

That S is well defined is easy to show by induction. Clearly, S can be obtained in polynomial time. The following lemmas are easy to obtain.

Lemma 3.1. $|S| = 3q + 12M$.

Lemma 3.2. $S_j \subset [0, N_j]$ *and hence if* $x, y \in S_j$, *then* $|x - y| \leqslant N_j$.

Lemma 3.3. $(S_j - S_{j-1}) \cap [0, N_{j-1}] = \emptyset$, *i.e.*, $(S_j - S_{j-1})$ *is totally "to the right" of* S_{j-1}.

We now determine the structure of the arithmetic progressions of S. In particular, we aim to show that 3-APs of a certain form exist and no others (Lemma 3.6). To begin, the following lemma gives the order of the elements of U_j.

Lemma 3.4. *For* $1 \leqslant j \leqslant M$,

$900N_{j-1} < \alpha_j \leqslant 901N_{j-1}$,

$909N_{j-1} < \beta_j \leqslant 910N_{j-1}$,

$917N_{j-1} < \epsilon_j < \delta_j < \zeta_j < \gamma_j \leqslant 919N_{j-1}$.

Proof. $900N_{j-1} < \alpha_j \leqslant 901N_{j-1} < 909N_{j-1} < \beta_j \leqslant 910N_{j-1} < 918N_{j-1} < \gamma_j \leqslant 919N_{j-1}$ follows from $0 < a_j, b_j, c_j \leqslant N_{j-1}$ and the definitions of α_j, β_j and γ_j.

Expanding the definition of δ_j, we obtain

$$\begin{aligned}\delta_j &= 2b_j - a_j + 918N_{j-1}\\ &= 2 \times 3 \times 4^s - 3 \times 4^r + 918N_{j-1}\\ &> 918N_{j-1}\end{aligned}$$

since $s > r$. Also,

$$\begin{aligned}\delta_j &< 3 \times 4^{s+1} - 3 \times 4^r + 918N_{j-1}\\ &\leqslant 3 \times 4^t + 918N_{j-1}\\ &= \gamma_j.\end{aligned}$$

Hence, $918N_{j-1} < \delta_j < \gamma_j$.

Expanding the definition of ϵ_j, we obtain

$$\begin{aligned}\epsilon_j &= \tfrac{4}{3}\delta_j - \tfrac{1}{3}\gamma_j\\ &< \tfrac{4}{3}\delta_j - \tfrac{1}{3}\delta_j\\ &= \delta_j.\end{aligned}$$

Also,

$$\begin{aligned}\epsilon_j &= \tfrac{1}{3}(8b_j - 4a_j - c_j) + 918N_{j-1}\\ &= 8 \times 4^s - 4 \times 4^r - 4^t + 918N_{j-1}\\ &> -4^t + 918N_{j-1}\\ &> 917N_{j-1}.\end{aligned}$$

Hence, $917N_{j-1} < \epsilon_j < \delta_j$.

Since ζ_j is a convex combination of δ_j and γ_j, $\delta_j < \zeta_j < \gamma_j$. □

The following lemma states that there are large gaps between any two of the sets S_{j-1}, T_j, U_j and V_j. Such gaps will make 3-APs of certain forms impossible.

Lemma 3.5. *For* $1 \leqslant j \leqslant M$, $S_{j-1} \subset [1, N_{j-1}]$, $T_j \subset [600N_{j-1}, 613N_{j-1}]$, $U_j \subset [900N_{j-1}, 919N_{j-1}]$, *and* $V_j \subset [1200N_{j-1}, 1225N_{j-1}]$.

Proof. $S_{j-1} \subset [1, N_{j-1}]$ follows from Lemma 3.2. $T_j \subset [600N_{j-1}, 613N_{j-1}]$ follows from $0 < a_j < b_j < c_j \leqslant N_{j-1}$ and from the definitions of a'_j, b'_j, c'_j.

$V_j \subset [1200N_{j-1},\ 1225N_{j-1}]$ follows by a similar argument. $U_j \subset [900N_{j-1},\ 919N_{j-1}]$ follows from Lemma 3.4. □

To summarize, we have shown that the elements of $S_j - S_{j-1}$ occur in the order

$a'_j,\ b'_j,\ c'_j,\ \alpha_j,\ \beta_j,\ \epsilon_j,\ \delta_j,\ \zeta_j,\ \gamma_j,\ a''_j,\ b''_j,\ c''_j.$

We are now prepared to define 3-sets of S that are 3-APs. For $1 \leqslant j \leqslant M$, define

$$A_j = \{a_j,\ a'_j,\ a''_j\}, \quad B_j = \{b_j,\ b'_j,\ b''_j\},$$
$$C_j = \{c_j,\ c'_j,\ c''_j\},$$
$$A'_j = \{a'_j,\ \alpha_j,\ a''_j\}, \quad B'_j = \{b'_j,\ \beta_j,\ b''_j\},$$
$$C'_j = \{c'_j,\ \gamma_j,\ c''_j\},$$
$$D_j = \{\alpha_j,\ \beta_j,\ \delta_j\}, \quad E_j = \{\epsilon_j,\ \zeta_j,\ \gamma_j\},$$
$$F_j = \{\epsilon_j,\ \delta_j,\ \zeta_j\}.$$

The main result of this section is the following lemma.

Lemma 3.6. *The only 3-APs in S are A_j, B_j, C_j, A'_j, B'_j, C'_j, D_j, E_j, and F_j, $1 \leqslant j \leqslant M$.*

Proof. It is easy to check that each of the given sets is a 3-AP. The proof that only these sets are 3-APs is by contradiction. Assume that S contains a 3-AP $P = \{x,\ y,\ z\}$ such that $x < y < z$ and such that P is not one of the given sets.

Let j be smallest such that $z \in S_j$. As already noted, no three elements of S_0 constitute a 3-AP. Thus $j \neq 0$. From the bounds in Lemma 3.5, it is clear that the only possibilities for P are

(1) $P = T_j$;
(2) $P = V_j$;
(3) $x \in T_j,\ y \in U_j,\ z \in V_j$;
(4) $x \in S_{j-1},\ y \in T_j,\ z \in V_j$;
(5) $P \subset U_j$.

We now derive contradictions for each of these possibilities.

(1) $P = T_j$. Then $x = a'_j$, $y = b'_j$, $z = c'_j$ and $y = \frac{1}{2}(x + z)$. Expanding x and z, we obtain

$$\begin{aligned} y &= \tfrac{1}{2}(x + z) \\ &= \tfrac{1}{2}(a_j + 600N_{j-1} + c_j + 612N_{j-1}) \\ &= \tfrac{1}{2}(a_j + c_j) + 606N_{j-1} \\ &= b_j + 606N_{j-1} \end{aligned}$$

which implies $b_j = \frac{1}{2}(a_j + c_j)$. But this is a contradiction to the fact that $\{a_j,\ b_j,\ c_j\}$ does not constitute a 3-AP.

(2) $P = V_j$. The contradiction is the same as for (1) $P = T_j$.

(3) $x \in T_j,\ y \in U_j,\ z \in V_j$. Since A'_j, B'_j and C'_j are 3-APs meeting this condition, we need not consider them here. There remain 6 subcases:

(i) $x = a'_j$, $z = b''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(a'_j + b''_j) \\ &= \tfrac{1}{2}(a_j + 600N_{j-1} + b_j + 1212N_{j-1}) \\ &= \tfrac{1}{2}(a_j + b_j + 1812N_{j-1}) \\ &\in [906N_{j-1},\ 907N_{j-1}]. \end{aligned}$$

(ii) $x = a'_j$, $z = c''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(a_j + 600N_{j-1} + c_j + 1224N_{j-1}) \\ &\in [912N_{j-1},\ 913N_{j-1}]. \end{aligned}$$

(iii) $x = b'_j$, $z = a''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(b_j + 606N_{j-1} + a_j + 1200N_{j-1}) \\ &\in [903N_{j-1},\ 904N_{j-1}]. \end{aligned}$$

(iv) $x = b'_j$, $z = c''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(b_j + 606N_{j-1} + c_j + 1224N_{j-1}) \\ &\in [915N_{j-1},\ 916N_{j-1}]. \end{aligned}$$

(v) $x = c'_j$, $z = a''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(c_j + 612N_{j-1} + a_j + 1200N_{j-1}) \\ &\in [906N_{j-1},\ 907N_{j-1}]. \end{aligned}$$

(vi) $x = c'_j$, $z = b''_j$. Then,

$$\begin{aligned} y &= \tfrac{1}{2}(c_j + 612N_{j-1} + b_j + 1212N_{j-1}) \\ &\in [912N_{j-1},\ 913N_{j-1}]. \end{aligned}$$

For each subcase, the inequalities of Lemma 3.4 show that $y \notin U_j$, a contradiction to $y \in U_j$.

(4) $x \in S_{j-1},\ y \in T_j,\ z \in V_j$. Consider the following bounds:

$$x \in [0,\ N_{j-1}],$$
$$a'_j \in [600N_{j-1},\ 601N_{j-1}],$$

$b_j' \in [606N_{j-1}, 607N_{j-1}]$,

$c_j' \in [612N_{j-1}, 613N_{j-1}]$,

$a_j'' \in [1200N_{j-1}, 1201N_{j-1}]$,

$b_j'' \in [1212N_{j-1}, 1213N_{j-1}]$,

$c_j'' \in [1224N_{j-1}, 1225N_{j-1}]$.

If $z = c_j''$, then $y = \frac{1}{2}(x+z)$ and $612N_{j-1} \leqslant y \leqslant 613N_{j-1}$, hence $y = c_j'$, $x = c_j$ and $P = C_j$. Similarly, if $z = b_j''$, then $P = B_j$, and if $z = a_j''$, then $P = A_j$. All cases are contradictions to P not being one of the 3-APs named in the lemma.

(5) $P \subset U_j$. From the bounds in Lemma 3.4 and from $P \neq D_j$, we can further assume that x, y, $z \in \{\epsilon_j, \delta_j, \zeta_j, \gamma_j\}$. From the inequalities of Lemma 3.4, x and y must be chosen from $\{\epsilon_j, \delta_j, \zeta_j\}$, and the only choice not covered by E_j or F_j is $x = \delta_j$, $y = \zeta_j$. From the order properties of Lemma 3.4, $z = \gamma_j$. But $\zeta_j = \frac{2}{3}\delta_j + \frac{1}{3}\gamma_j \neq \frac{1}{2}(\delta_j + \gamma_j)$.

For all possible values of P, there is a contradiction. Thus the only 3-APs in S are those given in the statement of the lemma. □

4. NP-completeness

Theorem 4.1. X3C *reduces to* XAP.

Proof. Suppose Y' is an exact cover for X. We construct an exact cover by APs $Z' \subset Z$. For each $Y_j \in Y'$, put A_j, B_j, C_j, D_j and E_j in Z'. For each $Y_j \notin Y'$, put A_j', B_j', C_j' and F_j in Z'. It is easy to check that Z' covers S and that no element of S is in two elements of Z'. The cardinality of Z' is $q + 4M = J$. Hence, Z' is an exact cover by J APs for S.

Now suppose that Z' is an exact cover by J APs for S. By Lemma 3.6, we need only consider 3-APs, since S contains no 4-APs. Let $Y' = \{Y_j \mid E_j \in Z'\}$. If $E_j \in Z'$, then C_j', $F_j \notin Z'$, since Z' is a partition of X. But then D_j must be in Y' for δ_j to be covered. $D_j \in Z'$ implies A_j', $B_j' \notin Z'$. Thus, A_j, B_j, $C_j \in Z'$. If $E_j \notin Z'$, C_j', $F_j \in Z'$. Thus, $D_j \notin Z'$ and A_j', $B_j' \in Z'$. A_j, B_j, $C_j \notin Z'$. Thus A_j, B_j, $C_j \in Z'$ if and only if $E_j \in Z'$. Now, A_j, B_j, $C_j \in Z'$ corresponds to covering the three integers in S_0 that encode the three elements of Y_j. Thus $\bigcup_{Y_j \in Y'} Y_j = X$. Y' is an exact cover of X since Z' is an exact cover by J APs of S. □

Corollary 4.2. XAP *is* NP-*complete*. XAP *restricted to instances where S contains no AP longer than* 3 *is* NP-*complete*.

Theorem 4.3. X3C *reduces to* CAP.

Proof. Same as Theorem 4.1 since a cover for S of size J by 3-APs cannot cover any element more than once and hence is also an exact cover. □

Corollary 4.4. CAP *is* NP-*complete*. CAP *restricted to instances where S contains no AP longer than* 3 *is* NP-*complete*.

5. Conclusion

Because the integers used in our proofs are exponentially larger than $|X|$, we have not shown our problems to be NP-complete in the strong sense [2]. Therefore, there is hope for a pseudo-polynomial time algorithm for each problem.

Let $|S| = n$. If we restrict Z to only those APs having a fixed increment m, then we find a minimum cover from Z in $O(n \log n)$ time as follows. Associate each $b \in S$ with the remainder/quotient pair $(b \bmod m, \lfloor b/m \rfloor)$. Sort the pairs lexicographically. A single scan of the sorted pairs finds all maximal APs in S having increment m. More generally, if we restrict Z to contain only APs having increments from a fixed set $\{m_1, \ldots, m_t\}$, then there is a dynamic programming algorithm to find a minimum cover or an exact cover in $O(n \log n + 2^t n)$ time. Since there can be $\binom{n}{2}$ different increments among the APs of S, a dynamic programming algorithm can find a minimum or exact cover in

$$O\left(n \log n + 2^{\binom{n}{2}} n\right)$$

time. We challenge the reader to find more efficient solutions for these new set covering problems.

Acknowledgment

The author expresses his appreciation to Arnold Rosenberg of the University of Massachusetts for suggesting the problem and many helpful discussions. Research supported in part by NSF grant MCS-83-01213 and by Air Force Contract AFOSR-86-0078.

References

[1] P. Erdös and P. Turan, On some sequences of integers, *J. London Math. Soc.* **11** (1936) 261–264.

[2] M.R. Garey and D.S. Johnson, *Computers and Intractability* (Freeman, San Francisco, CA, 1979).

[3] W.D. Grobman and T.W. Studwell, Data compaction and vector scan e-beam system performance improvement using a novel algorithm for recognition of pattern step and repeats, *J. Vacuum Sci. and Technology* **16** (1979) 1803–1808.

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

[5] R. Salem and D.C. Spencer, On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. U.S.A.* **28** (1942) 561–563.

[6] E. Szemerédi, On sets of integers contains no k elements in arithmetic progression, *Acta Arith.* **27** (1975) 199–245.