Chapter 5

COVERS FOR FUNCTIONAL DEPENDENCIES

In this chapter we shall explore methods to represent sets of FDs succinctly. For example, any FD implied by the set $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C, A B \rightarrow C, A \rightarrow B C\}$ is also implied by the set $G = \{A \rightarrow B, B \rightarrow C\}$, since all the FDs in F can be derived from FDs in G.

Why do we want shorter representations? We have already seen two algorithms, SATISFIES in Section 4.1 and MEMBER in Section 4.6, whose running times depend on the size of the set of FDs used as input. A smaller set of FDs guarantees faster execution. We shall see other algorithms with running times dependent upon the number of FDs in the input. FDs are used in database systems to help ensure consistency and correctness. Fewer FDs mean less storage space used and fewer tests to make when the database is modified.

5.1 COVERS AND EQUIVALENCE

Definition 5.1 Two sets of FDs F and G over scheme R are *equivalent*, written $F \equiv G$, if $F^+ = G^+$. If $F \equiv G$, then F is a *cover* for G.

The definition of cover makes no mention of the relative sizes of F and G. However, we shall soon consider restricted types of covers where F will be no larger than G in numbers of FDs.

If $F \equiv G$, then for every FD $X \rightarrow Y$ in G^+, $F \models X \rightarrow Y$, since $F^+ = G^+$. In particular, $F \models X \rightarrow Y$ for every FD $X \rightarrow Y$ in G. We extend our notation for implication to sets of FDs and write this last condition as $F \models G$. Since the definition of equivalent is symmetric in F and G, $F \equiv G$ also implies $G \models F$.

If $F \models G$, then $G \subseteq F^+$, since F^+ includes every FD $X \rightarrow Y$ such that $F \models X \rightarrow Y$. Taking the closure of both sides of the inequality, we get $G^+ \subseteq$

$(F^+)^+ = F^+$ (see Exercises 4.6 and 4.16). Similarly, $G \models F$ implies $G^+ \supseteq F^+$. We have proved the following result.

Lemma 5.1 Given sets of FDs F and G over scheme R, $F \equiv G$ if and only if $F \models G$ and $G \models F$.

Example 5.1 The sets $F = \{A \rightarrow B\,C,\ A \rightarrow D,\ C\,D \rightarrow E\}$ and $G = \{A \rightarrow B\,C\,E,\ A \rightarrow A\,B\,D,\ C\,D \rightarrow E\}$ are equivalent. F is not equivalent to the set $G' = \{A \rightarrow B\,C\,D\,E\}$ since $G' \not\models C\,D \rightarrow E$.

Lemma 5.1 provides a simple means to test equivalence for two sets of FDs. The function DERIVES in Algorithm 5.1 tests whether $F \models G$.

Algorithm 5.1 DERIVES
Input: Two sets of FDS F and G.
Output: *true* if $F \models G$, *false* otherwise.

```
DERIVES(F, G)
  begin
  v := true;
  for each FD X → Y in G do
    v := v and MEMBER(F, X → Y);
  return(v)
  end.
```

The function EQUIV in Algorithm 5.2 tests the equivalence of two sets of FDs.

Algorithm 5.2 EQUIV
Input: Two sets of FDs F and G.
Output: *true* if $F \equiv G$, *false* otherwise.

```
EQUIV(F, G)
  begin
  v := DERIVES(F, G) and DERIVES(G, F);
  return(v)
  end.
```

5.2 NONREDUNDANT COVERS

Definition 5.2 A set F of FDs is *nonredundant* if there is no proper subset F' of F with $F' \equiv F$. If such an F' exists, F is redundant. F is a *nonredundant cover* for G if F is a cover for G and F is nonredundant.

Example 5.2 Let $G = \{AB \to C, A \to B, B \to C, A \to C\}$. $F = \{AB \to C, A \to B, B \to C\}$ is equivalent to G but redundant, since $F' = \{A \to B, B \to C\}$ is also a cover for G. F' is a nonredundant cover for G.

An alternative characterization of nonredundancy is that F is nonredundant if there is no FD $X \to Y$ in F such that $F - \{X \to Y\} \models X \to Y$ (see Exercise 5.3). Call an FD $X \to Y$ in F *redundant in* F if $F - \{X \to Y\} \models X \to Y$. This alternative characterization provides the basis for the redundancy test for F given in Algorithm 5.3.

Algorithm 5.3 REDUNDANT
Input: A set of FDs F.
Output: *true* if F is redundant, *false* otherwise.
REDUNDANT(F)

```
begin
v := false;
for each FD X → Y in F do
    if MEMBER(F − {X → Y}, X → Y) then v := true;
return(v)
end.
```

For any set of FDs G, there is some subset F of G such that F is a nonredundant cover for G. If G is nonredundant, $F = G$. If G is redundant, then there is an FD $X \to Y$ in G that is redundant in G. Let $G' = G - \{X \to Y\}$, and note $(G')^+ = G^+$. If G' is redundant, there is an FD $W \to Z$ that is redundant in G'. Let $G'' = G' - \{W \to Z\}$; $(G'')^+ = (G')^+ = G^+$. This process of removing redundant FDs must terminate eventually. The result is a nonredundant cover F for G. This process is the basis for the algorithm NONREDUN, Algorithm 5.4, which computes a nonredundant cover for a set of FDs.

Algorithm 5.4 NONREDUN
Input: A set G of FDs.
Output: A nonredundant cover for G.
NONREDUN(G)

```
begin
F := G;
for each FD X → Y in G do
    if MEMBER(F − {X → Y}, X → Y) then
        F := F − {X → Y};
return(F)
end.
```

Example 5.3 Let $G = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, A \rightarrow C\}$. The result of NONREDUN(G) is $\{A \rightarrow B, B \rightarrow A, A \rightarrow C\}$. If G is presented in the order $\{A \rightarrow B, A \rightarrow C, B \rightarrow A, B \rightarrow C\}$, the result of NONREDUN(G) is $\{A \rightarrow B, B \rightarrow A, B \rightarrow C\}$.

Example 5.3 shows that a set G of FDs can contain more than one nonredundant cover. There can also be nonredundant covers for G that are not contained in G. $F = \{A \rightarrow B, B \rightarrow A, A\,B \rightarrow C\}$ is a nonredundant cover for the set G in Example 5.3.

5.3 EXTRANEOUS ATTRIBUTES

If F is a nonredundant set of FDs, there are no "extra" FDs in F, and in this sense F cannot be made smaller by removing FDs. Removing any FD from F would give a set of FDs that was not equivalent to F. However, it may be possible to reduce the size of F by removing attributes from FDs in F.

Definition 5.3 Let F be a set of FDs over scheme R and let $X \rightarrow Y$ be an FD in F. Attribute A in R is *extraneous in* X $\rightarrow$ Y with respect to F if

1. $X = A\,Z$, $X \neq Z$, and $(F - \{X \rightarrow Y\}) \cup \{Z \rightarrow Y\} \equiv F$, or
2. $Y = A\,W$, $Y \neq W$, and $(F - \{X \rightarrow Y\}) \cup \{X \rightarrow W\} \equiv F$.

The definition says that A is extraneous in $X \rightarrow Y$ if A can be removed from the left side or right side of $X \rightarrow Y$ without changing the closure of F.

Example 5.4 Let $G = \{A \rightarrow B\,C, B \rightarrow C, A\,B \rightarrow D\}$. Attribute C is extraneous in the right side of $A \rightarrow B\,C$ and attribute B is extraneous in the left side of $A\,B \rightarrow D$.

Definition 5.4 Let F be a set of FDs over scheme R and let $X \rightarrow Y$ be in F. $X \rightarrow Y$ is *left-reduced* if X contains no attribute A extraneous in $X \rightarrow Y$. $X \rightarrow Y$ is *right-reduced* if Y contains no attribute A extraneous in $X \rightarrow Y$. $X \rightarrow Y$ is *reduced* if it is left-reduced and right-reduced, and $Y \neq \emptyset$. A left-reduced FD is also called a *full* FD.

Definition 5.5 A set F of FDs is *left-reduced* (*right-reduced, reduced*) if every FD in F is left-reduced (respectively, right-reduced, reduced).

Example 5.5 The set $G = \{A \rightarrow B\,C, B \rightarrow C, A\,B \rightarrow D\}$ is neither left-reduced nor right-reduced. $G_1 = \{A \rightarrow B\,C, B \rightarrow C, A \rightarrow D\}$ is left-reduced but not right-reduced, while $G_2 = \{A \to B, B \to C, A B \to D\}$ is right-reduced but not left-reduced. The set $G_3 = \{A \to B, B \to C, A \to D\}$ is left- and right-reduced, hence reduced, since no right side is $\emptyset$.

We might imagine that we can compute reduced covers for a set G in a manner similar to NONREDUN: look for extraneous attributes and remove them. However, whether we reduce left sides or right sides of FDs first makes a difference. Reducing right sides first will not work. The set of FDs $G = \{B \to A, D \to A, B A \to D\}$ is right-reduced. Removing extraneous attributes from left sides yields the set $F = \{B \to A, D \to A, B \to D\}$, which is not right-reduced. Therefore, we will reduce left sides before right sides.

There is a problem, however, if G contains a redundant FD, say $X \to Y$. Every attribute in Y is extraneous, and eliminating them all leaves $X \to \emptyset$. It might seem we could save ourselves work if we first eliminate all redundant FDs from G before removing extraneous attributes. Unfortunately, even if we start with a nonredundant cover, we can run into the problem just described (see Exercise 5.7). Hence, as the last step in producing a reduced cover, we must remove any FD of the form $X \to \emptyset$.

Before we write an algorithm to find reduced covers, let us show that if we first remove all extraneous attributes from left sides of FDs and then from right sides of FDs, we are left with no extraneous attributes anywhere, as long as we discard FDs of the form $X \to \emptyset$.

Suppose we start with a nonredundant set of FDs G and produce an equivalent set of FDs F by removing extraneous attributes, first from left sides and then from the right sides of FDs in G. If F is not reduced, it can only be because there is an FD $X \to Y$ in F with $Y \neq \emptyset$ that is not left reduced. Assume A is an extraneous attribute in X. Let G' be G at the point immediately after all extraneous attributes were removed from left sides of FDs in the formation of F. Assume $X \to Y$ comes from $X \to YZ$ in G'. Let $X' = X - A$. Since A is extraneous in X in F, $F - \{X \to Y\} \cup \{X' \to Y\} \equiv F$, so $F \models X' \to Y$. Let H be an F-based DDAG for $X' \to Y$. If $X \to Y$ is not in $U(H)$, then $X \to Y$ is redundant in F and, specifically, $X \to Y$ is not right-reduced, since $Y \neq \emptyset$. Therefore $X \to Y$ is in $U(H)$ and $F \models X' \to X$, by Lemma 4.3. Hence, $G' \models X' \to X$, since $F \equiv G'$. Clearly, $X' \to X$ can be derived from G' without using $X \to YZ$. It follows that $G' - \{X \to YZ\} \cup \{X' \to YZ\} \equiv G'$ and that G' was not left-reduced, a contradiction. We see that F is reduced if FDs of the form $X \to \emptyset$ are excluded.

Notice that $\{X \to Y\} \models X A \to Y$, for any FD $X A \to Y$. Whenever we remove an attribute from a left side in a set of FDs, the result is always a stronger set of FDs. That is, let $G = F \cup \{XA \to Y\}$ and $G' = F \cup \{X \to Y\}$. G' always implies G. To test $G' \equiv G$, we need only test $G \models G'$, which reduces to testing $G \models X \rightarrow Y$. The algorithm LEFTRED in Algorithm 5.5 uses this method to detect extraneous attributes on the left sides of FDs. If G is nonredundant, this test can be simplified to $G \models X \rightarrow XA$ or just $G \models X \rightarrow A$. (Why?)

Algorithm 5.5 LEFTRED
Input: A set of FDs G.
Output: A left-reduced cover for G.
LEFTRED(G)

```
  begin
  F := G;
  for each FD X → Y in G do
    for each attribute A in X do
      if MEMBER(F, (X − A) → Y) then
        remove A from X in X → Y in F;
  return(F)
  end.
```

For removing extraneous attributes from the right sides of FDs, we note that if $G = F \cup \{X \rightarrow YA\}$ and $G' = F \cup \{X \rightarrow Y\}$, then G always implies G'. To test $G' \equiv G$, we only need to test $G' \models G$, which reduces to testing $G' \models X \rightarrow YA$. Since $X \rightarrow Y \in G'$, this test further reduces to $G' \models X \rightarrow A$. The algorithm for right-reduction is given as Algorithm 5.6.

Algorithm 5.6 RIGHTRED
Input: A set of FDs G.
Output: A right-reduced cover for G.
RIGHTRED(G)

```
  begin
  F := G;
  for each FD X → Y in G do
     for each attribute A in Y do
        if MEMBER(F − {X → Y} ∪ {X → (Y − A)}, X → A) then
          remove A from Y in X → Y in F;
  return(F)
  end.
```

We can now easily obtain the algorithm for reduced covers shown as Algorithm 5.7.

Algorithm 5.7 REDUCE
Input: A set G of FDs.
Output: A reduced cover for G.
REDUCE(G)

```
begin
F := RIGHTRED(LEFTRED(G));
remove all FDs of the form X → ∅ from F;
return(F)
end.
```

Example 5.6 Let $G' = \{A \rightarrow C, A\,B \rightarrow D\,E, A\,B \rightarrow C\,D\,I, A\,C \rightarrow J\}$. LEFTRED($G'$) yields $G'' = \{A \rightarrow C, A\,B \rightarrow D\,E, A\,B \rightarrow C\,D\,I, A \rightarrow J\}$ and RIGHTRED(G'') yields $F = \{A \rightarrow C, A\,B \rightarrow E, A\,B \rightarrow DI, A \rightarrow J\}$, which is reduced.

Lemma 5.2 The time complexity of REDUCE is $O(n^2)$ for inputs of length n.

Proof Left to the reader (see Exercise 5.8).

5.4 CANONICAL COVERS

Definition 5.6 A set of FDs F is *canonical* if every FD in F is of the form $X \rightarrow A$ and F is left-reduced and nonredundant.

Since a canonical set of FDs is nonredundant and every FD has a single attribute on the right side, it is right-reduced. Since it is also left-reduced, it is reduced.

Example 5.7 The set $F = \{A \rightarrow B, A \rightarrow C, A \rightarrow D, A \rightarrow E, B\,I \rightarrow J\}$ is a canonical cover for $G = \{A \rightarrow B\,C\,E, A\,B \rightarrow D\,E, B\,I \rightarrow J\}$.

The following lemma relates reduced and canonical covers.

Lemma 5.3 Let F be a reduced cover. Form G by taking each FD $X \rightarrow A_1 A_2 \cdots A_m$ and splitting it into $X \rightarrow A_1, X \rightarrow A_2, \ldots, X \rightarrow A_m$. G is a canonical cover. Conversely, if G is a canonical cover, it is a reduced cover. If we form F by combining all FDs with equal left sides into a single FD, then F is also a reduced cover. In both cases, F and G are equivalent.

Proof Let G be derived from F by splitting FDs. If $X \rightarrow A_i$ is redundant, then A_i is extraneous in $X \rightarrow A_1 A_2 \cdots A_m$. If $X \rightarrow A_i$ has an extraneous attribute B on the left side, then $G \models (X - B) \rightarrow A_i$, which means that $G \models$

$(X - B) \to X$, since $X \to A_i$ is nonredundant. (See the discussion preceding LEFTRED.) It follows that $F \models (X - B) \to X$ and hence B is extraneous in the left side of $X \to A_1 A_2 \cdots A_m$ in F.

The remainder of the proof is left to the reader (see Exercise 5.9).

5.5 THE STRUCTURE OF NONREDUNDANT COVERS

What can be said about two nonredundant covers F and F' for a set G of FDs, other than $F \equiv F'$? The following definition and lemma will point us toward some similarities in structure between F and F'.

Definition 5.7 Two sets of attributes X and Y are *equivalent* under a set F of FDs, written $X \leftrightarrow Y$, if $F \models X \to Y$ and $F \models Y \to X$.

Lemma 5.4 Let F and G be equivalent, nonredundant sets of FDs over scheme R. Let $X \to Y$ be an FD in F. There is an FD $V \to W$ in G with $X \leftrightarrow V$ under F (hence under G).

Proof Consider a G-based DDAG H for $X \to Y$. Look at the FDs in $U(H)$. Each has an F-based DDAG. Some FD $V \to W$ in $U(H)$ must have an F-based DDAG J that uses $X \to Y$. If not, there is an $(F - \{X \to Y\})$-based DDAG for $X \to Y$ and so $X \to Y$ is redundant in F (see Exercise 4.19). Since J uses $X \to Y$, by Lemma 4.3, $F \models V \to X$. Since H uses $V \to W$, $G \models X \to V$, hence $F \models X \leftrightarrow V$.

We may restate Lemma 5.4 as follows. Given equivalent, nonredundant covers F and G, for every left side X of an FD in F, there is an equivalent left side V of an FD in G.

Example 5.8 Let $F = \{A \to B\,C,\ B \to A,\ A\,D \to E\}$ and $G = \{A \to A\,B\,C,\ B \to A,\ B\,D \to E\}$. F and G are nonredundant and equivalent to each other. Note that $A \leftrightarrow A$, $B \leftrightarrow B$, and $A\,D \leftrightarrow B\,D$.

For a set of FDs F over scheme R and a set $X \subseteq R$, let $E_F(X)$ be the set of FDs in F with left sides equivalent to X. Let $\bar{E}_F$ be the set

$$\{E_F(X) \mid X \subseteq R \text{ and } E_F(X) \neq \emptyset\}.$$

$E_F(X)$ is empty when no left side of any FD in F is equivalent to X. $\bar{E}_F$ is always a partition of F.

Given equivalent, nonredundant sets F and G, Lemma 5.4 implies that

$E_F(X)$ is non-empty exactly when $E_G(X)$ is. Therefore, the number of sets in $\bar{E}_F$ is the same as the number in $\bar{E}_G$.

Example 5.9 Let F and G be as in Example 5.8. Then $\bar{E}_F$ is

$$E_F(A) = \{A \rightarrow B\ C,\ B \rightarrow A\}$$
$$E_F(A\ D) = \{A\ D \rightarrow E\},$$

and $\bar{E}_G$ is

$$E_G(A) = \{A \rightarrow A\ B\ C,\ B \rightarrow A\}$$
$$E_G(A\ D) = \{B\ D \rightarrow E\}.$$

5.6 MINIMUM COVERS

A nonredundant cover of a set G of FDs does not necessarily have as few FDs as any cover for G (see Exercise 5.15). This fact prompts the following definition.

Definition 5.8 A set of FDs F is *minimum* if F has as few FDs as any equivalent set of FDs.

A minimum set of FDs can have no redundant FDs (why?), so it is also nonredundant.

Example 5.10 The set $G = \{A \rightarrow B\ C,\ B \rightarrow A,\ A\ D \rightarrow E,\ B\ D \rightarrow I\}$ is nonredundant but not minimum, since $F = \{A \rightarrow B\ C,\ B \rightarrow A,\ A\ D \rightarrow E\ I\}$ is equivalent to G but has fewer FDs. F is a minimum cover for G.

5.6.1 Direct Determination

Unlike nonredundant covers, the definition of minimum covers provides no guide for finding minimum covers or even for testing minimality. In this section we introduce a restricted form of functional determination that gives us the means to compute minimum covers.

Definition 5.9 Given a set of FDs G, X *directly determines* Y under G, written $X \stackrel{\cdot}{\rightarrow} Y$, if there is a nonredundant cover F for G with an F-based DDAG H for $X \rightarrow Y$ such that $U(H) \cap E_F(X) = \emptyset$.

In other words, we can find a nonredundant cover F for G in which $X \rightarrow Y$ can be derived using only FDs in $F - E_F(X)$. Observe that $X \stackrel{\cdot}{\rightarrow} X$ always holds, that $X \dot{\rightarrow} Y$ implies $X \rightarrow Y$, and that $E_F(X)$ can be empty. Also note that $A \dot{\rightarrow} B$ only when $A = B$.

Example 5.11 Let $G = F = \{A \rightarrow C\,D,\ A\,B \rightarrow E,\ B \rightarrow I,\ D\,I \rightarrow J\}$. Then $A\,B \dot{\rightarrow} J$ under G, as the DDAG H in Figure 5.1 shows.

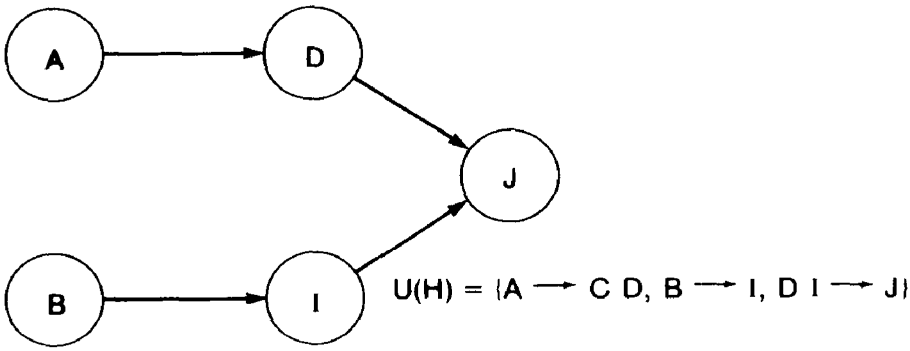

Figure 5.1

As the definition stands, it is not very useful. In order to test direct determination, we might have to find every nonredundant cover of G, which can be a lengthy task (see Exercise 5.11). The following lemma shows that life is not so hard.

Lemma 5.5 $X \dot{\rightarrow} Y$ under a set of FDs G if and only if for *every* nonredundant cover F for G there is an F-based DDAG H for $X \rightarrow Y$ with $U(H) \cap E_F(X) = \emptyset$.

Proof The if direction is trivial. We prove the only if direction. Let F be a nonredundant cover for G where there is an F-based DDAG H for $X \rightarrow Y$ using no FDs from $E_F(X)$. Let F' be any other nonredundant cover for G. For each FD $W \rightarrow Z$ in $U(H)$, we shall construct an F'-based DDAG using no FDs from $E_{F'}(X)$. We shall then splice these DDAGs together to get an F'-based DDAG for $X \rightarrow Y$ using no FDs in $E_{F'}(X)$, using Lemma 4.2.

If $W \rightarrow Z$ is in $U(H)$, then $F \models X \rightarrow W$ by Lemma 4.3. Suppose some F'-based DDAG for $W \rightarrow Z$ uses FD $U \rightarrow V$ from $E_{F'}(X)$. Again using Lemma 4.3, $F' \models W \rightarrow U$, hence $F \models W \rightarrow U$. But $U \leftrightarrow X$ under F' and F, hence $W \leftrightarrow X$ under F, so $W \rightarrow Z$ is in $E_F(X)$, contradicting the nature of H. Therefore every FD $W \rightarrow Z$ in $U(H)$ has an F'-based DDAG using no FDs from $E_{F'}(X)$ (see Exercise 5.19 for a slightly stronger version of Lemma 5.5).

Lemma 5.6 If $X \leftrightarrow Y$, $X \stackrel{\cdot}{\rightarrow} Y$, and $Y \stackrel{\cdot}{\rightarrow} Z$ under a set of G of FDs, then $X \stackrel{\cdot}{\rightarrow} Z$ under G.

Proof Left to the reader (see Exercise 5.20).

Definition 5.10 Let F be a set of FDs over scheme R. The set of all left sides in $E_F(X)$ is denoted by $e_F(X)$.

Lemma 5.7 Let F be a nonredundant set of FDs. Pick X, a left side of some FD in F, and any Y such that $X \leftrightarrow Y$ under F. There exists a Z in $e_F(X)$ such that $Y \stackrel{\cdot}{\rightarrow} Z$.

Proof If Y is in $e_F(X)$, then $Y \stackrel{\cdot}{\rightarrow} Y$ and we are done. Otherwise, since $Y \rightarrow Z$ for every Z in $e_F(X)$, there is an F-based DDAG for $Y \rightarrow Z$ for every Z in $e_F(X)$. Choose the Z in $e_F(X)$ which has a DDAG for $Y \rightarrow Z$ with the smallest number of nodes. Call this DDAG H. Suppose $U(H)$ contains $U \rightarrow V$ from $E_F(X)$. By Lemma 4.3 and its corollary, H is a DDAG for $Y \rightarrow U$, and furthermore, there is a node in H labeled by some attribute in V that can be removed from H and still leave a DDAG for $Y \rightarrow U$. Let H' be H with this node removed. H' has fewer nodes than H. Since U is in $e_F(X)$, the minimality of H is contradicted. There cannot be any FDs from $E_F(X)$ in $U(H)$, so $Y \stackrel{\cdot}{\rightarrow} Z$.

Example 5.12 Let $F = \{A \rightarrow B\ C,\ B\ C \rightarrow A,\ A\ D \rightarrow E,\ A\ D \rightarrow I,\ E \rightarrow B\}$. $B\ C\ D \leftrightarrow A\ D$, and Figure 5.2 shows an F-based DDAG for $B\ C\ D \rightarrow A\ D$ that uses no FDs from $E_F(A\ D)$.

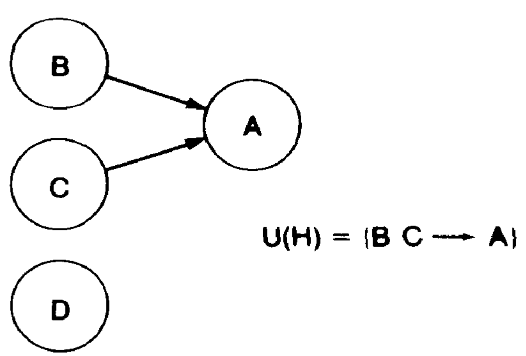

Figure 5.2

Lemma 5.8 Let F be a minimum set of FDs. There are no distinct FDs $Y \rightarrow U$ and $Z \rightarrow V$ in any $E_F(X)$ such that $Y \overset{\cdot}{\rightarrow} Z$.

Proof We shall show that if such FDs exist, we can find a set F' equivalent to F but with fewer FDs. Let H be an F-based DDAG for $Y \rightarrow Z$ that uses no FDs in $E_F(X)$ and let F' be F with $Y \rightarrow U$ and $Z \rightarrow V$ replaced by $Z \rightarrow U\ V$. Clearly, $F' \models Z \rightarrow V$. Since H is also an F'-based DDAG for $Y \rightarrow Z$ (it does not use $Y \rightarrow U$ or $Z \rightarrow V$), $F' \models Y \rightarrow Z$ and hence $F' \models Y \rightarrow U$. Every other FD in F is in F', so $F' \models F$. It is not difficult to show that $F \models F'$ and hence $F \equiv F'$. The minimality of F is contradicted.

Lemmas 5.7 and 5.8 are used to show the following result.

Theorem 5.1 Let F and G be equivalent, minimum sets of FDs. Then for any X, $|E_F(X)| = |E_G(X)|$.

Theorem 5.1 is stronger than Lemma 5.4. Not only does $|\bar{E}_F| = |\bar{E}_G|$ for minimum sets of FDs, but the sizes of corresponding equivalence classes are the same.

Proof Assume $E_F(X)$ and $E_G(X)$ are composed as follows, for m less than n.

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

Not all the Y_j's are the same as some X_i or else two Y_j's would be equal, contradicting Lemma 5.8. Thus there exists j such that $Y_j \neq X_i$, $1 \leq i \leq m$. By Lemma 5.7, there is a k such that $Y_j \overset{\cdot}{\rightarrow} X_k$. Renumber the FDs in $E_F(X)$ and $E_G(X)$ so that Y_j is Y_1 and X_k is X_1. In $E_G(X)$, replace $Y_1 \rightarrow \bar{Y}_1$ by $X_1 \rightarrow \bar{Y}_1$. Since $Y_1 \overset{\cdot}{\rightarrow} X_1$, $Y_1 \rightarrow \bar{Y}_1$ can still be derived in the modified G and the closure of G is unchanged. If $X_1 = Y_j$ for some j other than 1, combine $X_1 \rightarrow \bar{Y}_1$ and $Y_j \rightarrow \bar{Y}_j$ to get $X_1 \rightarrow \bar{Y}_1\ \bar{Y}_j$, which is a contradiction to the minimality of G.

Otherwise, $X_1 \neq Y_j$ for all j greater than 1, but the number of left sides in $e_G(X)$ that match left sides in $e_F(X)$ has increased by one. (We removed Y_1 and added X_1.) There must still be some Y_j not equal to any X_i in $e_F(X)$, by the remarks at the beginning of the proof. We return to the point in the proof where the renumbering took place.

If we never encounter a contradiction to the minimality of G, eventually every left side in $e_G(X)$ will be in $e_F(X)$, contradicting the observation that some Y_j must be different from every X_i. The assumption that m was less than n must be incorrect, and, in fact, $m = n$.

The correspondence between $E_F(X)$ and $E_G(X)$ goes further than there simply being the same number of FDs in each. Consider $E_F(X)$ and $E_G(X)$ again:

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

Choose any X_i in $e_F(X)$. There is some j such that $X_i \overset{\cdot}{\rightarrow} Y_j$, by Lemma 5.7. Also by Lemma 5.7, there must be a k such that $Y_j \overset{\cdot}{\rightarrow} X_k$. If $i \neq k$, then $X_i \overset{\cdot}{\rightarrow} X_k$ by Lemma 5.6, which is a contradiction to Lemma 5.8. Thus, $i = k$. We see that if $X_i \overset{\cdot}{\rightarrow} Y_j$, then $Y_j \overset{\cdot}{\rightarrow} X_i$.

Suppose $X_i \overset{\cdot}{\rightarrow} Y_j$ and $X_i \overset{\cdot}{\rightarrow} Y_h$, where $j \neq h$. By what we have just noted, $Y_j \overset{\cdot}{\rightarrow} X_i$, and by Lemma 5.6, $Y_j \overset{\cdot}{\rightarrow} Y_h$, again contradicting Lemma 5.8. We see there is a one-to-one correspondence between $e_F(X)$ and $e_G(X)$ induced by $\overset{\cdot}{\rightarrow}$. By the proof of Theorem 5.1, we see that X_i can be substituted for its corresponding Y_j in $E_G(X)$ without changing the closure of G.

Example 5.13 It is about time we got away from the As and Bs and got back to an example that resembles real life. Consider a relation *violations*(CAR-SERIAL# LICENSE# OWNER DATE TIME TICKET# OFFENSE) that holds a list of motor vehicle violations. One minimum cover for the set of FDs on this relation is $F =$

1. CAR-SERIAL# → LICENSE# OWNER
2. LICENSE# → CAR-SERIAL#
3. TICKET# → LICENSE# DATE TIME OFFENSE
4. LICENSE# DATE TIME → TICKET# OFFENSE.

An equivalent minimum set of FDs is

1. CAR-SERIAL# → LICENSE#

2. LICENSE# $\rightarrow$ CAR-SERIAL# OWNER
3. TICKET# $\rightarrow$ CAR-SERIAL# OWNER DATE TIME
4. CAR-SERIAL# DATE TIME $\rightarrow$ TICKET# OFFENSE.

E_F(LICENSE#) and E_G(LICENSE#) are composed of the first two FDs in each set, while E_F(TICKET#) and E_G(TICKET#) are composed of the last two FDs in each set. The FDs are arranged so that the left sides of same-numbered FDs directly determine each other. We can substitute CAR-SERIAL# DATE TIME for the left side of FD 4 in F without changing the closure of F. If we try to substitute where there is equivalence but not direct determination, such as LICENSE# for the left side of FD 1 in F, we change the closure.

The observations before the example indicate a means to combine equivalent minimum sets F and G to get an equivalent set of FDs with possibly fewer attribute symbols. Take $E_F(X)$ and $E_G(X)$ for some X and pair up the left sides using the correspondence induced by $\overset{\cdot}{\rightarrow}$. For each Y in $e_F(X)$ and the corresponding Z in $e_G(X)$, replace Y by Z if Z has fewer attributes than Y. If we can make such a substitution, the modified set F will have fewer attribute symbols than the original.

5.6.2 Computing Minimum Covers

The following Theorem will be our tool in developing an algorithm for minimum covers.

Theorem 5.2 Let G be a nonredundant set of FDs that is not minimum. There is some $E_G(X)$ containing distinct FDs $Y \rightarrow U$ and $Z \rightarrow V$ such that $Y \overset{\cdot}{\rightarrow} Z$.

This theorem is almost the converse of Lemma 5.8.

Proof Let F be a minimum cover for G. There must be some X such that $|E_F(X)| < |E_G(X)|$, by Theorem 5.1. Let $E_F(X)$ have FDs $X_1 \rightarrow \bar{X}_1$, $X_2 \rightarrow \bar{X}_2, \ldots, X_m \rightarrow \bar{X}_m$ and let $E_G(X)$ have FDs $Y_1 \rightarrow \bar{Y}_1$, $Y_2 \rightarrow \bar{Y}_2, \ldots, Y_n \rightarrow \bar{Y}_n$. For each Y_j in $e_G(X)$ there is an X_i in $e_F(X)$ with $Y_j \overset{\cdot}{\rightarrow} X_i$, by Lemma 5.7. Since m is less than n, there must be an i, j, and k such that $Y_j \overset{\cdot}{\rightarrow} X_i$ and $Y_k \overset{\cdot}{\rightarrow} X_i$, with $j \neq k$. In turn, $X_i \overset{\cdot}{\rightarrow} Y_h$ for some Y_h in $e_G(X)$. Either $h \neq j$ or $h \neq k$. If $h \neq j$, then by Lemma 5.6, $Y_j \overset{\cdot}{\rightarrow} Y_h$. Likewise, if $h \neq k$, $Y_k \overset{\cdot}{\rightarrow} Y_h$.

Theorem 5.2 says that if we have a nonredundant set G that is not minimum, we can find $Y \rightarrow U$ and $Z \rightarrow V$ in G with $Y \leftrightarrow Z$ and $Y \overset{\cdot}{\rightarrow} Z$ under G. Once we find these two FDs, we can replace them both by the single FD $Z \rightarrow U\ V$, as in the proof of Lemma 5.8. The result is an equivalent set with fewer FDs.

Algorithm 5.8 uses Theorem 5.2 in the manner just described and assumes a function DDERIVES that tests direct determination (see Exercise 5.18).

Algorithm 5.8 MINIMIZE
Input: A set of FDs G.
Output: A minimum cover for G.
MINIMIZE(G)
  begin
  $F :=$ NONREDUN(G);
  Find the sets of $\bar{E}_F$;
  for each $E_F(X)$ in $\bar{E}_F$ **do**
      for each $Y \rightarrow U$ in $E_F(X)$ **do**
          for each $Z \rightarrow V \not\equiv Y \rightarrow U$ in $E_F(X)$ **do**
              if DDERIVES(F, $Y \rightarrow Z$) **then**
                  replace $Y \rightarrow U$ and $Z \rightarrow V$ by $Z \rightarrow U\ V$ in F;
  return(F)
  end.

Theorem 5.3 MINIMIZE can be implemented to have time complexity $O(np)$ on inputs of length n with p FDs.

Proof Finding F takes $O(np)$ time (see Exercise 5.8). Finding the sets in $\bar{E}_F$ might seem to require $O(np^2)$ time, but this much time is not necessary. We can use a modified version of LINCLOSURE to mark, for a given X, every FD $Y \rightarrow Z$ in F such that $F = X \rightarrow Y$. (The marked FDs are those with COUNT $= 0$.) In $O(np)$ time we can run this modified algorithm on the left side of every FD in F to produce a p by p Boolean matrix M with rows and columns indexed by FDs in F. The entry $M[X \rightarrow Y,\ W \rightarrow Z]$ is *true* if $F \models X \rightarrow W$ and *false* otherwise. From M it is possible to find all the sets in $\bar{E}_F$ in $O(p^2)$ time (see Exercise 5.23).

Now, for each $E_F(X)$ in $\bar{E}_F$, look at each FD $Y \rightarrow U$ in turn. Run the modified version of LINCLOSURE on Y and $F - E_F(X)$, but keep track of COUNT[$Z \rightarrow Y$] for each $Z \rightarrow V$ in $E_F(X)$. If the count reaches 0 for some $Z \rightarrow V$ when the algorithm finishes, we know $Y \overset{\cdot}{\rightarrow} Z$ and we make the proper substitution of FDs. The modified LINCLOSURE algorithm is run once for each FD in F, giving $O(np)$ time complexity for this stage. Hence, the entire algorithm takes $O(np)$ time.

Corollary A reduced minimum cover can be found for a set of FDs G in $O(n^2)$ time for inputs of length n.

Proof Apply REDUCE to the output of MINIMIZE(G) (see Exercise 5.8).

5.7 OPTIMAL COVERS

We have been measuring our covers in terms of the number of FDs they contain. We can also measure them by the number of attribute symbols required to express them. For example, $\{A\ B \rightarrow C,\ C\ D \rightarrow E,\ A\ C \rightarrow I J\}$ has size 10 under this measure.

Definition 5.11 A set of FDs F is *optimal* if there is no equivalent set of FDs with fewer attribute symbols than F.

Example 5.14 The set $F = \{E\ C \rightarrow D,\ A\ B \rightarrow E,\ E \rightarrow A\ B\}$ is an optimal cover for $G = \{A\ B\ C \rightarrow D,\ A\ B \rightarrow E,\ E \rightarrow A\ B\}$. Notice that G is reduced and minimum, but not optimal.

Lemma 5.9 If F is an optimal set of FDs, then F is reduced and minimum.

Proof If F has an extraneous attribute, it is clearly not optimal. MINIMIZE always decreases the number of attribute symbols in a cover whenever it makes a change. Thus MINIMIZE(F) must return F and hence F is minimum.

Unfortunately, there is probably no polynomial time algorithm for finding an optimal cover for a set of FDs. This problem belongs to the class of NP-complete problems, for which no one has yet found any polynomial time algorithms. Another NP-complete problem concerning covers is, what is the smallest set F contained in G that is a cover for G? Size in this case is measured in FDs.

5.8 ANNULAR COVERS AND COMPOUND FUNCTIONAL DEPENDENCIES

We have seen that FDs in a set F can be partitioned on the basis of equivalent left sides. It is possible to represent the information in an equivalence class by a single, generalized FD.

Definition 5.12 A *compound functional dependency* (CFD) has the form $(X_1, X_2, \ldots, X_k) \rightarrow Y$, where $X_1, X_2, \ldots, X_k$ are all distinct subsets of a scheme R and Y is also a subset of R. A relation $r(R)$ satisfies the CFD $(X_1, X_2, \ldots, X_k) \rightarrow Y$ if it satisfies the FDs $X_i \rightarrow X_j$ and $X_i \rightarrow Y$, $1 \leq i, j \leq k$. In this CFD, $(X_1, X_2, \ldots, X_k)$ is the *left side*, $X_1, X_2, \ldots, X_k$ are the *left sets* and Y is the *right side*.

A CFD is nothing more than a shorthand way of writing a set of FDs with equivalent left sides. We do make one slight departure from our conventions in allowing $Y = \emptyset$. In this case we write the CFD as $(X_1, X_2, \ldots, X_k)$.

Definition 5.13 Let G be a set of CFDs over R and let F be a set of FDs or CFDs over R. G is *equivalent* to F, written $G \equiv F$, if every relation $r(R)$ that satisfies G satisfies F and vice versa.

This definition is consistent with equivalence for sets of FDs.

Definition 5.14 F is a *cover* for G if $F \equiv G$, where F and G may be either sets of FDs, sets of CFDs, or one set of each.

Example 5.15 The set of CFDs $G = \{(A, B), (A\ C, B\ C) \rightarrow D\ E\}$ is equivalent to the set of FDs $F = \{A \rightarrow B, B \rightarrow A, A\ C \rightarrow D, B\ C \rightarrow E\}$.

Definition 5.15 A set of FDs F is a *characteristic set* for the CFD $(X_1, X_2, \ldots, X_k) \rightarrow Y$, if $F \equiv \{(X_1, X_2, \ldots, X_k) \rightarrow Y\}$. If F uses each left set in the CFD as the left side of an FD exactly once (that is, F looks like $\{X_1 \rightarrow Y_1, X_2 \rightarrow Y_2, \ldots, X_k \rightarrow Y_k\}$), then F is a *natural* characteristic set for the CFD.

The definition of CFD gives us one characteristic set for $(X_1, X_2, \ldots, X_k) \rightarrow Y$, but the set is not natural. Another characteristic set is $\{X_1 \rightarrow X_2, X_2 \rightarrow X_3, \ldots, X_{k-1} \rightarrow X_k, X_k \rightarrow X_1\ Y\}$. This characteristic set is natural, and is the source of the term annular. The left sets in the CFD can be visualized in a ring, as shown in Figure 5.3.

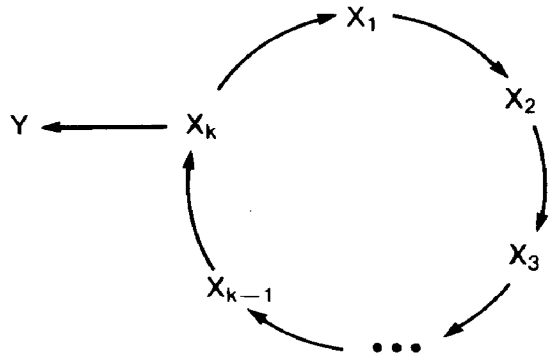

Figure 5.3

A set of CFDs can be treated as the union of characteristic sets for all the CFDs in the set. Treated as such, almost all the terminology from sets of FDs is applicable to sets of CFDs. In some cases we shall redefine our terms slightly for CFDs; when we do not, we use the corresponding definition for FDs. The only concept that does not carry over well is the closure of a set of CFDs. We shall interpret closure as the closure of an equivalent set of FDs.

Definition 5.16 A set F of CFDs is *annular* if there are no left sets X and Z in different left sides with $X \leftrightarrow Z$ under F.

Algorithm LINCLOSURE can be modified to run on sets of CFDs. Rather than keeping counts and lists for FDs, we keep them for left sets. When the count of some left set X_i in $(X_1, X_2, \ldots, X_k) \to Y$ reaches 0, we can add all the attributes in $X_1, X_2, \ldots, X_k$, and Y, to NEWDEP (see Exercise 5.24). MEMBER can therefore be modified to run on CFDs. DERIVES and EQUIV can be modified to work for a set of FDs and a set of CFDs or for two sets of CFDs by choosing characteristic sets for all the CFDs. The time complexity for all these algorithms remains the same (see Exercise 5.25).

Given a set of FDs G, it is possible to find an annular cover for G with no more than $|\bar{E}_F|$ CFDs, where F is a nonredundant cover for G. We combine all the FDs in one $E_F(X)$ into a single CFD. Every left side in $e_F(X)$ is a left set in the CFD, and the right side of the CFD is the union of all the right sides of FDs in $E_F(X)$.

Example 5.16 Let $G = F = \{A \to B\,C,\ B \to A\,D,\ A\,E \to I,\ B\,E \to J\,I\}$. An annular cover for G is the set $G' = \{(A, B) \to A\,B\,C\,D,\ (A\,E, B\,E) \to I\,J\}$.

The reverse process does not always work. If we have a nonredundant cover F for a set of FDs G and an annular cover G' for G with $|\bar{E}_F|$ CFDs, taking the union of natural characteristic sets for all the CFDs in G' does not necessarily yield a nonredundant cover for G.

Example 5.17 $G' = \{(A, AB, B) \rightarrow C\,D, (A\,E) \rightarrow I\,J\}$ is an annular cover for the set G in Example 5.16. When we form F' by combining natural characteristic sets, we can get $F' = \{A \rightarrow A\,B, A\,B \rightarrow B, B \rightarrow A\,C\,D, A\,E \rightarrow I\,J\}$. $A\,B \rightarrow B$ is redundant in F'.

Definition 5.17 Let G be a set of CFDs containing $(X_1, X_2, \ldots, X_k) \rightarrow Y$. Let X_i be one of the left sets and let A be an attribute in X_i. Attribute A is *shiftable* if A can be moved from X_i to Y while preserving equivalence. A left set X_i is *shiftable* if all the attributes in X_i are simultaneously shiftable.

Example 5.18 Let G' be as in Example 5.17. $A\,B$ in $(A, A\,B, B) \rightarrow C\,D$ is shiftable. The result of shifting it is $G'' = \{(A, B) \rightarrow A\,B\,C\,D, (A\,E) \rightarrow I\,J\}$. Note that A in $(A, AB, B) \rightarrow CD$ is not shiftable.

Definition 5.18 An annular set G is *nonredundant* if no CFDs can be removed from G without altering equivalence and no CFD in G contains a shiftable left set. Otherwise, G is redundant.

Example 5.19 The set G' in Example 5.17 is redundant, while G'' in Example 5.18 is nonredundant.

Lemma 5.10 Let G be a nonredundant annular set of CFDs. The union of natural characteristic sets for all the CFDs in G yields a nonredundant set of FDs equivalent to G. The proof is left to the reader (see Exercise 5.26).

We can also define the notions of reduced and minimum for annular covers.

Definition 5.19 Let G be a nonredundant annular set. A CFD $(X_1, X_2, \ldots, X_k) \rightarrow Y$ in G is *reduced* if no left set contains any shiftable attributes and the right side contains no extraneous attributes. The set G is *reduced* if every CFD in G is reduced.

Definition 5.20 Let G be a nonredundant annular set. G is minimum if it contains as few left sets as any equivalent annular set.

Example 5.20 $G = \{(A, B) \to C\,D, (A\,E) \to I\,J\}$ is a reduced, minimum annular cover for the set G' in Example 5.16.

LEFTRED and RIGHTRED can be modified to get a version of REDUCE for CFDs that runs in $O(n^2)$ time on input of length n. To aid in reduction, we can use the observation that $(X_1 \cup X_2 \cup \cdots \cup X_k) \cap Y = \emptyset$ for any reduced CFD$(X_1, X_2, \ldots, X_k) \to Y$ (see Exercise 5.27).

To find a minimum annular cover for a set G of FDs, we first find a minimum cover F for G. We then combine FDs with equivalent left sides into single CFDs. The question arises, is a reduced, minimum annular set really the same as a reduced minimum set of FDs? That is, can we get a reduced, minimum annular set by combining FDs from a reduced, minimum set of FDs? The answer is no, as the next example shows.

Example 5.21 Consider the set of FDs $F = \{B_1\,B_2 \to A,\ D_1\,D_2 \to B_1\,B_2,\ B_1 \to C_1,\ B_2 \to C_2,\ D_1 \to A,\ D_2 \to A,\ A\,B_1\,C_2 \to D_2,\ A\,B_2\,C_1 \to D_1\}$. F is minimum and reduced. The only equivalent left sides are $B_1\,B_2$ and $D_1\,D_2$. Let us combine FDs into CFDs to get $G = \{(B_1\,B_2, D_1\,D_2) \to A,\ (B_1) \to C_1,\ (B_2) \to C_2,\ (D_1) \to A,\ (D_2) \to A,\ (A\,B_1\,C_2) \to D_2,\ (A\,B_2\,C_1) \to D_1\}$. We have left $B_1\,B_2$ off the right side of the first CFD by the observation above that left sets and the right side should not intersect. Even so, the A on the right side of the first CFD is extraneous. It is not extraneous in $B_1\,B_2 \to A$ in F, since it is needed to prove $B_1\,B_2 \to D_1\,D_2$ (see Exercise 5.28).

We see that after converting from a reduced, minimum set of FDs to a minimum annular set, it is still necessary to perform a reduction step to get a reduced, minimum annular set.

We shall use annular covers again in Chapter 6, where we use them for synthesizing database schemes.

5.9 EXERCISES

5.1 Find a nonredundant cover for the set $G = \{A \to C,\ A\,B \to C,\ C \to D\,I,\ C\,D \to I,\ E\,C \to A\,B,\ E\,I \to C\}$.

5.2 Show how DERIVES (Algorithm 5.1) can be modified to run more quickly in some cases.

5.3 Show that a set of FDs F is redundant if and only if there is an FD $X \to Y$ in F such that $F - \{X \to Y\} \models X \to Y$.

5.4 Consider the following alternative to the algorithm NONREDUN (Algorithm 5.4).

```
REPUGNANT(G)
  begin
  F := ∅;
  for each FD X → Y in G do
     if MEMBER (G − {X → Y}, X → Y) then
        F := F ∪ {X → Y};
  return(G − F)
  end.
```

Does REPUGNANT correctly compute a nonredundant cover for G?

5.5 Give an example of a set of FDs that contains an FD $X \rightarrow Y$ with every attribute in X and Y extraneous.

5.6 Find sets of FDs F and G such that F is a nonredundant cover for G, but G has fewer FDs than F.

5.7 Show that starting with a nonredundant set of FDs, removing extraneous attributes from the left sides of FDs can yield a redundant set of FDs.

5.8 Prove that the algorithm NONREDUN (Algorithm 5.4) has time complexity $O(np)$ for inputs of length n with p FDs. Use this result to prove that the algorithm REDUCE (Algorithm 5.7) has time complexity $O(n^2)$ on input of length n.

5.9 Complete the proof of Lemma 5.3.

5.10 Let F be the set of all possible FDs over a relation scheme $R = A_1 A_2 \cdots A_n$, except those of the form $\emptyset \rightarrow Y$. Find a nonredundant cover for F.

5.11* What is the maximum number of nonredundant covers a set of n FDs may contain?

5.12 Show that an FD $X \rightarrow Y$ is redundant in F if and only if there is an F-based DDAG H for $X \rightarrow Y$ with $X \rightarrow Y$ not in $U(H)$.

5.13 Show that Lemma 5.4 can fail if F is redundant.

5.14 Show that for equivalent nonredundant sets of FDs F and G, it is possible that for some X, $E_F(X)$ has a different number of FDs than $E_G(X)$.

5.15 Find two equivalent nonredundant sets of FDs with different numbers of FDs.

5.16 Give a set of FDs G and sets of attributes X and Y such that $X \overset{\cdot}{\rightarrow} Y$ does not hold under G, but $X \rightarrow Y$ does.

5.17 Show that in a minimum set of FDs there cannot be two distinct FDs $X \rightarrow Y$ and $X \rightarrow Z$.

5.18* Find an algorithm to test direct determination.

5.19 Prove: If $X \dot{\rightarrow} Y$ under G, then for any cover F for G there is an F-based DDAG H for $X \rightarrow Y$ with $U(H) \cap E_F(X) = \emptyset$.

5.20 Prove Lemma 5.6.

5.21 Give an example of a set of FDs where some FD has two extraneous attributes, but only one can be removed if equivalence is to be preserved.

5.22 Prove that Theorem 5.1 does not hold if F and G are only nonredundant.

5.23 Given the p by p Boolean matrix M in the proof of Theorem 5.3, show how to find the sets of $\bar{E}_F$ in $O(p^2)$ time.

5.24 Let $(X_1, X_2, \ldots, X_k) \rightarrow Y$ be a CFD and let $S = X_1 \cup X_2 \cup \cdots \cup X_k \cup Y$. Show that $F = \{X_1 \rightarrow S, X_2 \rightarrow S, \ldots, X_k \rightarrow S\}$ is a natural characteristic set for the CFD.

5.25 Show that for any set of CFDs there is an equivalent set of FDs that uses no more than twice the number of attribute symbols.

5.26 Prove Lemma 5.10.

5.27 Show that in a reduced CFD $(X_1, X_2, \ldots, X_k) \rightarrow Y$, $(X_1 \cup X_2 \cup \cdots \cup X_k) \cap Y = \emptyset$.

5.28 Let F be the set of FDs in Example 5.21. Show that A is not extraneous in $B_1 B_2 \rightarrow A$.

5.29 Find a reduced, minimum annular cover for the set G in Exercise 5.1.

5.10 BIBLIOGRAPHY AND COMMENTS

Armstrong [1974] investigated equivalent sets of FDs and alternative representations for sets of FDs. Bernstein [1976b] demonstrated the usefulness of nonredundancy for database normalization (see Chapter 6). Canonical covers were introduced by Paredaens [1977]. Lewis, Sekino, and Ting [1977] examined a representation for all the nonredundant covers of a set of FDs. Lucchesi and Osborn [1978] present some NP-completeness results concerning covers and key finding. Beeri and Bernstein [1979] presented an efficient algorithm for computing a nonredundant cover, and also some NP-completeness results involving FDs. Direct determination and the algorithm MINIMIZE are from Maier [1980b], who also shows that finding optimal covers is NP-complete.

The reader is directed to Garey and Johnson [1979] for background on the theory of NP-complete problems.

Jou [1980] and Steiner [1981] give alternative formalisms for discussing covers and implications of FDs.