Minimum Covers in the Relational Database Model

DAVID MAIER

Princeton University, Princeton, New Jersey

ABSTRACT Numerous algorithms concerning relational databases use a cover for a set of functional dependencies as all or part of their input. Examples are Beeri and Bernstein's synthesis algorithm and the tableau modification algorithm of Aho et al. The performance of these algorithms may depend on both the number of functional dependencies in the cover and the total size of the cover. Starting with a smaller cover will make such algorithms run faster. After Bernstein, many researchers believe that the problem of finding a minimum cover is NP-complete. It is shown here that minimum covers can be found in polynomial time, using the notion of *direct determination*. The proof details the structure of minimum covers, refining the structure Bernstein and Beeri show for nonredundant covers. The kernel algorithm of Lewis, Sekino, and Ting is improved using these results.

KEY WORDS AND PHRASES: relational database, functional dependency, minimum cover, nonredundant cover, efficient algorithms

CR CATEGORIES: 4.33, 5.23

1. *Introduction*

Consider the following simple problem for databases. We are given a relation r and a set of functional dependencies (FDs) F to enforce on r. After any update to r, we wish to determine whether the relation satisfies the FDs in F. One way to proceed with the problem is to take each FD $X \rightarrow Y$ in F in turn, sort the relation to bring equal values of X together, and check if these equal values of X correspond to equal values of Y. If not, r violates F. If r is the relation

A	B	C
a_1	b_1	c_1
a_1	b_2	c_1
a_2	b_1	c_2
a_2	b_2	c_1

we see that r satisfies the FD $AB \rightarrow C$, since r is already sorted by AB-values. Testing r against the FD $B \rightarrow C$, we sort by B-values to get

A	B	C
a_1	b_1	c_1
a_2	b_1	c_2
a_1	b_2	c_1
a_2	b_2	c_1

to see that r violates this FD.

The time required to check an FD $X \rightarrow Y$ against the relation r directly depends on the number of attribute symbols in X and in Y. The sorting process is repeated as many times as there are FDs in F. For any cover F' of F, if r satisfies F', then r satisfies F. To solve the

This work was supported by an IBM Research Fellowship.

Author's present address: Department of Computer Science, State University of New York at Stony Brook, Stony Brook, NY 11794.

satisfaction problem more quickly, we can seek covers for F with fewer attribute symbols or fewer FDs. If F is the set $\{AB \rightarrow C, A \rightarrow B\}$, then $F' = \{A \rightarrow C, A \rightarrow B\}$ covers F and has one fewer attribute symbol. Or, given $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$, we can use the cover $F' = \{A \rightarrow B, B \rightarrow C\}$, which has fewer FDs.

The next section defines several kinds of minimality for covers and presents some basic results. Direct determination is introduced in Section 3 and is used there to elucidate the structure of minimum covers. In Section 4 we show how to find minimum covers in polynomial time. Section 5 uses the results of Sections 3 and 4 to improve an algorithm of Lewis, Sekino, and Ting [11]

2. *Notions of Minimality*

The reader should be familiar with the notation of the relational model and functional dependencies. For an introduction, see Date [10], Beeri and Bernstein [5], Bernstein [6, 7], or Ullman [17]. Throughout this paper we assume that all attributes are chosen from some fixed universe U. Let F be a set of FDs. The *closure* of F, written F^+, is the set of all FDs that can be inferred from the FDs in F. The set F^+ can be computed by repeated application of a complete set of inference axioms to F. The following set of inference axioms can be proved complete using Armstrong's axioms [4, 13, 14].

For V, W, X, Y, Z, subsets of U,

A1. (reflexivity) $X \rightarrow X$.
A2. (projectivity) $X \rightarrow YZ$ implies $X \rightarrow Y$.
A3. (accumulation) $X \rightarrow YZ$ and $Z \rightarrow VW$ imply $X \rightarrow YZV$.

The convention for attribute symbols above and elsewhere is that capital letters from the beginning of the alphabet represent single attributes, capital letters from the end of the alphabet stand for sets of attributes, and concatenation is used for union.

Definition. Given sets of FDs F and G, F is a *cover* for G if $F^+ = G^+$. That is, F and G imply the same set of FDs. We also say that F and G are *equivalent*, written $F \equiv G$, if $F^+ = G^+$.

Saying that F is a cover of G says nothing about the relative sizes of F and G. We now define various restrictions of FDs that will guarantee different sorts of minimality.

Definition. A set of FDs F is *nonredundant* if there is no set of FDs G properly contained in F with $G^+ = F^+$. A nonredundant cover is also called a minimal cover (but not here).

Definition. The sets of attributes X and Y are *equivalent* under a set of FDs F, written $X \leftrightarrow Y$, if $X \rightarrow Y$ and $Y \rightarrow X$ are in F^+.

An important property of nonredundant covers is given by the following lemma of Bernstein [7].

LEMMA 1. *If G and F are equivalent, nonredundant sets of FDs and there is an FD $X \rightarrow W$ in G, then there is an FD $Y \rightarrow Z$ in F with $X \leftrightarrow Y$ under F.*

Lemma 1 implies that given a set of FDs G, if the FDs of any nonredundant cover F of G are partitioned on the basis of equivalent left sides, the number of cells in the partition is independent of the choice of F. In such a partition for a set of attributes X, let $E_F(X)$ be the set of all FDs in F with left sides equivalent to X and let $e_F(X)$ be the set of left sides of FDs in $E_F(X)$. Let $\bar{E}_F$ be the collection of all nonempty $E_F(X)$'s. (That is, X is equivalent to some left side of an FD in F.) For example, if $F = \{A \rightarrow BC, B \rightarrow A, AD \rightarrow E, BD \rightarrow C\}$, then $\bar{E}_F = \{E_F(A), E_F(AD)\}$, where

$$E_F(A) = \{A \rightarrow BC, B \rightarrow A\} \qquad \text{and} \qquad E_F(AD) = \{AD \rightarrow E, BD \rightarrow C\}.$$

$$\text{optimal} \not\Leftarrow \text{LR-minimum} \underset{\not\Leftarrow}{\Rightarrow} \text{L-minimum} \underset{\not\Leftarrow}{\Rightarrow} \text{minimum} \underset{\not\Leftarrow}{\Rightarrow} \text{nonredundant}$$

FIGURE 1

Definition (Paredaens [15]). A set of FDs F is *canonical* if F is nonredundant and, for every FD $X \rightarrow Y$ in F,

(1) Y is a single attribute, and
(2) there is no X' properly contained in X with $X' \rightarrow Y$ in F^+.

Definition. A set of FDs F is *minimum* if there is no set G with fewer FDs than F such that $G \equiv F$.

Definition. A set of FDs F is *L-minimum* if

(1) F is minimum, and
(2) for every FD $X \rightarrow Y$ in F, there is no X' properly contained in X with $X' \rightarrow Y$ in F^+.

Definition. A set of FDs F is *LR-minimum* if it is L-minimum and replacing FD $X \rightarrow Y$ in F by $X \rightarrow Y'$, with Y' properly contained in Y, alters the closure of F.

Definition. A set of FDs F is *optimal* if there is no set of FDs G with fewer attribute symbols such that $G \equiv F$. Repeated symbols are counted as many times as they occur. For example, $F = \{A \rightarrow BC, B \rightarrow A, AD \rightarrow C\}$ uses eight symbols.

Canonical, L-minimum, LR-minimum and optimal sets have no unnecessary symbols in the left sides of their FDs. Canonical, LR-minimum, and optimal sets have no unnecessary symbols in the right sides as well. Figure 1 shows the relationship between the definitions.

The implications come directly from the definitions. The following counterexamples show the nonimplications.

(1) $\{A \rightarrow B, A \rightarrow C\}$ is nonredundant but not minimum; $\{A \rightarrow BC\}$ has fewer FDs.
(2) $\{ABC \rightarrow D, A \rightarrow B\}$ is minimum but not L-minimum; the B can be removed from the left side of the first FD.
(3) $\{A \rightarrow AB\}$ is L-minimum but not LR-minimum; the A can be removed from the right side.
(4) $\{ABC \rightarrow D, BC \rightarrow E, E \rightarrow BC\}$ is LR-minimum but not optimal; $\{AE \rightarrow D, BC \rightarrow E, E \rightarrow BC\}$ uses fewer attribute symbols.

The missing parts of the diagram are canonical sets and the implication or nonimplication from optimal to LR-minimum. Canonical sets are treated shortly; optimal sets are taken up at the end of Section 3.

Beeri and Bernstein [5] introduce the notion of a *G-based derivation tree* for an FD $X \rightarrow B$, where G is a set of FDs and B is a single attribute. This tree is a chart of applications of axiom A3 used to derive B from X using FDs from G. We extend the notion to a *G-based derivation DAG* (G-based DDAG) for an FD $X \rightarrow Y$, where Y is a set of attributes. A G-based derivation DAG is defined constructively according to the following rules:

R1. Any set of unconnected nodes labeled with attributes from U is a G-based DDAG.
R2. If H is a G-based DDAG, $v_1, v_2, \ldots, v_n$ are nodes in H labeled $B_1, B_2, \ldots, B_n$, and $B_1B_2 \cdots B_n \rightarrow CZ$ is an FD in G, then the DAG H' obtained from H by adding a node u labeled C and edges $(v_1, u), (v_2, u), \ldots, (v_n, u)$ is a G-based DDAG.
R3. Nothing else is a G-based DDAG.

Rule R2 ensures that the graphs constructed are actually DAGs. An *initial node* of a

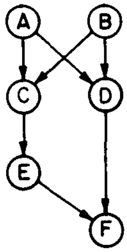

FIGURE 2

FIGURE 3

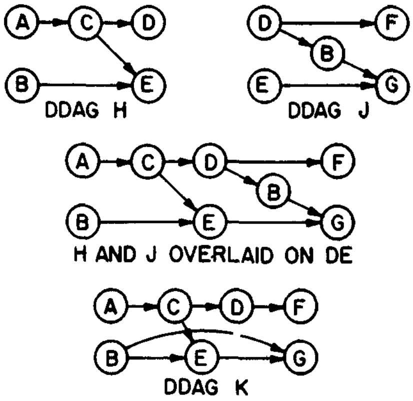

FIGURE 4

DDAG is a node with no incoming edges. A DDAG H *represents* a derivation for $X \rightarrow Y$ if initial nodes have labels in X and every attribute of Y labels some node of H.

Figure 2 shows a G-based DDAG for $AB \rightarrow CF$, where $G = \{AB \rightarrow CD, A \rightarrow J, C \rightarrow E, DE \rightarrow FJ\}$.

For any G-based DDAG H we want to know which FDs of G were used to construct H. Call this set $U(H)$, the *use set of* H. It contains all those FDs $B_1B_2 \cdots B_n \rightarrow CZ$ used in applying R2 in the definition of a DDAG while constructing H. For the DDAG H in Figure 2, $U(H) = \{AB \rightarrow CD, C \rightarrow E, DE \rightarrow FJ\}$. We should actually say *a* use set for H, since there may be more than one, but we shall not. However, if G is minimum, then there is only one choice for $U(H)$, since G cannot contain distinct FDs $B_1B_2 \cdots B_n \rightarrow CZ$ and $B_1B_2 \cdots B_n \rightarrow CW$. Although axioms A1 and A2 do not appear explicitly in the definition of DDAGs, they are implicitly incorporated. For example, Figure 3 is a G-based DDAG for $AB \rightarrow AB$ for any attributes A and B in the universe U.

There is a direct correspondence between G-based DDAGs for an FD $X \rightarrow Y$ in G^+ and derivations of $X \rightarrow Y$ in G using axioms A1–A3. The DDAG is essentially a diagram showing what applications of axiom A3 are used to derive Y from X. Given a derivation of $X \rightarrow Y$, we can use the applications of A3 to construct a DDAG for $X \rightarrow Y$ using rule R2. The following lemma is similar to one Beeri and Bernstein present for G-based derivation trees [5].

LEMMA 2. *If $X \rightarrow Y$ is in $U(H)$ for some G-based DDAG H of $V \rightarrow Z$, then $V \rightarrow X$ is in G^+.*

PROOF. If $X \rightarrow Y$ is in $U(H)$, then all the attribute symbols of X must appear as labels of nodes of H. Thus H is also a G-based DDAG for $V \rightarrow X$. For example, in Figure 2 $DE \rightarrow FG \in U(H)$, and H is a G-based DDAG for $AB \rightarrow DE$. □

LEMMA 3. *Take an LR-minimum set F, and form F' by splitting the right sides of FDs into single attributes ($\{AB \rightarrow CD, E \rightarrow AD\}$ becomes $\{AB \rightarrow C, AB \rightarrow D, E \rightarrow A, E \rightarrow D\}$). F' is a canonical cover of F^+.*

PROOF. Suppose $XY \rightarrow AZ$ is in F, and after splitting right sides, $XY \rightarrow A$ is redundant in F'. Then F is not LR-minimum, since $XY \rightarrow Z$ can replace $XY \rightarrow AZ$ in F without altering its closure.

Suppose $XY \rightarrow A$ can be replaced by $X \rightarrow A$ in F'. Then $X \rightarrow A$ is in F^+. The dependency $X \rightarrow A$ cannot be derived from $F - \{XY \rightarrow AZ\}$, since otherwise A would not appear in $XY \rightarrow AZ$. So there is an F-based DDAG for $X \rightarrow A$ using $XY \rightarrow AZ$. By Lemma 2, $X \rightarrow XY$ must be in F^+, and therefore $X \rightarrow Y$ is in F^+. Thus $XY \rightarrow AZ$ cannot be in F, since Y is superfluous. □

Observations. If H is a G-based DDAG for $X \to Y$, then there is a G-based DDAG H' for $X \to Y$ with every node having a distinct label. Suppose v and w are nodes in H both labeled C. Assume v was added before w or at the same time, so there is no directed path from w to v. Remove w and all its incoming edges. Attach the outgoing edges of w to v. Repeat the process for all pairs of same-labeled nodes to get H'. Note that $U(H) \supseteq U(H')$.

If H and J are G-based DDAGs for FDs $X \to Y$ and $Y \to Z$, we can construct a G-based DDAG K for $X \to Z$ having no duplicate labels with $U(K)$ contained in the union of $U(H)$ and $U(J)$. Assume that H and J have no duplicate labels. K is formed by *splicing* H and J together: Overlay H and J so the initial nodes in J coincide with the nodes labeled Y in H. If the result has duplicate nodes, remove them as described above. The result is K. In Figure 4, G is $\{A \to C, C \to D, BC \to E, D \to FB, BE \to G\}$, and the DDAGs are for $AB \to DE$ and $DE \to FG$. The DDAGs are combined and excess nodes eliminated to form a DDAG for $AB \to FG$. Actually, $Y \to Z$ can be replaced with $W \to Z$ for any set W of attributes labeling nodes in H, since H must also be a DDAG for $X \to W$.

3. *Direct Determination*

Definition. Given a set of FDs G with $X \to Y$ in G^+, *X directly determines Y* under G, written $X \xrightarrow{\cdot} Y$, if there exists an F-based DDAG H for $X \to Y$ with $U(H) \cap E_F(X) = \varnothing$ for some nonredundant cover F of G. That is, no FDs with left sides equivalent to X are used in H.

Note. $E_F(X)$ may itself be empty, and always $X \xrightarrow{\cdot} X$ As an example, if $F = G = \{A \to B, C \to D, AC \to E\}$, then $AC \xrightarrow{\cdot} BD$ under G.

As the definition stands it is not particularly useful, for the existence of a DDAG not using FDs from $E_F(X)$ might depend on which cover F is chosen. Checking direct determination could become computationally very hard. The next lemma proves that the choice of a cover for G is immaterial.

LEMMA 4. *X directly determines Y under G if and only if for every nonredundant cover F for G there exists an F-based DDAG H for $X \to Y$ with $U(H) \cap E_F(X) = \varnothing$.*

PROOF. Let F be a nonredundant cover for G for which there is an F-based DDAG H of $X \to Y$ using no FDs in $E_F(X)$. For every FD $Z \to W$ in $U(H)$, Lemma 2 states that $X \to Z$. Let F' be another nonredundant cover for G. Suppose some F'-based DDAG for $Z \to W$ uses an FD in $E_{F'}(X)$, say $T \to S$. By Lemma 2, $Z \to T$. But T is equivalent to X, so $Z \to X$, and hence Z is equivalent to X, which contradicts the assumption about H. Therefore every $Z \to W$ in $U(H)$ has a F'-based DDAG that does not use FDs from $E_{F'}(X)$. We obtain the required F'-based DDAG for $X \to Y$ by splicing together the DDAGs for each $Z \to W$ in $U(H)$. □

COROLLARY. *$X \xrightarrow{\cdot} Y$ under G if and only if for every cover F for G there exists an F-based DDAG H for $X \to Y$ with $U(H) \cap E_F(X) = \varnothing$.*

PROOF. Every cover F for G contains a nonredundant cover as a subset. □

The next lemma gives a limited transitivity rule for a direct determination

LEMMA 5. *If $X \xrightarrow{\cdot} Y$, $Y \xrightarrow{\cdot} Z$, and $Y \to X$ under G, then $X \xrightarrow{\cdot} Z$ under G.*

PROOF. Let F be a nonredundant cover for G, and let H and J be DDAGs for $X \to Y$ and $Y \to Z$ such that $U(H)$ and $U(J)$ contain no FDs from $E_F(X)$ ($= E_F(Y)$, since $X \leftrightarrow Y$). Splicing H and J will form an F-based DDAG K for $X \to Z$ that uses no FDs in $E_F(X)$, by the observation at the end of the last section. □

LEMMA 6. *Let F be nonredundant. Pick an X that is a left side in F and any set Y equivalent to X. There is some Z in $e_F(X)$ such that $Y \xrightarrow{\cdot} Z$.*

PROOF. Assume Y is not in $e_F(X)$. Otherwise $Y \xrightarrow{\cdot} Y$, and the lemma is proved. Since $X \leftrightarrow Y$, for every Z in $e_F(X)$ there must be a derivation in F for $Y \to Z$ and hence an F-

based DDAG for $Y \rightarrow Z$. Choose the Z in $e_F(X)$ with a DDAG H for $Y \rightarrow Z$ with the least number of nodes. Suppose there is an FD $T \rightarrow S$ in $E_F(X)$ used in H. Then H is a DDAG for $Y \rightarrow T$, and furthermore, there is some node in H labeled by an attribute of S that can be removed and still leave a DDAG for $Y \rightarrow T$. If H' is H with this node removed, the minimality of H is contradicted, since $T \in e_F(X)$. Thus there are no FDs from $E_F(X)$ in $U(H)$ and $Y \dot{\rightarrow} Z$. □

LEMMA 7. *If F is minimum, there are no distinct FDs $Y \rightarrow Q$ and $Z \rightarrow R$ in $E_F(X)$ such that $Y \dot{\rightarrow} Z$.*

PROOF. Suppose H is an F-based DDAG for $Y \rightarrow Z$ using no FDs in $E_F(X)$. Form F' by replacing the two FDs $Y \rightarrow Q$ and $Z \rightarrow R$ by $Z \rightarrow QR$. The FD $Y \rightarrow Z$ can still be derived in F', since none of the FDs in $U(H)$ have been altered. However, F' has one fewer FD than F but the same closure, a contradiction. □

Lemmas 6 and 7 are the tools needed to show the following property of minimum covers. Let $|S|$ denote the cardinality of a set S.

THEOREM 1. *Given equivalent minimum sets of FDs F and G, $|E_F(X)| = |E_G(X)|$ for any X. Thus the size of the equivalence classes in $\bar{E}_F$ is the same for all minimum F with the same closure.*

PROOF Let $m < n$, and let $E_F(X)$ and $E_G(X)$ be composed as shown below.

$E_F(X)$	$E_G(X)$
$X_1 \rightarrow \bar{X}_1$	$Y_1 \rightarrow \bar{Y}_1$
$X_2 \rightarrow \bar{X}_2$	$Y_2 \rightarrow \bar{Y}_2$
$\vdots$	$\vdots$
$X_m \rightarrow \bar{X}_m$	$Y_n \rightarrow \bar{Y}_n$

Some Y_j is not the same as some X_i, or two Y_j's would be equal, contradicting Lemma 7. Thus there exists a j such that $Y_j \neq X_i$, $1 \leq i \leq m$. By Lemma 6 there exists a k such that $Y_j \dot{\rightarrow} X_k$. Renumber the FDs in the two equivalence classes so that $Y_1 \dot{\rightarrow} X_1$. In $E_G(X)$ (and G itself) replace $Y_1 \rightarrow \bar{Y}_1$ with $X_1 \rightarrow \bar{Y}_1$. Note that $Y_1 \rightarrow \bar{Y}_1$ can still be derived from G, since $Y_1 \rightarrow X_1$ and $X_1 \rightarrow \bar{Y}_1$ can both still be derived. If $X_1 = Y_j$ for some $j > 1$, $X_1 \rightarrow \bar{Y}_1$ can be combined with $Y_j \rightarrow \bar{Y}_j$ to form $X_1 \rightarrow \bar{Y}_1\bar{Y}_j$, which shows that G is not minimum. If $X_1 \neq Y_j$ for all $j > 1$, the number of left sides of FDs in $E_G(X)$ that match left sides of FDs in $E_F(X)$ has increased. (We removed Y_1 and added X_1.) By the remarks at the beginning of the proof, there is still a j such that $Y_j \neq X_i$, $1 \leq i \leq m$, and a k such that $Y_j \dot{\rightarrow} X_k$. Repeat the preceding process from where the renumbering took place.

If we never encounter a contradiction to the minimality of G, eventually all the X_i's must become left sides of FDs in $E_G(X)$, contradicting our opening remark. Therefore the assumption that $m < n$ must be incorrect, and in fact $m = n$. □

Lemma 1 implies that for any equivalent nonredundant sets F and G, $|\bar{E}_F| = |\bar{E}_G|$. Theorem 1 goes a step further and shows that if F and G are minimum, not only is the number of classes of FDs with equivalent left sides the same in each set, but the sizes of corresponding classes are the same. The correspondence goes one step further. Let F and G both be minimum, and look at $E_F(X)$ and $E_G(X)$.

$E_F(X)$	$E_G(X)$
$X_1 \rightarrow \bar{X}_1$	$Y_2 \rightarrow \bar{Y}_1$
$X_2 \rightarrow \bar{X}_2$	$Y_2 \rightarrow \bar{Y}_2$
$\vdots$	$\vdots$
$X_m \rightarrow \bar{X}_m$	$Y_m \rightarrow \bar{Y}_m$

Every X_i directly determines some Y_j, and this Y_j directly determines some X_k by Lemma 6 (recalling that Lemma 4 states direct determination is independent of the choice of nonredundant cover). If $i \neq k$, since $X_i \xrightarrow{\cdot} Y_j$, $Y_j \xrightarrow{\cdot} X_k$, and $Y_j \rightarrow X_i$, we can apply Lemma 5 to get $X_i \xrightarrow{\cdot} X_k$, which contradicts Lemma 7. Hence $i = k$. It follows that for every X_i in $e_F(X)$ there is exactly one Y_j in $e_G(X)$ such that $X_i \xrightarrow{\cdot} Y_j$ and $Y_j \xrightarrow{\cdot} X_i$. This relationship allows X_i to be substituted for Y_j without changing the closure of G, and Y_j for X_i in F, since one left side can still be derived from the other after the substitution. For example, if A is social security number and B is student number, then $A \leftrightarrow B$. Whenever we have a left side of the form AX we can replace it with BX, and vice versa, since $AX \xrightarrow{\cdot} BX$ and $BX \xrightarrow{\cdot} AX$, provided A does not determine X.

The observations above show how to combine two equivalent minimum sets F and G to get possibly a new minimum cover for both with fewer attribute symbols than either. Suppose G has no more attribute symbols than F. Start with a pair of corresponding equivalence classes, $E_F(X) = \{X_1 \rightarrow \bar{X}_1, X_2 \rightarrow \bar{X}_2, \ldots, X_m \rightarrow \bar{X}_m\}$ and $E_G(X) = \{Y_1 \rightarrow \bar{Y}_1, Y_2 \rightarrow \bar{Y}_2, \ldots, Y_m \rightarrow \bar{Y}_m\}$, and number the FDs so that X_i and Y_i directly determine each other Modify $E_G(X)$ by substituting X_i for Y_i whenever X_i is smaller than Y_i. The new G will have no more attribute symbols than the old G, and possibly fewer. The next section demonstrates that this combination can be done in polynomial time.

Suppose now that G is only nonredundant, but F is still minimum, and that $|E_F(X)| < |E_G(X)|$ for some X. Say the FDs in $E_G(X)$ go up to $Y_n \rightarrow \bar{Y}_n$, $n > m$. There must be Y_j and Y_k, $j \neq k$, in $e_G(X)$ such that $Y_j \xrightarrow{\cdot} X_i$ and $Y_k \xrightarrow{\cdot} X_i$ for some X_i in $e_F(X)$. In turn, $X_i \xrightarrow{\cdot} Y_h$ for some Y_h in $e_G(X)$. Either $h \neq j$ or $h \neq k$. Assume the first case, and apply Lemma 5 to get $Y_j \xrightarrow{\cdot} Y_h$, $j \neq h$. In the second case, $Y_k \xrightarrow{\cdot} Y_h$, $k \neq h$. We have proved the following result

THEOREM 2. *Let F be minimum and G be nonredundant, with $F \equiv G$. For any $E_G(X)$ with more FDs than $E_F(X)$, there are Y_i and Y_j in $e_G(X)$ belonging to different FDs, with $Y_i \xrightarrow{\cdot} Y_j$.*

The existence of Y_i and Y_j means that G can be improved by replacing $Y_i \rightarrow \bar{Y}_i$ and $Y_j \rightarrow \bar{Y}_j$ with $Y_j \rightarrow \bar{Y}_i\bar{Y}_j$. Furthermore, we need not know F to make this improvement. This result is very important in the next section.

COROLLARY. *An optimal set of FDs is LR-minimum.*

PROOF. If a set of FDs G were optimal but not minimum, it could be shortened as described in the preceding paragraph, since G must be nonredundant. If G has superfluous attribute symbols on the right or left sides of its FDs, it is not optimal. Hence G is LR-minimum. □

4. *Complexity Results*

Beeri and Bernstein [5] present a membership algorithm that determines in linear time if $X \rightarrow Y$ is in G^+, given a set of FDs G and an FD $X \rightarrow Y$. (All complexity results are for the RAM model [3].) The algorithm actually finds all FDs $W \rightarrow Z$ such that $X \rightarrow W$ and all attributes A such that $X \rightarrow A$ is in G^+. The set of all such A's is called the *closure* of X and written X^+. Beeri and Bernstein also give an $O(np)$ algorithm for finding a nonredundant cover for G, where n is the length of G (in attribute symbols) and p is the number of FDs in G.

The membership and nonredundant cover algorithms can be used to decide direct determination: Given a set of FDs G and an FD $X \rightarrow Y$, does $X \xrightarrow{\cdot} Y$? Direct determination can be tested in $O(np)$ time:

DIRECT(G, $X \rightarrow Y$)

1. Find a nonredundant cover for G [$O(np)$]
2 Determine $e_F(X)$. First find X^+ [$O(n)$]. Then for every FD in F with left side Z contained in X^+, determine if $Z \rightarrow X$ is in F^+. If so, Z is in $e_F(X)$ [$O(np)$]

3 Run the membership algorithm on $F - E_F(X)$, $X \to Y$ The FDs in $E_F(X)$ can be marked while finding $e_F(X)$ in step 2 If $X \to Y$ is in the closure of $F - E_F(X)$, output "yes" and stop, otherwise output "no" and stop $[O(n)]$

A test can be incorporated before step 1 to determine if $X \to Y$ is in G^+. If not, output "no" and ignore the rest of the procedure.

THEOREM 3. *Given a set of FDs G, finding a minimum cover F for G can be done in O(np) time.*

PROOF. Lemma 7 and Theorem 2 together say that a nonredundant cover F is minimum if and only if there are no FDs $X \to \bar{X}$ and $Y \to \bar{Y}$ in F such that $X \leftrightarrow Y$ and $X \dot{\to} Y$. Furthermore, if such a pair of FDs exists in F, we can reduce the size of F by replacing the pair with $Y \to \bar{X}\bar{Y}$. Thus the minimum cover algorithm proceeds by finding such pairs of FDs in G and replacing them with a single FD until no more pairs remain.

MINIMIZE(G)

1 Find a nonredundant cover F for G
2 Determine all the classes in $\bar{E}_F$
3 For each class $E_F(X)$ in $\bar{E}_F$,
   for each $Y \to \bar{Y}$ in $E_F(X)$,
      compute Y^+ under $F - E_F(X)$ If there is a $Z \to \bar{Z}$ in $E_F(X)$ with Z in Y^+, remove $Y \to \bar{Y}$ from F and add $\bar{Y}$ to the right side of $Z \to \bar{Z}$
4 Output F

Finding F takes $O(np)$ time. Finding the equivalence classes in $\bar{E}_F$ might seem to require $O(np^2)$ time, since for each pair of FDs $X \to \bar{X}$ and $Y \to \bar{Y}$ in F we need to test if $X \leftrightarrow Y$ under F. However, in one run of Beeri and Bernstein's membership algorithm, for a given X, we can mark every FD $Y \to \bar{Y}$ in F such that $X \to Y$. In $O(np)$ time we can run the membership algorithm on the left side of every FD in F to produce a $p \times p$ (at most) Boolean matrix M with rows and columns indexed by FDs in F. The entry $M[X \to \bar{X}, Y \to \bar{Y}]$ equals 1 if $X \to Y$ is in F^+; it equals 0 otherwise. From M it is possible to find all the sets in $\bar{E}_F$ in $O(p^2)$ time.

For step 3, for each $Y \to \bar{Y}$ in $E_F(X)$, a similar use of the membership algorithm can mark every FD $Z \to \bar{Z}$ such that $Y \to Z$ is in $(F - E_F(X))^+$. That is, $Y \dot{\to} Z$. The membership algorithm is run at most once for each FD in F, giving $O(np)$ time complexity for step 3. Since no step of MINIMIZE takes more than $O(np)$ time, the complexity of the entire algorithm is $O(np)$. □

COROLLARY. *Given a set of FDs G, L-minimum and LR-minimum covers for G can be found in $O(n^2)$ time.*

PROOF. First find a minimum cover F for G in $O(np)$ time. Beeri and Bernstein give an $O(n^2)$ procedure for removing extraneous attributes from left sides of FDs [5]. Applying this procedure makes F L-minimum. To make F LR-minimum, remove extraneous attributes from right sides, as follows:

Take $X \to Y$ in F Suppose $Y = B_1 B_2 \cdots B_m$ Let F' be F with $X \to Y$ replaced by $X \to (Y - \{B_i\})$ Test if $X \to Y$ is in the closure of F' If so, let $F = F'$ Repeat this process for each B_i in Y and all FDs in F $[O(n^2)]$.

To see that the above process works correctly, we must prove that after removing extraneous attributes from right sides of FDs, no new attributes are made extraneous on left sides.

Suppose after eliminating extraneous attributes from right sides there is an FD $X \to Y$ in F with extraneous attribute A in X. Let F' be the version of F immediately after removing extraneous attributes from left sides of FDs. Assume that $X \to Y$ comes from $X \to YZ$ in F'. Let $X' = X - A$. Since A is extraneous in X, $F - \{X \to Y\} \cup \{X' \to Y\}$

$\equiv F$, so $X' \rightarrow Y$ is in F^+. Let H be an F-based DDAG for $X' \rightarrow Y$. If $X \rightarrow Y$ is not in $U(H)$, $X \rightarrow Y$ is redundant in F, contradicting the minimality of F. Therefore $X \rightarrow Y$ is in $U(H)$ and $X' \rightarrow X$ is in F^+ by Lemma 2. Since $F' \equiv F$, $X' \rightarrow X$ is in $(F')^+$. Clearly, $X' \rightarrow X$ can be derived from F' without using $X \rightarrow YZ$. It follows that $F' - \{X \rightarrow YZ\} \cup \{X' \rightarrow YZ\} \equiv F'$. We see that F' is not L-minimum, a contradiction. □

The *yes/no minimum cover problem* is: Given a set of FDs G and an integer k, is there a cover F for G with no more than k FDs? A theorem of Bernstein [6] states that the above problem is NP-complete. However, we have not shown that P = NP. What Bernstein actually proved is that the *yes/no contained cover problem* is NP-complete. The contained cover problem is the minimum cover problem with the added restriction that F is contained in G.

An analogous situation to the minimum cover and contained cover problems arises with a pair of graph problems. A *transitive reduction* of a directed graph H is a graph J with fewest nodes that has the same transitive closure as H. This problem is solvable in polynomial time [2]. A *minimum equivalent graph* of a directed graph H is a *subgraph* J of H with fewest nodes that has the same transitive closure as H. Sahni shows that finding the size of a minimum equivalent graph is NP-complete [16]. The analogy is not surprising, for the transitive reduction and minimum equivalent graph problems are special cases of the minimum cover and contained cover problems. (All FDs have single attributes on both the right and left sides.) Indeed, Bernstein uses the minimum equivalent graph problem to obtain his result.

The *optimal cover problem* is the same as the minimum cover problem, except that F must have fewer than k attribute symbols (rather than FDs) This is most likely a much harder problem.

THEOREM 4. *The optimal cover problem is NP-complete.*

PROOF. Given a set of FDs G and a set of attributes X, a *key* for X is a subset Y of X such that $Y \rightarrow X$ is in G^+, but not $Y' \rightarrow X$, for any Y' properly contained in Y. Simply, a key is a minimal subset of X that functionally determines X. Lucchesi and Osborn [12] show that the following *key of cardinality k problem* is NP-complete. Given a set of FDs G and an integer k, let X be the set of all attribute symbols in G. Does X have a key of cardinality no larger than k?

We can solve the key of cardinality k problem in polynomial time using a polynomial-time algorithm for the optimal cover problem. First we need to prove two claims.

Definition. An FD $X \rightarrow Y$ in G^+ is *reduced if* $X \cap Y = \emptyset$ and for no proper subset X' of X is $X' \rightarrow Y$ in G^+. Let RED(G) be the set of reduced FDs in G^+.

CLAIM 1. *Let G be a set of FDs with attribute symbols X, and let A and B be attribute symbols not in X. Let $G' = G \cup \{AX \rightarrow B\}$. Then*

$$RED(G') = RED(G) \cup \{AY \rightarrow BZ \mid Y \text{ is a key of } X \text{ and } Y \cap Z = \emptyset\}.$$

PROOF. Let $T \rightarrow S \in$ RED(G'), and let H be the smallest G'-based DDAG for $T \rightarrow S$. Consider the following two cases:

(1) $AX \rightarrow B$ is not in $U(H)$. If A is in H it can have no incoming or outgoing edges. Therefore A must belong to T. Since $T \rightarrow S$ is reduced, A is not in S, so A can be removed from both H and T, contradicting the assumption that $T \rightarrow S$ is reduced. Hence A must not be in H or $T \rightarrow S$. Neither is B, by a similar argument, so $T \rightarrow S$ is in RED(G).

(2) $AX \rightarrow B$ is in $U(H)$. Once again A can have no incoming edges, so A is in T. The labeled B has no outgoing edges, so it must be in S or it could be erased from H, which is supposed to be minimal. Let Y be the set of nodes with no incoming edges, except for A. Then $T \rightarrow S$ is actually $AY \rightarrow BZ$ for some Z, $Z \rightarrow Y = \emptyset$. Removing A and B from H yields a DDAG with all the attributes of X. From Lemma 2 we have $Y \rightarrow X$. Since $Y \rightarrow X$ is reduced, Y must be a key of X.

This argument shows containment in one direction. Containment in the other direction is simple. □

Note that any LR-minimum cover contains only reduced FDs.

CLAIM 2. *Let G and G' be as in Claim 1. Then F' is an LR-minimum cover for G' if and only if $F' = F \cup \{AY \rightarrow B\}$, where F is an LR-minimum cover of G and Y is a key of X.*

PROOF. We show only the only if condition. Let F' be given, and let $F = F' \cap \text{RED}(G)$. We must show that F is a cover of G. Let $T \rightarrow S$ be in G^+, and let H be an F'-based DDAG for it. Suppose $U(H)$ contains an FD of the form $AY \rightarrow BZ$. Since A has no incoming edges (no reduced FD has A on the right), A is in T, a contradiction to $T \rightarrow S$ being in G^+. Hence H is also an F-based DDAG for $T \rightarrow S$, and it follows that F is a cover of G. F is easily seen to be LR-minimum by the LR-minimality of F'.

Let F'' be $F' - F$. F'' consists of FDs of the form $AY \rightarrow BZ$. Since F is a cover of G, Claim 1 tells us that Y is a key of X, and hence $Y \rightarrow Z$ in F^+. Since F' is LR-minimum, $Z = \emptyset$, so all the FDs of F'' have the form $AY \rightarrow B$, Y a key of X. Suppose F'' contains $AY_1 \rightarrow B$ and $AY_2 \rightarrow B$, $Y_1 \neq Y_2$. $AY_1 \rightarrow B$ is redundant in F', since $Y_1 \rightarrow X$ and $X \rightarrow Y_2$ in F^+, so $AY_1 \rightarrow B$ can be derived from $F' - \{AY_1 \rightarrow B\}$. Thus F'' contains a single FD. □

PROOF OF THEOREM 4 (CONT.). We want to find if X has a key of cardinality no greater than k under G. Let $G' = G \cup \{AX \rightarrow B\}$ for A, B not in X. Use repeated applications of the optimal cover algorithm to find the size s of an optimal cover for G. Now find the size t of an optimal cover for G'. The sizes of the two covers differ by the number of symbols in $AY \rightarrow B$, where Y is a smallest key of X. Hence $|Y| = t - (s + 2)$.

The argument above shows that the optimal cover problem is NP-hard. It is in NP, since a cover for G can be guessed and checked in polynomial time. □

5. *The Kernel Algorithm*

Lewis et al. [11] have proposed a representation for a set of FDs G that they term the kernel. The *kernel* is a unique canonical form and embodies all nonredundant covers of G. The kernel consists of sets of equivalent left sides that may appear in a nonredundant cover of G, together with a list of possible right-side attributes for each set. The algorithm they present for finding the kernel of G takes exponential time—$O(n^n)$ at least, on inputs of size $n^2\log_2 n$. Such a time complexity hampers the usefulness of the kernel.

The algorithm begins by finding a nonredundant cover for G with no extraneous attributes on left or right sides. The authors use the term minimal for redundant, but they blur the distinction between minimal and minimum. To find the nonredundant cover, the algorithm generates G^+, which can be huge compared to G itself. This step is totally unnecessary, since the LR-minimum cover algorithm can replace the part of the kernel algorithm that finds the nonredundant cover. This change reduces the time complexity of this portion of the algorithm to be polynomial and throws in a minimum rather than nonredundant cover as part of the bargain.

Unfortunately, this change does not create a polynomial-time algorithm. The next part of the kernel algorithm starts by finding all left sides equivalent to left sides in the nonredundant cover for G and leaves all those in the kernel that cannot be derived by augmentation from any of the others. The number of such left sides can be much larger than the number of FDs in G. For example let $G = \{A_i \rightarrow B_i, B_i \rightarrow A_i \mid 1 \leq i \leq m\} \cup \{A_1A_2 \cdots A_m \rightarrow C\}$. The set of G is LR-minimum and has $2m + 1$ FDs The set of left sides equivalent to $A_1A_2 \cdots A_m$ is $\{D_1D_2 \cdots D_m \mid D_i = A_i \text{ or } B_i\}$. This set has 2^m elements. Examples exist where G has m^2 FDs and there are m^m equivalent left sides.

Thus the kernel inherently takes long to compute, since it can be more than exponentially larger than its input G. The kernel algorithm may not even be polynomial in the size of its output because of other steps in the algorithm. Although the kernel is a unique representation of a set G, we maintain it is not a very useful one, since it can take so long to compute and is not necessarily a very succinct representation.

6. *Summary and Further Questions*

We have compared and related different notions of minimality of covers for sets of FDs. Using direct determination, we showed it is possible to find covers with the smallest number of FDs in polynomial time. We also demonstrated that it is unlikely that covers with the smallest number of attribute symbols can be found in polynomial time.

One question, raised in the abstract, is how much the use of a minimum cover improves the run time of various algorithms that use a set of FDs as an input. In the case of relational synthesis algorithms, the use of minimum covers instead of nonredundant covers can improve the database scheme synthesized [13, 14]. The use of minimum covers in connection with the tableau modification algorithm of Aho et al. [1] should also be investigated. Finding optimal covers is NP-complete, but LR-minimality takes us part of the way there by giving a necessary condition for optimality. What bound can be placed on the ratio of the size of an optimal cover to the size of a LR-minimum cover? This paper mainly deals with equivalence and transformations of left sides of FDs. What sort of transformations can be found for right sides?

ACKNOWLEDGMENTS. The authors thanks Jeff Ullman for his comments on an earlier version of this paper and Catriel Beeri for an improved definition of DDAG and shorter proofs of Lemma 4, Lemma 6, and Theorem 4.

REFERENCES

(Note. References [8, 9] are not cited in the text)

1 AHO, A V , BEERI, C , AND ULLMAN, J D. The theory of joins in relational databases *ACM Trans Database Syst 4*, 3 (Sept 1979), 297–314

2 AHO, A V., GAREY, M R , AND ULLMAN, J D The transitive reduction of a directed graph *SIAM J Comput 1*, 2 (June 1972), 131–137

3 AHO, A V , HOPCROFT, J F , AND ULLMAN, J D *The Design and Analysis of Computer Algorithms* Addison-Wesley, Reading, Mass , 1974

4 ARMSTRONG, W W Dependency structures of data base relationships *Proceedings IFIP 1974*, North-Holland, Amsterdam, pp 580–583

5 BEERI, C , AND BERNSTEIN, P A Computational problems related to the design of normal form relational schemas *ACM Trans Database Syst 4*, 1 (March 1979), 30–59.

6 BERNSTEIN, P A Normalization and functional dependencies in the relational data base model Ph D Dissertation, Univ of Toronto, Toronto, Ontario, Canada, November 1975 See also Tech Rep CSRG-60, Computer Systems Research Group, Univ of Toronto, 1975.

7 BERNSTEIN, P A Synthesizing third normal form relations from functional dependencies *ACM Trans Database Syst 1*, 4 (Dec 1976), 277–298

8 CODD, E F A relational model of data for large shared data banks *Commun ACM 13*, 6 (June 1970), 377–387

9 CODD, E F Further normalization of the data base relational model In *Data Base Systems*, R Rustin, Ed., Prentice-Hall, Englewood Cliffs, N J , pp 33–64

10 DATE, C J *An Introduction to Database Systems* Addison-Wesley, Reading, Mass 1975

11 LEWIS, E A , SEKINO, L C , AND TING, P D A canonical representation for the relational schema and logical data independence IEEE COMPSAC '77 Conference, Chicago, Ill , November 1977, pp 276–280

12 LUCCHESI, C L , AND OSBORNE, S L Candidate keys for relations *J Comput Syst Sci 17*, 2 (Oct 1978), 270–279

13 MAIER, D Relational synthesis with annular covers Tech Rep , State Univ of New York at Stony Brook, Stony Brook, N Y., Dec 1979

14 MAIER, D *The Theory of Relational Databases* In preparation

15 PAREDAENS, J About functional dependencies in a database structure and their coverings Rep R 342, Philips M B L E Research Lab , Brussels, Belgium, March 1977

16 SAHNI, S Some related problems from network flows, game theory and integer programming Proc. 13th IEEE Symp on Switching and Automata Theory, College Park, Md , 1972, pp 130–138

17 ULLMAN, J D *Principles of Database Systems* Computer Science Press, Potomac, Maryland, 1980

RECEIVED JANUARY 1980, REVISED APRIL 1980, ACCEPTED MAY 1980