Comment on "Decomposition of a Data Base and the Theory of Boolean Switching Functions"

Two of the authors of this letter have presented a theorem (Appendix C of [1], hereinafter called "Theorem C") relating Codd's third normal form for a data collection [2] to the minimum cover for an associated Boolean function. Actually this theorem is incorrect as stated. The third author (Bernstein) in his dissertation [3] has produced a counterexample to the theorem and has developed techniques for decomposing a data base into third normal form under general conditions.

The major results of [1] concerned the mathematical similarity between functional dependencies in a data base and a class of Boolean functions. These results are unaffected by the correction given here. Indeed, the fundamental equivalence has been redemonstrated in a different manner by Armstrong [4] and by Fagin [5]. Theorem C actually concerns a property somewhat stronger than Third Normal Form, since the latter allows a transitive dependency $A \rightarrow B \rightarrow C$ to exist in a relation R if attribute C belongs to a minimal key for R.

Bernstein's work may also be used to correct Theorem C. If his Property P is included in the hypothesis, then the theorem becomes true. Here we present an alternative way of correcting it.

Bernstein's counterexample and a graphical representation of his functional relations (FR's) are shown in Fig. 1 (see [1] for the definition of terms such as "functional relation"). The graph of the counterexample is seen to contain a cycle along the path R_3-X-R_4-C-R_2-B-R_3. This occurrence is related to the failure of Theorem C of [1], as is evidenced by the proof we have included at the end of this letter. If one adds to the hypothesis of the original Theorem C the condition that the graph of the minimum cover be acyclic, then the theorem is true. The property of acyclicity is a stronger condition for third normal form than Bernstein's Property P, although the former is probably easier to check algorithmically for a given set of FR's.

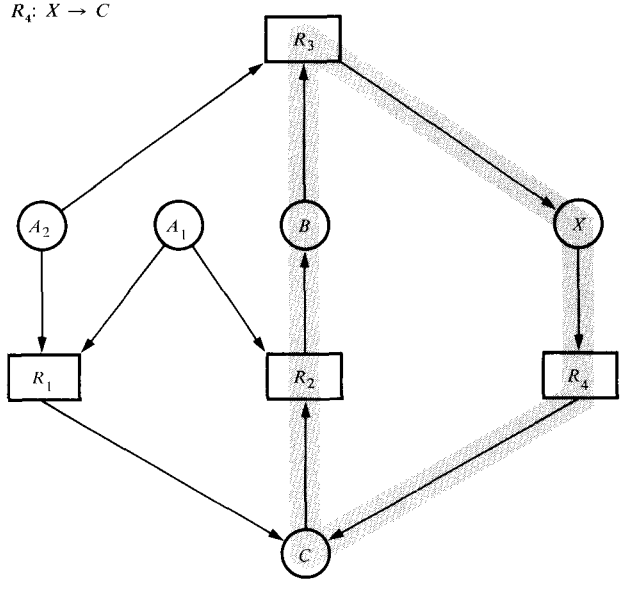
Figure 1 A counterexample. The FR's given in [2], and a graphical representation of them, are shown. Data base attributes are represented by circles, the given functional relations by boxes, and connecting arrows denote the left and right sides of the respective FR's. The existence of a circuit in the graph is indicated by shading.

R_1: $A_1A_2 \rightarrow C$

R_2: $A_1C \rightarrow B$

R_3: $A_2B \rightarrow X$

R_4: $X \rightarrow C$



Theorem If MC is a minimum cover of a set F of functional relations, and if the graph of MC is acyclic, then there is no attribute $X \neq \phi$ such that the following statements are true:

S_1: $A_1 \rightarrow X \in F$;

S_2: $A_2X \rightarrow C \in F$;

S_3: $A_1A_2 \rightarrow C \in MC$,

where A_1, A_2, and C are attributes (or attribute sets) such that

$A_1 \cap A_2 = \phi$;

$X \cap A_2 = \phi$.

Proof Assume the contrary, i.e., there is such an X. We shall show that S_1, S_2, S_3 then imply either that the graph contains a cycle or else that MC is not a minimum cover. This constitutes proof by contradiction.

The following four cases are examined:

1. $S_1, S_2 \in MC$;
2. $S_1 \notin MC$, $S_2 \in MC$,
3. $S_1 \in MC$, $S_2 \notin MC$,
4. $S_1, S_2 \notin MC$.

Case 1 As developed in [1] the FR's S_1-S_3 are not independent. That is, S_1 and S_2 imply S_3. Thus all three cannot belong to the same minimum cover, so this case leads to an immediate contradiction.

Case 2 Since S_1 belongs to F, it must be derivable from the FR's of the minimum cover:

$S_1 = f(V_1, V_2, \cdots, V_n)$, where $V_i \in MC$.

Let the symbol $\oplus$ represent the pseudotransitivity operation as defined in [1]. That is, if we have three functional relations R_1, R_2, and R_3 such that

R_1: $E \rightarrow F$,

R_2: $F, G \rightarrow H$,

R_3: $E, G \rightarrow H$,

then we shall write

$R_1 \oplus R_2 = R_3$.

It is important to note that the closure of a given set of FR's (i.e., the set of all FR's derivable from them) can be obtained by successive application of $\oplus$ to members of the given set. The function f defined above is an instance of such an application.

From the definition of $\oplus$ it follows that

$S_3 = S_1 \oplus S_2$.

Suppose that $S_3 \neq V_i$ for any i. Then

$S_3 = f(V_1, V_2, \cdots, V_n) \oplus S_2$,

which shows that S_3 can be derived from members of MC and thus cannot belong to MC itself, a contradiction.

On the other hand, consider $S_3 = V_i$ for some i. Since $S_2 \in MC$, there is a path from X to C in the graph of MC. But since

$S_1 = f(V_1, V_2, \cdots, V_{i-1}, S_3, V_{i+1}, \cdots, V_n)$,

then there is a path from C to X, i.e., a cycle, contradicting our hypothesis. Cases 3 and 4 are proved the same way.

Acknowledgment

M. Leonard, of the University of Grenoble, assisted in the statement of the acyclic property used to amend Theorem C.

C. Delobel
Univ. of Grenoble
Grenoble, France

R. G. Casey
IBM Research Division laboratory
5600 Cottle Road
San Jose, California 95193

P. A. Bernstein
Aiken Computation Laboratory
Harvard Univ.
Cambridge, Massachusetts 02138

February 15, 1977

References

1. C. Delobel and R. G. Casey, "Decomposition of a Data Base and the Theory of Boolean Switching Functions," *IBM J. Res. Develop.* **17**, 374 (1973).
2. E. F. Codd, "Further Normalization of the Data Base Relation Model," in *Data Base Systems*, Courant Computer Science Symposia Series, Vol. 16, Prentice-Hall, Inc., Englewood Cliffs, NJ 1972.
3. P. A. Bernstein, "Normalization and Functional Dependencies in the Relational Data Base Model," Ph.D. Thesis, University of Toronto, Canada, 1975.
4. W. W. Armstrong, "Dependency Structures of Data Base Relationships," *Proceedings of IFIP '74*, North-Holland Publishing Company, Amsterdam, 1974, p. 580.
5. R. Fagin, "Functional Dependencies in a Relational Database and Propositional Logic," *IBM J. Res. Develop.*, to be published.