IBM Research

RC 5431
(#23737)
5/27/75
Mathematics

5 pages

THE SET BASIS PROBLEM IS NP-COMPLETE

L. J. Stockmeyer
Mathematical Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Yorktown Heights, New York

San Jose, California

Zurich, Switzerland

RC 5431
(#23737)
5/27/75
Mathematics

5 pages

THE SET BASIS PROBLEM IS NP-COMPLETE

L. J. Stockmeyer
Mathematical Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication elsewhere and has been issued as a Research Report for early dissemination of its contents. As a courtesy to the intended publisher, it should not be widely distributed until after the date of outside publication.

Copies may be requested from:
IBM Thomas J. Watson Research Center
Post Office Box 218
Yorktown Heights, New York 10598

Given a collection of sets $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$, a basis $\mathcal{B}$ is defined as a collection of sets $\mathcal{B} = \{B_1, B_2, \cdots, B_m\}$ such that for each S_i in $\mathcal{S}$ there exists a subset of $\mathcal{B}$ whose union equals S_i. The problem of finding a basis of least cardinality arises in several applications [2, 4]. Kou and Wong [6] show that this problem can be efficiently reduced to the clique cover problem.

The purpose of this note is to prove that the set basis problem is NP-complete [3, 5, cf.1], thus providing evidence that there is no efficient algorithm which finds a minimum basis in all cases. Moreover, this problem is NP-complete even with the restriction that each of the sets S_i is of cardinality three or less. However, for technical completeness, we note that if the problem is restricted further by requiring the S_i to be of cardinality two or less, then a minimum basis can be found in a computationally straightforward manner. We assume the reader is familiar with the terminology concerning NP-complete problems; see [1, 5].

For positive integer b, let b-SET BASIS denote the set of pairs $(\mathcal{S}, k)$ such that $\mathcal{S}$ is a collection of sets $\mathcal{S} = \{S_1, \cdots, S_n\}$ with[(1)] $\#S_i \le b$ for $1 \le i \le n$, k is an integer, and $\mathcal{S}$ possesses a basis $\mathcal{B} = \{B_1, \cdots, B_m\}$ where $m \le k$.

Let NODE COVER denote the set of pairs (G, ℓ) such that G is an undirected graph, ℓ is an integer, and there is a subset R of the nodes of G such that every edge of G is incident with some node in R (i.e., R is a node cover of G), and $\#R \le \ell$.

Fact [5]. NODE COVER is NP-complete.

---

(1) $\#S$ denotes the cardinality of the set S.

<u>Theorem 1</u> 3-SET BASIS is NP-complete.

<u>Proof</u>. 3-SET BASIS can obviously be recognized by a nondeterministic Turing machine within polynomial time. We now show that NODE COVER is polynomially transformable to 3-SET BASIS.

Let an undirected graph G and a positive integer ℓ be given. Say G has nodes N and edges $E = \{e_1, e_2, \cdots, e_m\}$. For each i with $1 \le i \le m$, let p(i) and q(i) be the endpoints of e_i. Assume $a_i, b_i \notin N$ for $1 \le i \le m$, and form the following set basis problem:

$$\mathcal{S} = \{ \{a_i, p(i), q(i)\}, \{a_i, p(i), b_i\}, \{a_i, q(i), b_i\} \mid 1 \le i \le m \};$$

$$k = \ell + 2m.$$

Clearly the transformation mapping (G, ℓ) to $(\mathcal{S}, k)$ can be computed within polynomial time. It remains to verify that G has a node cover of cardinality $\le \ell$ if and only if $\mathcal{S}$ has a basis of cardinality $\le k$.

I. (only if). Let R be a node cover of G such that $\#R \le \ell$. Let

$$\mathcal{B} = \{ \{a_i, b_i\} \mid 1 \le i \le m \} \cup \{ \{u\} \mid u \in R \}$$

$$\cup \{ \{a_i, p(i), q(i)\} - R \mid 1 \le i \le m \}.$$

Now $\#\mathcal{B} = \#R + 2m \le k$. Since, for each i, at least one of p(i) or q(i) belongs to R, it easily follows that $\mathcal{B}$ is a basis for $\mathcal{S}$.

II. (if). Let $\mathcal{B}$ be a basis for $\mathcal{S}$ such that $\#\mathcal{B} \le k$.

Lemma 1. For each j with $0 \le j \le m$ there is a basis $\mathcal{B}_j$ for S such that: (i) $\#\mathcal{B}_j \le \#\mathcal{B}$; and (ii) for each i with $1 \le i \le j$, either $\{p(i)\} \in \mathcal{B}_j$ or $\{q(i)\} \in \mathcal{B}_j$ (or both).

Proof. The $\mathcal{B}_j$ are constructed inductively. $\mathcal{B}_0 = \mathcal{B}$. Assume $\mathcal{B}_{j-1}$ has been constructed for some $j \le m$. Since $\{a_j,p(j),q(j)\}$ must be expressible as a union of sets in $\mathcal{B}_{j-1}$, we have one of several cases.

(1). If $\{p(j)\} \in \mathcal{B}_{j-1}$ or $\{q(j)\} \in \mathcal{B}_{j-1}$, then take $\mathcal{B}_j = \mathcal{B}_{j-1}$.

(2). Assume (1) does not hold, and suppose $U_1 \in \mathcal{B}_{j-1}$ where either $U_1 = \{a_j,p(j),q(j)\}$ or $U_1 = \{p(j),q(j)\}$. Since U_1 cannot be used in the unions for $\{a_j,p(j),b_j\}$ or for $\{a_j,q(j),b_j\}$, we must have $U_2,U_3 \in \mathcal{B}_{j-1}$, where $U_2 = \{p(j)\} \cup C_2$ and $U_3 = \{q(j)\} \cup C_3$ for some $C_2,C_3 \subseteq \{a_j,b_j\}$. Furthermore, $C_2 \ne \emptyset$ and $C_3 \ne \emptyset$ because (1) does not hold. Let

$$\mathcal{B}_j = (\mathcal{B}_{j-1} - \{U_1,U_2,U_3\}) \cup \{\ \{a_j,p(j)\},\ \{q(j)\},\ \{a_j,b_j\}\ \}.$$

Certainly $\#\mathcal{B}_j \le \#\mathcal{B}_{j-1}$. Now $\mathcal{B}_j$ is a basis for S. This is true because, since $C_2,C_3 \ne \emptyset$, the sets U_1,U_2,U_3 can only be used in unions for $T_1 = \{a_j,p(j),q(j)\}$, $T_2 = \{a_j,p(j),b_j\}$, and $T_3 = \{a_j,q(j),b_j\}$. But the three new sets added to $\mathcal{B}_j$ are a basis for $\{T_1,T_2,T_3\}$.

(3). If (1) and (2) do not hold, the only remaining possibility is $V_1,V_2 \in \mathcal{B}_{j-1}$ where $V_1 = \{a_j,p(j)\}$ and $V_2 = \{a_j,q(j)\}$. $\{a_j,p(j),b_j\} \in S$ implies that $V_3 \in \mathcal{B}_{j-1}$ where $V_3 = \{b_j\} \cup C$ for some set C. As in case (2), V_1, V_2, and V_3 can only be used in unions for T_1, T_2, and T_3. Thus $\mathcal{B}_j$ is a basis where

$$\mathcal{B}_j = (\mathcal{B}_{j-1} - \{V_1,V_2,V_3\}) \cup \{\ \{a_j,p(j)\},\ \{q(j)\},\ \{a_j,b_j\}\ \}.$$

This completes the proof of the lemma. $\square$

Let $R = \{ u \mid \{u\} \in \mathcal{B}_m$ for some $u \in N \}$; and $\mathcal{B}_c = \{ \{u\} \mid u \in R \}$. Since $\mathcal{B}_m$ satisfies (ii) of Lemma 1, R is a node cover of G. Let

$$\mathcal{B}_a = \{ B \in \mathcal{B}_m \mid (\exists i)[a_i \in B] \text{ and } (\forall i)[b_i \notin B] \},$$

$$\mathcal{B}_b = \{ B \in \mathcal{B}_m \mid (\exists i)[b_i \in B] \}.$$

Since $\{a_i, p(i), q(i)\} \in \mathcal{S}$ for $1 \le i \le m$, we must have $\#\mathcal{B}_a \ge m$. Since $\{a_i, p(i), b_i\} \in \mathcal{S}$ for $1 \le i \le m$, we have $\#\mathcal{B}_b \ge m$. But $\mathcal{B}_a$, $\mathcal{B}_b$, and $\mathcal{B}_c$ are pairwise disjoint. Therefore

$$2m + \#R \le \#\mathcal{B}_a + \#\mathcal{B}_b + \#\mathcal{B}_c \le \#\mathcal{B}_m \le \#\mathcal{B} \le k = 2m + \ell.$$

So $\#R \le \ell$ which completes the proof of Theorem 1. □

For technical completeness, it is appropriate to point out that 2-SET COVER can be recognized within deterministic polynomial time. A family of sets $\{S_1, \cdots, S_n\}$ is said to be <u>connected</u> iff for each $u, v \in \bigcup_{i=1}^{n} S_i$ there are $i(1), i(2), \cdots, i(k)$ such that $u \in S_{i(1)}$, $v \in S_{i(k)}$, and $S_{i(j)} \cap S_{i(j+1)} \neq \emptyset$ for $1 \le j < k$. Clearly a minimum basis for a (possibly non-connected) family $\mathcal{S}$ is the union of minimum bases for the connected components of $\mathcal{S}$. However, if the S_i are of cardinality ≤ 2, it is trivial to find a minimum basis for a connected family because:

<u>Lemma 2</u>. Let $\mathcal{S} = \{S_1, \cdots, S_n\}$ be connected, and $\#S_i \le 2$ for $1 \le i \le n$. Let $D = \bigcup_{i=1}^{n} S_i$. If $\#D \le n$, then $\{ \{d\} \mid d \in D \}$ is a minimum basis $\mathcal{S}$. If $\#D > n$, then $\mathcal{S}$ is a minimum basis for $\mathcal{S}$.

The proof of Lemma 2 is not difficult and is left to the reader.

REFERENCES

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Mass., 1974, 363-404.

2. H. D. Block, N. J. Nilsson, and R. O. Duda, Determination and detection of features in patterns, Computer and Information Sciences, J T. Tou and R. H. Wilcox, eds., Spartan Books, Washington, D. C., 1964, 75-110.

3. S. A. Cook, The complexity of theorem proving procedures, Proc. Third Annual ACM Symposium on Theory of Computing (1971), 151-158.

4. J. F. Gimpel, The minimization of spatially-multiplexed character sets, CACM 17, 6 (June 1974), 315-318.

5. R. M. Karp, Reducibility among combinatorial problems, in Complexity of Computer Computations, R. E. Miller and J. W. Thatcher, eds., Plenum Press, New York, 1972, 85-104.

6. L. T. Kou and C. K. Wong, A note on the set basis problem related to the compaction of character sets, IBM Report RC-5244, January, 1975.