# Relations and minimal keys*

J. DEMETROVICS and V. D. THI

## Abstract

The main purpose of this paper is to prove that the time complexity of finding a relation representing a given Sperner-system $K$ is exactly exponential in the number of elements of $K$. Conversely, we show that if $\mathrm{NP} \neq \mathrm{P}$ then the time complexity of finding a set of all minimal keys of given relation $R$ is also exactly exponential in the size of $R$.

## § 1. Introduction

The minimal keys play important roles for the logic and structural investigation of relational datamodel. In this datamodel the form of data storage is matrix (relation), rows of which represent records and columns represent attributes. A set of minimal keys of a relation forms a Sperner-system. Sets of minimal keys and Sperner-systems are equivalent in the sense that for an arbitrary Sperner-system $K$ there exists a relation $R$ such that the minimal keys of $R$ are exactly the elements of $K$ (cf. [3]), i.e. $R$ represents $K$.

In this paper we prove that the time complexity of finding a relation representing a given Sperner-system $K$ is exactly exponential in the number of elements of $K$, i.e. we shall show that there is an algorithm that determines a relation representing a given Sperner-system $K$ in time exponential in the number of elements of $K$, and there is no algorithm which finds a relation representing $K$ and the time complexity of which is polynomial in the number of elements of $K$. Let P denote the class of problems that can be solved in deterministic polynomial time and let NP denote the class of problems that can be solved in nondeterministic polynomial time. It is shown that if $\mathrm{NP} \neq \mathrm{P}$ then the time complexity of finding a set of all minimal keys of a given relation $R$ is exactly exponential in the number of rows and columns of $R$.

We start with some necessary definitions, and in § 2 formulate our results.

*This paper was supported by a grant from the Hungarian Academy of Sciences OTKA Nr. 1066.

**Definition 1.1.** Let $R=\{h_1, \ldots, h_m\}$ be a relation over $\Omega$, and $A, B \subseteq \Omega$. Then we say that $B$ functionally depends on $A$ in $R$ (denoted $A \xrightarrow[R]{f} B$) iff

$$(\forall h_i, h_j \in R)((\forall a \in A)(h_i(a) = h_j(a)) \to ((\forall b \in B)(h_i(b) = h_j(b))).$$

Let $F_R = \{(A, B) : A \xrightarrow[R]{f} B\} \cdot F_R$ is called the full family of functional dependencies of $R$.

**Definition 1.2.** Let $\Omega$ be a finite set, and denote $P(\Omega)$ its power set. Let $F \subseteq P(\Omega) \times P(\Omega)$. We say that $F$ is an $f$-family over $\Omega$ iff for all $A, B, C, D \subseteq \Omega$:

(F1) $(A, A) \in F$;
(F2) $(A, B) \in F, (B, C) \in F \to (A, C) \in F$;
(F3) $(A, B) \in F, A \subseteq C, D \subseteq B \to (C, D) \in F$;
(F4) $(A, B) \in F, (C, D) \in F \to (A \cup C, B \cup D) \in F$.

Clearly, $F_R$ is an $f$-family over $\Omega$.

It is known [2] that if $F$ is an arbitrary $f$-family, then there is a relation $R$ over $\Omega$ such that $F_R = F$.

**Definition 1.3.** The mapping $L\colon P(\Omega) \to P(\Omega)$ is called a closure operation over $\Omega$ iff for every $A, B \subseteq \Omega$

(1) $A \subseteq L(A)$;

(2) $A \subseteq B \to L(A) \subseteq L(B)$;

(3) $L(L(A)) = L(A)$.

**Definition 1.4.** Let $R$ be a relation, $L$ be a closure operation over $\Omega$, and $A \subseteq \Omega$. $A$ is a key of $R$ (a key of $L$) if $A \xrightarrow[R]{f} \Omega$ $(L(A) = \Omega)$. $A$ is a minimal key of $R$ (a minimal key of $L$) if $A$ is a key of $R$ (a key of $L$), but $B \not\xrightarrow[R]{f} \Omega$ $(L(B) \neq \Omega)$ for any proper subset $B$ of $A$. Denote $K_R$ $(K_L)$ the set of all minimal keys of $R$ $(L)$. Clearly, $K_R$, $K_L$ are Sperner-systems over $\Omega$.

**Definition 1.5.** Let $K$ be a Sperner-system over $\Omega$. We define the set of antikeys of $K$, denoted by $K^{-1}$, as follows:

$$K^{-1} = \{A \subsetneqq \Omega\colon (B \in K) \to (B \nsubseteq A) \quad \text{and} \quad (A \subsetneqq C) \to (\exists B \in K)\ (B \subseteq C)\}.$$

It is easy to see that $K^{-1}$ is also a Sperner-system over $\Omega$.

**Theorem 1.1.** ([2], [3]) If $K$ is an arbitrary Sperner-system, then there is a closure operation $L$ for which $K_L = K$. □

In this paper we always assume that if a Sperner-system plays the role of the set of minimal keys (antikeys), then this Sperner-system is not empty (doesn't contain $\Omega$).

**Definition 1.6.** ([2]) Let $F$ be an $f$-family over $\Omega$, and $(A, B) \in F$. We say that $(A, B)$ is a maximal right side dependency of $F$ iff $\forall B'$ $(B \subseteq B')\colon (A, B') \in F \to B = B'$. Denote by $M(F)$ the set of all maximal right side dependencies of $F$. We say that $B$

is a maximal side of $F$ iff there is an $A$ such that $(A, B) \in M(F)$. Denote by $I(F)$ the set of all maximal sides of $F$.

In this paper we regard the comparison of two attributes to be the elementary step of algorithms. Thus, if we assume that subsets of $\Omega$ are represented as sorted lists of attributes, then a Boolean operation on two subsets of $\Omega$ requires at most $|\Omega|$ elementary steps.

**Definition 1.7.** Let $R$ be a relation, and $K$ be a Sperner-system over $\Omega$. We say that $R$ represents $K$ iff $K_R = K$.

## § 2. Results

**Definition 2.1.** Let $L$ be a closure operation over $\Omega$.
Denote $Z(L) = \{A \in P(\Omega): L(A) = A\}$,

$$T(L) = \{A \in P(\Omega): L(A) = A \quad \text{and} \quad A \subsetneqq B \rightarrow L(B) = \Omega\}.$$

The elements of $Z(L)$ are called closed sets. $T(L)$ is called a maximal family of $L$.

**Lemma 2.1.** ([5]) Let $L$ be a closure operation over $\Omega$. Then $K_L^{-1} = T(L)$. □

**Theorem 2.1.** ([4]) Let $K$ be a Sperner-system over $\Omega$. Let $s(K) = \min \{m: |R| = m, K_R = K, R$ is a relation over $\Omega\}$. Then $\sqrt{2|K^{-1}|} \leqq s(K) \leqq |K^{-1}| + 1$. □

**Theorem 2.2.** The time complexity of finding a relation representing a given Sperner-system $K$ is exactly exponential in the number of elements of $K$.

*Proof.* We have to prove that:

(1) There exists an algorithm that determines a relation representing a given Sperner-system $K$ in time exponential in the number of elements of $K$.

(2) There is no algorithm that finds a relation representing $K$ in time polynomial in the number of elements of $K$. Based on (1) and (2) it is clear that the time complexity of any algorithm that determines a relation representing a given Sperner-system is at least exponential in the number of elements of this Sperner-system.

For (1): First we construct an algorithm which finds the set of antikeys from a given Sperner-system, as follows:

Let us given an arbitrary Sperner-system $K = \{B_1, \ldots, B_m\}$ over $\Omega$. We set $K_1 = \{\Omega \setminus \{a\}: a \in B_1\}$. It is obvious that $K_1 = \{B_1\}^{-1}$. Let us suppose that we have constructed $K_q = \{B_1, \ldots, B_q\}^{-1}$ for $q < m$. We assume that $X_1, \ldots, X_{t_q}$ are the elements of $K$ containing $B_{q+1}$. So $K_q = F_q \cup \{X_1, \ldots, X_{t_q}\}$, where $F_q = \{A \in K_q: B_{q+1} \nsubseteq A\}$. For all $i$ $(i = 1, \ldots, t_q)$ we construct the antikeys of $\{B_{q+1}\}$ on $X_i$ in an analogous way as $K_1$, which are the maximal subsets of $X_i$ not containing $B_{q+1}$. We denote them by $A_1^i, \ldots, A_{r_i}^i$ $(i = 1, \ldots, t_q)$. Let

$$K_{q+1} = F_q \cup \{A_p^i: A \in F_q \rightarrow A_p^i \not\subset A, \quad 1 \leqq i \leqq t_q, \quad 1 \leqq p \leqq r_i\}.$$

Clearly, because $K$ and $K^{-1}$ are uniquely determined by one another, the determination of $K^{-1}$ based on our algorithm does not depend on the order of $B_1, \ldots, B_m$.

In [5] we proved that for every $q$ $(1 \leqq q \leqq m)$, $K_q = \{B_1, \ldots, B_q\}^{-1}$, i.e. $K_m = K^{-1}$, and the worst-case time of this algorithm is exponential not only in the number of elements of $K$, but also in the number of attributes. Now we construct the following algorithm:

*Step 1:* Based on the above algorithm we construct $K^{-1}$.

*Step 2:* Let $K^{-1}=\{A_1, \dots, A_t\}$ be a set of antikeys. Let $R=\{h_0, h_1, \dots, h_t\}$ be a relation over $\Omega$ given as follows:

For all $a \in \Omega$, $h_0(a)=0$,

$$\text{for } i \ (1 \leqq i \leqq t), \quad h_i(a) = \begin{cases} 0 & \text{if } a \in A_i, \\ i & \text{otherwise.} \end{cases}$$

In [4], it has been proved that $R$ represents $K$. It is clear that the complexity of this algorithm is the complexity of the algorithm that finds the set of antikeys.

For (2): Let us take a partition $\Omega = X_1 \cup \dots \cup X_m \cup W$, where $m = \left[\frac{n}{3}\right]$, and $|X_i|=3$ $(1 \leqq i \leqq m)$. Let

$$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i\} \quad \text{if} \quad |W| = 0,$$

$$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i: 1 \leqq i \leqq m-1 \text{ or } B \subseteq X_m \cup W\} \text{ if } |W| = 1,$$

$$K = \{B: |B| = 2, B \subseteq X_i \text{ for some } i: 1 \leqq i \leqq m \text{ or } B = W\} \text{ if } |W| = 2.$$

It is clear that

$$K^{-1} = \{A: |A \cap X_i| = 1, \forall i\} \quad \text{if} \quad |W| = 0,$$

$$K^{-1} = \{A: |A \cap X_i| = 1 \ (1 \leqq i \leqq m-1) \text{ and } |A \cap (X_m \cup W)| = 1\} \text{ if } |W| = 1,$$

$$K^{-1} = \{A: |A \cap X_i| = 1 \ (1 \leqq i \leqq m) \text{ and } |A \cap W| = 1\} \text{ if } |W| = 2.$$

Let $f: N \to N$ ($N$ is the set of natural numbers) be the function defined as follows:

$$f(n) = \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \ (\text{mod } 3), \\ 3^{[n/3]} \cdot 4/3 & \text{if } n \equiv 1 \ (\text{mod } 3), \\ 3^{[n/3]} \cdot 2 & \text{if } n \equiv 2 \ (\text{mod } 3). \end{cases}$$

It can be seen that $f(n)=|K^{-1}|$. Clearly, $n-1 \leqq |K| \leqq n+2$ and $3^{[n/4]} < f(n)$, where $n=|\Omega|$, i.e. $3^{[n/4]} < |K^{-1}|$. According to Theorem 2.1 we have $\sqrt{2} \cdot 3^{[n/8]} \leqq s(K)$, i.e. the number of rows of minimal relation representing $K$ is greater than $\sqrt{2} \cdot 3^{[n/8]}$. Thus, for an arbitrary set of attributes we always can construct an example, in which the number of $K$ is not greater than $|\Omega|+2$, but the number of rows of any relation representing $K$ is exponential not only in the number of attributes, but also in the number of elements of $K$. Hence, there is no algorithm which finds a relation representing a given Sperner-system and the time complexity of which is polynomial in the number of elements of Sperner-system. The theorem is proved. □

Now we give a necessary and sufficient condition for a relation to represent a given Sperner-system. We define the following concept.

**Definition 2.2.** Let $R=\{h_1, \ldots, h_m\}$ be a relation over $\Omega$. Let $E_R=\{E_{ij}: 1\leqq i<j\leqq m\}$, where $E_{ij}=\{a\in\Omega: h_i(a)=h_j(a)\}$. Let

$$M_R = \{A\in P(\Omega): \exists E_{ij}\in E_R: E_{ij} = A \text{ and } \bar{\exists} E_{st}\in E_R: A\subsetneqq E_{st}\}.$$

$M_R$ is called the maximal equality system of $R$.

**Theorem 2.3.** Let $K$ be a non-empty Sperner-system, and $R$ be a relation over $\Omega$. Then $R$ represents $K$ iff $K^{-1}=M_R$, where $M_R$ is the maximal equality system of $R$.

*Proof.* Because $K$ is a non-empty Sperner-system, $K^{-1}$ exists. On the other hand, $K$ and $K^{-1}$ are uniquely determined by each other, hence $K_R=K$ holds iff $K_R^{-1}=K^{-1}$ holds. Consequently, we must prove that $K_R^{-1}=M_R$. It is obvious that $F$ is a $f$-family. Now we suppose that $A$ is an antikey of $K$. It can be seen that $A\neq\Omega$. If there exists a $B$ such that $A\subsetneqq B$ and $A\xrightarrow[R]{f} B$, then by the definition of antikey we have $B\xrightarrow[R]{f}\Omega$. Hence $A\xrightarrow[R]{f}\Omega$ holds. This contradicts $C\in K_R: C\nsubseteq A$. So $A\in I(F_R)$ holds. If there is a $B'$ so that $B'\neq\Omega$, $B'\in I(F_R)$, and $A\subsetneqq B'$ then $B'$ is a key of $R$. This contradicts to $B'\neq\Omega$. Consequently, $A\in I(F_R)\setminus\Omega$ and $\bar{\exists} B'$ $(B'\in I(F_R)\setminus\Omega)$: $A\subsetneqq B'$. On the other hand, according to the of relation $\Omega\notin M_R$. It is easy to see that $E_{ij}\in I(F_R)$. Thus, $M_R\subseteqq I(F_R)$ holds. If $D$ is a set such that $\forall C\in M_R: D\nsubseteq C$, then by the definition of functional dependency, $D$ is a key of $R$. Consequently, $M_R$ is the set of maximal distinct elements of $I(F_R)$. So we have $A\in M_R$.

Conversely, we assume that $A\in M_R$. According to the definition of relation and $M_R$ we obtain $A\not\xrightarrow[R]{f}\Omega$, i.e. $\forall B\in K_R: B\nsubseteq A$. On the other hand, because $A$ is a maximal equality set, for all $D$ $(A\subsetneqq D)$ $D\xrightarrow[R]{f}\Omega$ holds. Consequently, by the definition of antikey $A\in K_R^{-1}$. The theorem is proved. □

It can be seen that the time complexity of finding the set of antikeys of $R$ is polynomial in the number of rows and columns of $R$. We construct the following algorithm for finding a minimal key. Let $H$ be a Sperner-system. We take a $B$ $(B\in H)$ and an $a\in\Omega\setminus B$. We suppose that $B=\{b_1, \ldots, b_m\}$. Let $G=\{B_j\in H: a\notin B_j\}$ and $T_0=B\cup\{a\}$. We define

$$T_{q+1} = \begin{cases} T_q\setminus\{b_{q+1}\} & \text{if } \forall B_i\in H\setminus G: T_q\setminus\{b_{q+1}\}\nsubseteq B_i, \\ T_q & \text{otherwise.}\end{cases}$$

**Lemma 2.2.** ([5]) If $H$ is a set of antikeys, then $T_0$, $T_1, \ldots, T_m$ are the keys and $T_m$ is a minimal key. □

It is easy to see that the worst-case time of finding $T_m$ is $O(|\Omega|^2\cdot|H|)$.

**Lemma 2.3.** Let $H$ be a Sperner-system over $\Omega$, and let $H^{-1}=\{B_1, \ldots, B_m\}$ be a set of antikeys of $H$, $T\subseteqq H$. Then $T\subsetneqq H$ and $T\neq\emptyset$ if and only if there is a $B\subseteqq\Omega$ such that $B\in T^{-1}$ and $B\nsubseteq B_i$ $(\forall i: 1\leqq i\leqq m)$.

*Proof.* Suppose that there exists a $B$ such that $B\in T^{-1}$ and $B\nsubseteq B_i$ $(\forall i: 1\leqq i\leqq m)$. From the definition of the set of antikeys and by $T^{-1}\neq\emptyset$, we have $T\neq\emptyset$, and for all $C$ $(C\in T)$, $B$ does not contain $C$. If there is a $B_i$ such that $B_i\in H^{-1}$ and $B_i\subset B$, then it is obvious that $B$ is a key. If $H^{-1} \cup B$ is a Sperner-system, then by Theorem 1.1 there exists a closure operation $L$ such that $H = K_L$. It is clear that if $L(B) \neq \Omega$, then from Lemma 2.1 there is a $B_i$ $(B_i \in H^{-1})$ such that $L(B) \subseteq B_i$. Consequently, $B \subseteq B_i$. This conflicts with the fact that $B \nsubseteq B_i$ $(\forall i: 1 \leqq i \leqq m)$. That is, $B$ is a key. Hence there is an $A$ $(A \subseteq \Omega)$ such that $A \subseteq B$ and $A \in H \setminus T$. It is easy to see that $T \subset H$.

Conversely, we suppose that $T \subset H$ and $T \neq \emptyset$. It is obvious that there is an $A$ such that $A \in H \setminus T$. From $H$ is a Sperner-system we have $A \cup T$ is a Sperner-system. Denote $B$ the biggest set such that $A \subseteq B$ and $B \cup T$ is also a Sperner-system. It is clear that, $B$ always exists and from the definition of antikeys we have $B \in T^{-1}$. By $A \in H$ it can be seen that $A \nsubseteq B_i$ $(\forall i: 1 \leqq i \leqq m)$. By $A \subseteq B$ we have $B \nsubseteq B_i$ $(\forall i: 1 \leqq i \leqq m)$. The theorem is proved. □

Let $K = \{B_1, \ldots, B_m\}$ be a Sperner-system over $\Omega$. We have to construct $H$, where $H^{-1} = K$. We construct $H$ by induction.

**Algorithm 2.1.** *Step 1:* Using a minimal key algorithm we construct an $A_1$, $(A_1 \in H)$. We set $K_1 = \{A_1\}$.

*Step* $i+1$: If there is a $B \in K_i^{-1}$ such that $B \nsubseteq B_j (\forall j: 1 \leqq j \leqq m)$, then by algorithm which finds a minimal key we determine an $A_{i+1}$ $(A_{i+1} \in H)$ and $A_{i+1} \subseteq B$. After that, let $K_{i+1} = K_i \cup \{A_{i+1}\}$. In the converse case we set $H = K_i$. □

Based on Lemma 2.3 there is a natural number $p$ so that $K_p = H$. It can be seen that the time complexity of Algorithm 2.1 is also exponential in the number of attributes.

**Lemma 2.4.** The following problem is NP-complete:

Given a Sperner-system $K = \{B_1, \ldots, B_m\}$ over $\Omega = \{a_1, \ldots, a_n\}$ and integer $k$ $(k \leqq n)$, decide whether there exists an $A \subseteq \Omega$ such that $|A| \leqq k$ and $A \nsubseteq B_i$ $(i = 1, \ldots, m)$, i.e. decide whether there exists a key having cardinality not greater than $k$, if $K$ is the set of antikeys.

*Proof.* We nondeterministically choose a subset $A$ of $\Omega$ so that $|A| \leqq k$ and decide whether $A$ is not a subset of $B_i (i = 1, \ldots, m)$. It is obvious that this algorithm is nondeterministic polynomial. Thus, the problem lies in NP. It is known [1] that the vertex cover problem is NP-complete:

Given integer $k$ and non-directed graph $G = \langle V, E \rangle$, where $V$ is a set of vertices and $E$ is set of edges, decide whether or not $G$ has a vertex cover having cardinality not greater than $k$.

We shall prove that the vertex cover problem is polynomially reducible to our problem.

Let $G = \langle V, E \rangle$ be a non-directed graph, $k \leqq |V|$. We set $\Omega = V$ and $K = \{\Omega \setminus \{a_i, a_j\} : (a_i, a_j) \in E\}$.

If $A \subseteq \Omega$, $|A| \leqq k$ and $A \nsubseteq B$ $(\forall i = 1, \ldots, m)$, then according to definition of $K$ we have $A \cap \{a_i, a_j\} \neq \emptyset (\forall (a_i, a_j) \in E)$. Consequently, $A$ is a vertex cover of $G$.

Conversely, if $A$ is a vertex cover of $G$, then by definition of $K$ and definition of vertex cover we have $A \nsubseteq B_i$ $(\forall i = 1, \ldots, m)$. Hence, $A \nsubseteq B_i$ $(\forall i = 1, \ldots, m)$ holds if and only if $A$ is a vertex cover of $G$. The Lemma is proved. □

Based on Lemma 2.4 and Step 2 of the algorithm which determines a relation representing a given Sperner-system in Theorem 2.2, the following corollary is obvious.

**Corollary 2.1.** The following problem is NP-complete: Given integer $k$ and relation, decide whether or not there exists a key having cardinality not greater than $k$. □

**Theorem 2.4.** The time complexity of finding a set of all minimal keys of a given relation $R$ is exactly exponential in the number of rows and columns of $R$.

*Proof.* For a given arbitrary relation $R$ we construct the following algorithm which determines the set of all minimal keys of $R$.

*Step 1:* According to Theorem 2.3 we construct the set of antikeys of $R$.

*Step 2:* Based on Algorithm 2.1 we determine the set of all minimal keys of $R$.
By Lemma 2.2, Lemma 2.3, Theorem 2.3 and Algorithm 2.1, it is clear that the worst-case time of this algorithm is exponential in the number of rows and columns of $R$.

According to Lemma 2.4 and Corollary 2.1, it can be seen that there is no algorithm which finds a set of all minimal keys of a given relation and the time complexity of which is polynomial in the size of this relation. The theorem is proved. □

Based on Theorem 2.1 and Theorem 2.4 it can be seen that the problem of finding a relation representing a given Sperner-system and finding a set of all minimal keys of a relation are inherently difficult.

## Резюме

В настоящей работе изучается связь между отношениями и минимальными ключами.

COMPUTER AND AUTOMATION INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
1132 BUDAPEST
VICTOR HUGO U. 18—22
HUNGARY

## References

[1] A. V. Aho, J. E. Hopcroft, J. D. Ullman, The design and analysis of computer algoritms. Addison-Wesley, Reading, Mass. 1974.
[2] W. W. Armstrong, Dependency structures of data base relationships. Inf. Proc. 74 North-Holland Pub. Co. (1974) 580—583.
[3] J. Demetrovics, On the equivalence of candidate keys with Sperner-systems. Acta Cybernetica 4 (1979) 247—252.
[4] J. Demetrovics, Z. Füredi, G. O. H. Katona, Minimum matrix representation of closure operations. Discrete Applied Mathematics 11 (1985) 115—128.
[5] V. D. Thi, Minimal keys and antikeys. Acta Cybernetica 7 (1986), 361—371.

*(Received March 12, 1987)*