Recovery from Inconsistency in Incomplete Relations

Mark Levene
Department of Computer Science
University College London
Gower Street, London WC1E 6BT, U.K.
email: mlevene@cs.ucl.ac.uk

Millist Vincent
Advanced Computing Centre
School of Computing and Information Science
University of South Australia
Adelaide, Australia, 5095
email: millist.vincent@unisa.edu.au

Abstract

It is not always possible to avoid inconsistency of a set of constraints in a database, for example in a legacy database where we are given an out of date set of constraints which may have changed due to maintenance. Thus we need some mechanism to re-assess the set of constraints in order to recover from this inconsistency. We tackle this problem concentrating on the situation when the constraints are a set of functional dependencies (FDs) and the database may be incomplete. The approach we take to alleviate this problem is to consider as the correct semantics the maximally consistent subsets of the given set of FDs in the incomplete relation. We show how to compute all the maximally consistent subsets, although, in general, this computation is shown to be intractable. We then investigate the dependency inference problem for FDs in incomplete relations, which can be viewed as a data mining application of our approach to dealing with inconsistency. Finally, we show that Lien and Atzeni's axiomatisation of FDs in incomplete relations can be given semantics in terms of a refined family of maximally consistent sets of FDs.

Keywords: incomplete relation; functional dependency; inconsistency

1 Introduction

Although the problem of dealing with inconsistent information has been thoroughly investigated within the logic community [Jaś69, KdC78, Gär88, EGH95], it has not been seriously considered by the database community within the context of relational databases. An exception is [BS95] wherein the relational model is generalised to allow both incomplete and inconsistent information based on a four-valued paraconsistent logic but [BS95] do not consider constraints in their model. When considering satisfaction of constraints in a relational database it is normally assumed that at all times the database satisfies the constraints and inconsistent states are rejected during the process of updating the database [Ull88, MR92, AA93]. In practice inconsistencies may arise for several reasons, especially in incomplete information databases. For example, in a legacy database we may be given a set of constraints that are part of the original documentation of the database but it may be the case that the actual database does not satisfy these constraints due to the evolution of the database.

We restrict our investigation to data dependencies and more specifically to *functional dependencies* (FDs) [Ull88, MR92, AA93], which generalise the notion of keys in relational databases. Moreover, we extend relations so that they may model incomplete information by allowing data values to represent the null value, *unk*, denoting a data value which exists but is unknown at the present time. Since each null value in a relation may be replaced by a non-null value, each incomplete relation induces a set of possible complete relations. Thus an FD $\mathrm{X} \rightarrow \mathrm{Y}$ is satisfied in an incomplete relation r if there exists a complete relation emanating from r that satisfies $\mathrm{X} \rightarrow \mathrm{Y}$ in the standard manner [LL97]; (cf. [Lie82, AM86] for an interpretation of FDs in incomplete relations which allows other types of null values). As a further example of when inconsistency may arise, consider the relation r shown in Table 1, over the relation schema {EMP, DEPT, MGR}, with the attributes in the schema having their natural meaning. Suppose that we are given this relation and are asked to infer the FDs that are satisfied in this relation; such a problem can be viewed as a data mining exercise [MR94]. Amongst other FDs, we will be told that EMP $\rightarrow$ DEPT and DEPT $\rightarrow$ MGR are satisfied in r. But it can easily be verified that r is inconsistent with respect to any set of FDs F containing these two FDs, since all possible complete relations emanating from r must violate EMP $\rightarrow$ MGR. (In this case we also say that F is inconsistent in r.) Thus we need to deal with the problem of recovering from this inconsistency in some reasonable manner by re-assessing the set of FDs that are satisfied in this relation.

EMP	DEPT	MGR
John	*unk*	Steve
John	*unk*	Peter

Table 1: An example incomplete relation

Assume that a set of FDs F over R is inconsistent in a relation r over R. The solution we present in this paper to the problem of recovering from inconsistency is to consider the maximal subsets of the closure of F (with respect to Armstrong's axiom system [Ull88, MR92, AA93]) that are consistent in r. The motivation behind this solution is that we consider F to be an over specification of the constraints that must hold in r and therefore we consider the sets of FDs obtained by removing from the closure of F as little as possible, to be the alternative constraint sets that hold in r. The intersection of the closures of the family of all maximally consistent subsets of F in r, called the *core* of F in r, can be viewed as the set of FDs that is indisputably satisfied in r. We note that our approach of taking the maximally consistent subsets as the possible sets of constraints, is compatible with the philosophy of minimal change which is fundamental in belief revision [Gär88].

After formalising the concept of the family of maximally consistent subsets of a set of FDs F in a relation r, we present an algorithm for computing this family. We show that there are, in general, an exponential number of sets of FDs in this family and investigate complexity issues related to finding maximally consistent subsets of a set of FDs in a relation. We also show that finding whether an FD is in the core of F in r is co-NP-complete. Finally, we are able to establish that the problem of finding whether a set G of FDs is a maximally consistent subset of F in r is at least as hard as the *FD relation equivalence problem* [GL90], whose complexity is still an open problem [EG95].

Algorithms for solving the dependency inference problem for FDs [MR94] are important in data mining applications when we have a relation available without knowledge of the data dependencies. When the input to such an algorithm is an incomplete relation, say r, we carry out the dependency inference process in two steps. Firstly we infer the set of FDs, denoted by $dep(r)$, such that for each FD in $dep(r)$ there is a possible complete relation emanating from r that satisfies it. Due to the semantics of incomplete relations $dep(r)$ is not the correct solution to the FDs inference problem, since as was shown in the example above $dep(r)$ may not be consistent in r. So we need to carry out a second step which returns the family of maximally consistent subsets of $dep(r)$ in r as the alternative FDs sets that are inferred from r. We show that if G is maximally consistent in $dep(r)$ then there is a possible complete relation s emanating from r such that G is an equivalent set of FDs to $dep(s)$. In addition, we show that the problem of finding whether a set G of FDs is a maximally consistent subset of $dep(r)$ in r is also at least as hard as the FD relation equivalence problem. We note that it is has already been established that the complexity of discovering $dep(r)$ is at worst of exponential time complexity as a function of the number of attributes in the relation schema under consideration and at best of polynomial time complexity as a function of the cardinality of the input relation [MR94].

We also show an interesting connection between Lien and Atzeni's axiomatisation of FDs in incomplete relations [Lie82, AM86] (see also [LL95]) and maximally consistent sets of FDs. In particular, we show that we can refine the set of maximally consistent subsets of a set of FDs F by choosing an appropriate relation r that satisfies F, so that an FD can be derived from F using Lien and Atzeni's axiom system if and only if there is some maximally consistent subset G of F in r such that the FD can be derived from G using Armstrong's axiom system.

The layout of the rest of the paper is as follows. In Section 2 we formalise the notion of incomplete relations and in Section 3 we extend FDs so that they hold in incomplete relations and give the chase procedure as an algorithm for testing whether a set of FDs F is consistent in an incomplete relation r. In Section 4 we present an algorithm for computing the maximally consistent subsets of a set of FDs F over R in a relation r over R and consider some of the related computational complexity issues. In Section 5 we show how maximally consistent sets of FDs can be used to solve the dependency inference problem for FDs in incomplete relations. In Section 6 we show that the axiomatisation of FDs in incomplete relations presented in [Lie82, AM86] can be given semantics in terms of maximally consistent subsets of FDs. Finally, in Section 7 we give our concluding remarks.

2 Relations that model incomplete information

Herein we extend relations so as to model incomplete information by including in the underlying database domain a null value of the type value at present exists but is unknown.

We use the notation |S| to denote the cardinality of a set S. If S is a subset of T we write S $\subseteq$ T and if S is a proper subset of T we write S $\subset$ T. Furthermore, S and T are *incomparable* if S $\not\subseteq$ T and T $\not\subseteq$ S. At times we denote the singleton {A} simply by A, and the union of two sets S and T, i.e. S $\cup$ T, simply by ST. We will refer to the length of some standard encoding [GJ79] of S as the *size* of S.

Definition 2.1 (Relation schema and relation) A *relation schema* R is a finite set of attributes.

We assume a countably infinite domain of constants, *Dom*, containing a distinguished constant *unk*, denoting the null value unknown.

A tuple over R is a total mapping t from R into *Dom* such that $\forall A_i \in$ R, $t(A_i) \in Dom$. The *projection* of a tuple t over R onto a set of attributes Y $\subseteq$ R, denoted by t[Y], is the restriction of t to Y.

An *incomplete relation* (or simply a relation) over R is a finite set of tuples over R. A relation over R having no occurrences of *unk* is called a *complete relation.*

From now on we let R *be a relation schema and r is a relation over* R. In addition, we let REL(R) denote the countably infinite set of relations over R. As usual uppercase letters (which may be subscripted) from the end of the alphabet such as X, Y, Z will be used to denote sets of attributes, while those from the beginning of the alphabet such as A, B, C will be used to denote single attributes.

In [Lev95, LL97] we defined the semantics of incomplete relations in terms of *possible worlds* by defining a partial order, $\sqsubseteq$ in *Dom*, such that $u \sqsubseteq v$ if and only if either $u = unk$ or $u = v$, where $u, v \in Dom$. The partial order $\sqsubseteq$ is extended to tuples over R in a natural way. The set of all possible worlds relative to r, denoted by POSS(r), is the set of all complete relations that emanate from all possible substitutions of occurrences of *unk* in r by non-null values in $Dom - \{unk\}$.

3 Functional dependencies in relations which may be incomplete

Herein we formalise the notion a functional dependency being satisfied in an incomplete relation and present the well-known chase procedure for testing whether an incomplete relation satisfies a set of FDs.

Definition 3.1 (Functional dependency) A *functional dependency* over R (or simply an FD) is a statement of the form X $\rightarrow$ Y, where X, Y $\subseteq$ R. We call an FD of the form X $\rightarrow$ Y, where Y $\subseteq$ X, a *trivial* FD.

We stress the fact that we allow FDs whose left-hand side is the empty set. We let FD(R) be the set of all FDs over R. From now on we let F *be a set of FDs over* R.

Definition 3.2 (Satisfaction of an FD) An FD X $\rightarrow$ Y is *weakly* satisfied (or simply satisfied) in a relation r, denoted by $r \approx\!\!| \ \mathrm{X} \rightarrow \mathrm{Y}$, whenever $\exists s \in \mathrm{POSS}(r)$ such that $\forall t_1, t_2 \in s$, if $t_1[\mathrm{X}] = t_2[\mathrm{X}]$ then $t_1[\mathrm{Y}] = t_2[\mathrm{Y}]$.

The reader can verify that when the relation, r, in Definition 3.2 is a complete relation then the definition of satisfaction of an FD in r reduces to the standard definition of satisfaction of an FD [Arm74, Ull88, MR92, AA93]. The following proposition shown in [LL97] gives a syntactic characterisation of the satisfaction of an FD.

Proposition 3.1 A relation r over R satisfies an FD X $\rightarrow$ Y if and only if $\forall t_1, t_2 \in r$, if $\forall$ A $\in$ X, $t_1[\mathrm{A}] \neq unk$ and $t_1[\mathrm{X}] = t_2[\mathrm{X}]$ then $\forall$ B $\in$ Y, either $t_1[\mathrm{B}] = unk$, $t_2[\mathrm{B}] = unk$ or $t_1[\mathrm{B}] = t_2[\mathrm{B}]$. $\square$

We now generalise the definition of satisfaction of a single FD to satisfaction of a set of FDs and discuss the semantics of this definition. We denote by SAT(F) the set of all relations $r \in \mathrm{REL(R)}$ such that $\forall$ X $\rightarrow$ Y $\in F, r \approx\!\!| \ \mathrm{X} \rightarrow \mathrm{Y}$.

Definition 3.3 (Satisfaction of a set of FDs) A set of FDs F over R is *weakly* satisfied (or simply satisfied) in a relation r over R, denoted by $r \approx\!\!| \ F$, if and only if $\exists s \in \mathrm{POSS}(r)$ such that $s \in \mathrm{SAT}(F)$. If $r \approx\!\!| \ F$ then we also say that F is *consistent* in r, otherwise we say that F is *inconsistent* in r.

We now define the notion of additive satisfaction which is the central theme investigated in [LL97].

Definition 3.4 (Additive satisfaction of a set of FDs) Satisfaction of a set of FD F over R in a relation r over R is said to be *additive* whenever $r \models\approx F$ if and only if $r \in \mathrm{SAT}(F)$.

The following example shows that, in general, satisfaction of a set of FDs in a relation is *not* additive.

Example 3.1 For example, the well known relation, r_1, shown in Table 2, is such that $r_1 \models\approx \mathrm{A} \rightarrow \mathrm{B}$ and $r_1 \models\approx \mathrm{B} \rightarrow \mathrm{C}$ but $r_1 \not\models\approx F_1 = \{\mathrm{A} \rightarrow \mathrm{B}, \mathrm{B} \rightarrow \mathrm{C}\}$. Thus satisfaction of F_1 in r_1 is not additive. Another example is the relation, r_2, shown in Table 3, where $r_2 \models\approx \mathrm{A} \rightarrow \mathrm{C}$ and $r_2 \models\approx \mathrm{B} \rightarrow \mathrm{C}$, but $r_2 \not\models\approx F_2 = \{\mathrm{A} \rightarrow \mathrm{C}, \mathrm{B} \rightarrow \mathrm{C}\}$. Thus satisfaction of F_2 in r_2 is not additive. Finally, another example is the relation, r_3, shown in Table 4, where $r_3 \models\approx \mathrm{B} \rightarrow \mathrm{A}$ and $r_3 \models\approx \mathrm{AC} \rightarrow \mathrm{B}$, but $r_3 \not\models\approx F_3 = \{\mathrm{B} \rightarrow \mathrm{A}, \mathrm{AC} \rightarrow \mathrm{B}\}$. Thus satisfaction of F_3 in r_3 is not additive.

A	B	C
0	*unk*	0
0	*unk*	1

Table 2: The counterexample relation r_1

A	B	C
0	*unk*	0
0	0	*unk*
unk	0	1

Table 3: The counterexample relation r_2

A	B	C
0	0	*unk*
unk	0	0
0	1	0

Table 4: The counterexample relation r_3

We assume that the reader is familiar with Armstrong's axiom system for FDs [Arm74, Ull88, MR92, AA93], consisting of the inference rules: reflexivity, augmentation and transitivity. (We observe that, in general as can be seen from Example 3.1, the transitivity inference rule is *unsound* for FDs holding in incomplete relations.) A fundamental result in relational database theory is that Armstrong's axiom system is sound and complete for FDs holding in complete relations. We denote the closure of a set of FDs F over R with respect to Armstrong's axiom system by F^+, and say that a set of FDs G over R is a *cover* of F if $G^+ = F^+$.

The following proposition shown in [LL97] follows from Definition 3.3 and the fact that Armstrong's axiom system is sound and complete for complete relations.

Proposition 3.2 A set of FDs F over R is consistent in a relation r over R if and only if F^+ is consistent in r. $\square$

We close this section by mentioning that the time complexity of deciding whether $r \approx\!\!\!\!\models F$ is polynomial in the sizes of r and F. A polynomial-time algorithm for deciding whether $r \approx\!\!\!\!\models F$, designated by CHASE$(r, F)$, can be derived directly from Theorem 6.4 in [AA93] on using Theorem 3 in [Hon82]. The algorithm is analogous to the standard chase procedure for FDs [Hon82] and is defined as follows.

Let Dum $= \{\perp_1, \perp_2, \ldots, \perp_q\}$ be a set of distinguished non-null values in *Dom* which do not appear in r, and let q denote the finite number of distinct occurrences of *unk* in r. For the purpose of defining the chase procedure we extend the partial order in *Dom* as follows: $\perp_i \sqsubseteq \perp_j$ if and only if i $\leq$ j, and for all non-null values v appearing in r, we have that $\forall \perp_i \in$ Dum, $\perp_i \sqsubseteq v$ but $v \not\sqsubseteq \perp_i$. The pseudo-code for the algorithm CHASE(r, F), which given the inputs r and F returns YES is F is consistent in r and NO if F is inconsistent in r is presented in Algorithm 3.1.

Algorithm 3.1 (CHASE(r, F))

begin
  Tmp := r;
  i := 1;
  for each A $\in$ R **do**
    for each $t \in$ Tmp such that t[A] = *unk* **do**
      t[A] := $\perp_i$;
      i := i + 1;
    end for
  end for
  while $\exists t_1, t_2 \in$ Tmp **and** $\exists$ X $\rightarrow$ Y $\in F$ such that t_1[X] = t_2[X] **and**
      $\exists$ A $\in$ Y such that t_1[A] $\neq$ t_2[A] **do**
    if t_1[A], t_2[A] $\notin$ Dum **then**
      return NO;
    else
      t_1[A], t_2[A] := the least upper bound of t_1[A] and t_2[A] with respect to $\sqsubseteq$;
    end if
  end while
  return YES;
end.

The following proposition states an important property of the chase procedure [AA93, Hon82], which will be useful in the next section. (Note that Dum $\subseteq$ *Dom*.)

Proposition 3.3 A set of FDs F over is consistent in a relation r over R if and only if CHASE(r, F) returns YES. $\square$

An immediate corollary of Proposition 3.3 is that if CHASE(r, F) returns YES then satisfaction of F in r is additive. We also observe that if CHASE(r, F) returns YES, then it is unique up to renaming of the constants in Dum.

4 Maximally consistent subsets of functional dependencies

Here we formalise the concept of maximally consistent subsets of a set of FDs F in an incomplete relation r which is central to this work. Intuitively, each maximally consistent subset of F is a possible interpretation of the set of FDs that are satisfied in r. The core of F in r, which is the intersection of all the maximally consistent subsets of F in r, contains all the FDs which undisputably are consistent in r. We investigate issues related to the computational complexity of finding maximally consistent subsets and conclude that these problems are, in general, intractable.

Definition 4.1 (Maximally consistent subset of FDs) We call a set of FDs G over R a *maximal subset* of a set of FDs F over R, or alternatively we call F a *minimal superset* of G, if $G^+ \subset F^+$ and there does not exist a set of FDs H over R such that $G^+ \subset H^+ \subset F^+$.

Given a family of sets of FDs, we say that a set of FDs G in this family is *maximal* if there does not exist a set of FDs H in the family such that $G^+ \subset H^+$.

G is said to be a *maximally consistent subset* of F in a relation r over R, if G is consistent in r and there does not exist a set of FDs H over R such that H is consistent in r and $G^+ \subset H^+ \subseteq F^+$. G is said to be the *core* of F in r if G is a cover of the intersection of the closures of all maximally consistent subsets of F in r; an FD $\mathrm{X} \rightarrow \mathrm{Y} \in G^+$ is said to be in the core of F in r.

If G is a maximally consistent subset of F in r then the FDs in $F^+ - G^+$ are said to have been *retracted* from F.

Without loss of generality, from now on when we refer to a maximally consistent sets of FDs, say G, we often do not distinguish between G and a cover of G, and assume that they are identical unless otherwise stated.

An immediate consequence of the Definition 4.1 is that, assuming that $r \in \mathrm{SAT}(F)$, then F has a unique maximally consistent subset in r if and only if F is consistent in r. Moreover, there must be at least one maximally consistent subset of F in r, since the emptyset of FDs is consistent in r. Finally, given any distinct maximally consistent subsets G_1 and G_2, their closures G_1^+ and G_2^+ must be incomparable.

The pseudo-code for the algorithm MAX_CON(r, F), which given the inputs r and F returns the set of all maximally consistent subsets of F in r is presented in Algorithm 4.1.

Algorithm 4.1 (MAX_CON(r, F))

```
begin
  max_con := ∅;
  if CHASE(r, F) returns YES then
    max_con := max_con ∪ {F};
  else % CHASE(r, F) returns NO
    for all maximal subsets G of F do
      max_con := max_con ∪ MAX_CON(r, G);
    end for
  end if
  return the maximal sets of FDs in max_con;
end.
```

Prior to proving the correctness of Algorithm 4.1 we illustrate it with an example.

Example 4.1 Let F = {A → B, B → C} be a set of FDs over R = {A, B, C}. The non-trivial FDs in F^+ are {A → B, AC → B, B → C, AB → C, A → C} and so the maximal subsets of F are:

1. F_1 = {AC → B, B → C}.
2. F_2 = {B → C, A → C}.
3. F_3 = {A → B, A → C}.

Considering the relation over R of Table 2, which we call r, it can be seen that F_1 is a maximally consistent subset of F in r, but F_2 and F_3 are not, since they are inconsistent in r. The only other maximally consistent subset of F in r is $F_4 = \{\mathrm{A} \rightarrow \mathrm{B}\}$ and thus the core of F in r is $F_1^+ \cap F_2^+ = \{\mathrm{AC} \rightarrow \mathrm{B}\}$.

To illustrate the need for line 10 of the algorithm, let us extend the schema R of r by two additional attributes D and E, extend the tuples of r with these two attributes by setting the attribute values in the extended tuples to be *unk* thus obtaining s, and finally let us add the FD D $\rightarrow$ E to F obtaining G. It can easily be verified that D $\rightarrow$ E is in the core of G in s but F_1 and F_4 will still be included in the contents of the variable max_con of the Algorithm 4.1.

Theorem 4.1 MAX_CON(r, F) returns the set of all maximally consistent subsets of F in r.

Proof. As a result of Definition 4.1, G is a maximally consistent subset of F in r if and only if CHASE(r, G) returns YES and there does not exist a minimal superset H of G, with $H^+ \subseteq F^+$, and such that H is consistent in r. Thus in order to test whether a G is a maximally consistent subset of F in r, only its minimal supersets need to be checked for consistency in r.

Now, suppose that G is an FD set in the family of FDs returned at line 10 of the algorithm. Then, G is consistent in r, since by line 3 of Algorithm 4.1 CHASE(r, G) returns YES. Now, during each recursive call to MAX_CON at line 7 of Algorithm 4.1, we consider only the maximal subsets of the current state, G of F. Moreover, without any loss of generality, we can assume that G is such that the FDs retracted from F are not in the core of F in r, since at line 6 all maximal subsets of F are considered. Therefore, for every minimal superset H of G, with $H^+ \subseteq F^+$, we have that H is inconsistent in r, as required. It follows by the above observation that G is a maximally consistent subset of F in r. The result follows, since at line 10 we return all and only the maximal sets of FDs in max_con. $\square$

We note that the core of F in r can easily be computed using Algorithm 4.1 by intersecting all the sets of FDs in the set returned by MAX_CON(r, F). This approach to computing the core is not always feasible, since the proposition given below implies that the complexity of computing MAX_CON(r, F) is in the worst case exponential in the size of F.

Proposition 4.2 There exists a relation schema R, a set of FDs F over R and a relation r over R such that the number of maximally consistent subsets of F in r is exponential in the size of F.

Proof. We construct a set of FDs F such that F is a set of non-trivial unary FDs over R, i.e. such that the left-hand sides of all FDs in F are singletons. Moreover, we let A $\rightarrow$ B $\in F^+ - F$ be an FD such that A $\rightarrow$ B $\in F^+$, can be derived from F solely by using the transitivity inference rule, where A and B are singleton attributes in R. For example, if $F = \{$A $\rightarrow A_1$, $A_1 \rightarrow A_2$, $A_2 \rightarrow$ B$\}$, then A $\rightarrow$ B $\in F^+$ can be derived by using two applications of the transitivity rule. We call the minimal number of times the transitivity rule is applied in such a derivation of A $\rightarrow$ B from F the *length* of the derivation, and assume that the minimal length of any such derivation is $m \geq 2$. We further assume that there are $k \geq 2$ distinct derivations, A $\rightarrow A_1^1$, $A_1^1 \rightarrow A_2^1$, ..., $A_m^1 \rightarrow$ B, A $\rightarrow A_1^2$, $A_1^2 \rightarrow A_2^2$, ..., $A_m^2 \rightarrow$ B, ..., A $\rightarrow A_1^k$, $A_1^k \rightarrow A_2^k$, ..., $A_m^k \rightarrow$ B, of A $\rightarrow$ B from F. Finally, we also assume that the set of attributes A, B, A_1^1, A_2^1, ..., A_m^1, A_1^2, A_2^2, ..., A_m^2, ..., A_1^k, A_2^k, ..., A_m^k in R are all pairwise distinct.

Now, let r be a relation over R having two tuples t_1 and t_2 such that t_1[A] $= t_2$[A] $= < 0 >$, t_1[R $-$ AB] $= t_2$[R $-$ AB] $= < unk, \ldots, unk >$, t_1[B] $=$ $< 0 >$ and t_2[B] $= < 1 >$. We claim that the cardinality of MAX_CON(r, F) is exactly m^k, where k can easily be seen to correspond to a small fraction of the size of F, when we set the cardinality of F to be mk.

In order to prove the claim, we note that in order to remove the inconsistency from r we must remove k distinct FDs from F, whose right-hand sides, are $A^i_{j_i}$, $1 \leq i \leq k$, $j_i \in \{1, 2, \ldots, m\}$. Moreover, by assumption, all the $A^i_{j_i}$ must be pairwise distinct. Let $G^i_{j_i}$ be a cover of the set of FDs resulting from removing from F^+ the FD $A^i_{j_i-1} \rightarrow A^i_{j_i}$ and adding to the remaining set of FDs, the set of FDs $\{A^i_{j_i-1}\mathrm{C} \rightarrow A^i_{j_i} \mid \mathrm{C} \in R - \{\mathrm{A}, A^i_1, A^i_2, \ldots, A^i_{j_i-1}, A^i_{j_i}\}\}$. By the construction of r and F, it can now be verified that $G(< j_1, j_2, \ldots, j_k >)$ is maximally consistent in r, where $G(< j_1, j_2, \ldots, j_k >) = \bigcup_{i \in \{1,2,\ldots,k\}} G^i_{j_i}$, since by the construction of F we cannot derive the FDs $A^i_{j_i-1} \rightarrow A^i_{j_i}$, $1 \leq i \leq k$, from $G(< j_1, j_2, \ldots, j_k >)$. The result now follows that there are exactly m^k distinct maximally consistent subsets $G(< j_1, j_2, \ldots, j_k >)$ of F in r, with $1 \leq j_i \leq m$, no two of which are covers of each other. $\Box$

As a result of Proposition 4.2 computing the core of F in r by intersecting the maximally consistent sets of FDs in MAX_CON(r, F) is, in general, of exponential time complexity. The next lemma shows that the problem of deciding whether a set of FDs is a maximally consistent subset of another set of FDs in a relation is in co-NP.

Lemma 4.3 Given sets of FDs G and F over R and a relation r over R the problem of deciding whether G is a maximally consistent subset of F in r is in co-NP.

Proof. We show that the problem is in co-NP by showing that its complement is in NP. We observe that by Definition 4.1 G is *not* a maximally consistent subset of F in r if and only if either G is inconsistent in r or G is consistent in r and there exists a set of FDs H over R, with $G^+ \subset H^+ \subseteq F^+$ such that H is consistent in r. We first check whether G is consistent in r or not by computing $\mathrm{CHASE}(r, G)$. Given that the answer is positive, we guess an FD $\mathrm{X} \to \mathrm{Y} \in \mathrm{FD(R)}$, and letting $H = G \cup \{\mathrm{X} \to \mathrm{Y}\}$, we can check in polynomial time whether $G^+ \subset H^+ \subseteq F^+$ [Ull88, MR92, AA93]. If this is the case then we proceed to check in polynomial time whether H is consistent in r or not by computing $\mathrm{CHASE}(r, H)$. If H is consistent in r, then G is not a maximally consistent subset of F in r. $\square$

In Section 5 (Proposition 5.4) we establish that the above decision problem is at least as hard as the *FD relation equivalence problem* [GL90], which is the problem of deciding whether a set of FDs is equal to the set of all FDs that can be inferred from a complete relation, whose complexity is still an unresolved problem [EG95].

The next theorem shows that the problem of deciding whether an FD is in the core of a set of FDs in a relation is co-NP-complete.

Theorem 4.4 Given sets of FDs F over R and a relation r over R the problem of whether an FD $\mathrm{X} \to \mathrm{A}$ is in the core of F in r is co-NP-complete.

Proof. We show that the problem is in co-NP by showing that its complement is in NP. We observe that by Definition 4.1 $\mathrm{X} \to \mathrm{A}$ is *not* in the core of F in r if and only if either $\mathrm{X} \to \mathrm{A} \notin F^+$ or for some set of FDs G over R, with $G^+ \subseteq F^+$, $\mathrm{CHASE}(r, G)$ returns YES but $\mathrm{CHASE}(r, G \cup \{\mathrm{X} \to \mathrm{A}\})$ returns NO. We first check in polynomial time whether $\mathrm{X} \to \mathrm{A} \in F^+$ or not. Given that the answer is positive, we guess a set of FDs G, and check whether $G^+ \subseteq F^+$ in polynomial time. The result follows since the chase can be computed in polynomial time in the size of its input.

To show that the complement of the problem is NP-hard a polynomial-time transformation from the hitting set problem, which is known to be NP-complete [Kar72], is given.

The hitting set problem: Given a family $S_1, S_2, \ldots, S_n$ of subsets of a set U, does there exist a subset W $\subseteq$ U, such that $\forall i \in$ {1,2,. . .,n}, $|\mathrm{W} \cap S_i| = 1$? (Such a subset W of U is called a hitting set, in other words W is a hitting set if for each i the cardinality of the intersection of W and S_i is one.)

We construct a relation schema, R such that R has one attribute A_j for each element in U, one attribute B_i for each set S_i and an additional attribute C. We then add the FD $B_1 B_2 \ldots B_n \rightarrow$ C to F, and an FD $A_j \rightarrow B_i$ to F, for each element in U represented by A_j that is in the set represented by B_i; for the rest of the proof we do not distinguish between the elements, sets and their respective representative attributes. Moreover, for each set of FDs, $A_{i1} \rightarrow B_i, A_{i2} \rightarrow B_i, \ldots, A_{ik} \rightarrow B_i$, in F, representing the elements included in S_i, we add k tuples $t_{i1}, t_{i2}, \ldots, t_{ik}$ to r such that $\forall j \in \{i1, i2, \ldots, ik\}$, t_j is given by

1. $t_j[A_j] =< 0 >$, $t_j[B_i] =< j >$ and
2. $t_j[\mathrm{R} - A_j B_i] = < unk, \ldots, unk >$.

Finally we add to r two distinguished tuples u_1 and u_2, where the cardinality of U is m, such that

1. $u_1[A_1, A_2, \ldots, A_m] = u_2[A_1, A_2, \ldots, A_m] =< 0, \ldots, 0 >$,
2. $u_1[B_1, B_2, \ldots, B_n] = u_2[B_1, B_2, \ldots, B_n] =< unk, \ldots, unk >$, and
3. $u_1[\mathrm{C}] = < 0 >$ and $u_2[\mathrm{C}] = < 1 >$.

It remains to show that there exists a hitting set of U if and only if $B_1, B_2, \ldots, B_n \rightarrow$ C is not in the core of F in r.

If. Suppose that $B_1, B_2, \ldots, B_n \rightarrow$ C is not in the core of F in r. Then, since $B_1, B_2, \ldots, B_n \rightarrow \mathrm{C} \in F$ there exists a set of FDs, say G, with $G^+ \subseteq F^+$ such that $r \models\approx G$, but $r \not\models\approx G \cup \{B_1, B_2, \ldots, B_n \rightarrow \mathrm{C}\}$. We claim that the set W $= \{A_j \mid A_j \rightarrow B_i \in G\}$ is a hitting set of U, i.e. W is the union of the left-hand sides of FDs in G.

Now, let $s_1, s_2 \in$ POSS(r), be the complete relations generated by CHASE(r, G) and CHASE$(r, G \cup \{B_1, B_2, \ldots, B_n \rightarrow \mathrm{C}\})$, respectively. Furthermore, let u_1' and u_2' be the states of u_1 and u_2 in s_1. Then by the construction of r it must be the case that $u_1'[B_1, B_2, \ldots, B_n] = u_2'[B_1, B_2, \ldots, B_n]$ but $u_1'[\mathrm{C}] \neq u_2'[\mathrm{C}]$ and thus $s_1 = s_2$ up to remaining of the constants in Dum. It follows that W must be a hitting set for U, otherwise G would not be consistent in r.

Only if. Suppose that W is a hitting set of U. Let $G = \bigcup_i \{A_j \to B_i \mid A_j = \mathrm{W} \cap S_i\}$, and let $s_1, s_2 \in \mathrm{POSS}(r)$ and u'_1 and u'_2 be as in the if part of the proof. Then by the construction of r it must be the case, as above, that $u'_1[B_1, B_2, \ldots, B_n] = u'_2[B_1, B_2, \ldots, B_n]$ but $u'_1[\mathrm{C}] \neq u'_2[\mathrm{C}]$ and thus $s_1 = s_2$ up to remaining of the constants in Dum. Moreover, by the construction of r, CHASE(r, G) returns YES since W is a hitting set of U. The result is now established, since it is evident that CHASE$(r, G \cup \{B_1, B_2, \ldots, B_n \to \mathrm{C}\})$ returns NO. $\square$

5 The dependency inference problem in incomplete relations

We now investigate the problem of inferring the set of FDs that hold in an incomplete relation r. Due to the fact that satisfaction of a set of FDs in an incomplete relation is not always additive we carry out the dependency inference procedure in two steps. The first step is the standard dependency inference algorithm [MR94] which returns $dep(r)$, the set of all FDs that are satisfied in r. Due to the fact that $dep(r)$ may be inconsistent in r, the second step computes the family of maximally consistent subsets of $dep(r)$ in r. We show that if G is maximally consistent in $dep(r)$ then there is a relation $s \in \mathrm{POSS}(r)$ such that G is a cover of $dep(s)$ but that the converse of this statement does not always hold. We also show that the problem of finding whether a set G of FDs is a maximally consistent subset of $dep(r)$ in r is at least as hard as the FD relation equivalence problem.

Definition 5.1 (The set of FDs that can be inferred from a relation) The set of FDs that can be inferred from a relation r over R, denoted by $dep(r)$, is a cover of the set of FDs given by

$$\{\mathrm{X} \to \mathrm{Y} \mid \mathrm{X} \to \mathrm{Y} \in \mathrm{FD(R)} \text{ and } r \models\!\approx \mathrm{X} \to \mathrm{Y}\}.$$

We briefly recall from [MR94] that in order to find a cover of the FDs in $dep(r)$ that functionally determine an attribute A, denoted by $lhs(r, \mathrm{A})$, we use the following procedure. First we compute for any two tuples $t, u \in r$ their *disagreement* set, modified to cater for incomplete relations, as follows:

$$disagree(t, u) = \{\mathrm{B} \in \mathrm{R} \mid t[\mathrm{B}] \neq unk \text{ and } u[\mathrm{B}] \neq unk \text{ and } t[\mathrm{B}] \neq u[\mathrm{B}]\}.$$

Secondly, we compute for any two tuples $t, u \in r$ their *maybe disagreement* set, which is novel for incomplete relations, as follows:

$$may\_disagree(t, u) = \{\mathrm{B} \in \mathrm{R} \mid t[\mathrm{B}] = unk \text{ or } u[\mathrm{B}] = unk \text{ or } t[\mathrm{B}] \neq u[\mathrm{B}]\}.$$

Then, we compute the *necessary* set for an attribute $\mathrm{A} \in \mathrm{R}$, which is given by

$$necessary(r, \mathrm{A}) = \{may\_disagree(t, u) - \{\mathrm{A}\} \mid t, u \in r \text{ and } \mathrm{A} \in disagree(t, u)\}.$$

Finally, $lhs(r, \mathrm{A})$ is the set of *minimal transversals* of the collection of minimal sets in $necessary(r, \mathrm{A})$. (Informally, a *transversal* of a collection $\mathcal{H}$ of subsets of R is a subset of R having a non-empty intersection with all the sets in $\mathcal{H}$; see [EG95] for a detailed exposition on the computational problems related to identifying minimal transversals and their applications in database theory.) Utilising [MR94, Theorem 5] it can easily be shown that $lhs(r, \mathrm{A})$, can still be computed in this manner for FDs being satisfied in incomplete relations, given our definition of $disagree(t, u)$ which is due to Proposition 3.1.

From the semantics of satisfaction of a set of FDs in an incomplete relation, it is obvious that $dep(r)$ is *not* the correct measure of the FDs that hold in r, since $dep(r)$ may not be consistent in r. The next theorem justifies taking MAX_CON(r) as representing all the sets of FDs that can be inferred from relations $s \in$ POSS(r), where MAX_CON(r) denotes MAX_CON$(r, dep(r))$.

Theorem 5.1 If $G \in$ MAX_CON(r) then G is a cover of $dep(s)$ for some $s \in$ POSS(r).

Proof. By Theorem 4.1 CHASE(r, G) returns YES and thus a complete relation $s \in$ POSS(r) is generated by the chase procedure such that $s \models_{\approx} G$. It follows that $G^+ \subseteq dep(s)^+$. To show that $dep(s)^+ \subseteq G^+$, suppose to the contrary that there exists an FD $\mathrm{X} \rightarrow \mathrm{Y} \in dep(s)^+ - G^+$. It follows that $r \models_{\approx} \mathrm{X} \rightarrow \mathrm{Y}$ and thus $\mathrm{X} \rightarrow \mathrm{Y} \in dep(r)^+$. Therefore we conclude that G is not maximally consistent in r, since $G \cup \{\mathrm{X} \rightarrow \mathrm{Y}\}$ is consistent in s, leading to a contradiction. $\square$

We observe that the converse of Theorem 5.1 is false. For example, let r be a relation over AB which contains the two tuples, $t_1 =< 0, unk >$ and $t_2 =< 0, 1 >$. It is easily seen that $\{\mathrm{A} \rightarrow \mathrm{B}, \mathrm{B} \rightarrow \mathrm{A}\}$ is the only set of FDs in MAX_CON(r) but there is a relation $s = \{t_1, t_2\} \in$ POSS(r) such that $t_1[\mathrm{B}] = 0$ and $t_2[\mathrm{B}] = 1$ and thus $s \not\models_{\approx} \mathrm{A} \rightarrow \mathrm{B}$.

Proposition 5.2 Let r be a relation over R. Then, $\forall s \in$ POSS(r), $\exists G \in$ MAX_CON(r) such that $dep(s)^+ \subseteq G^+$.

Proof. By Definition 3.3 we have that $dep(s)^+$ is consistent in r. The result follows, by Definition 4.1, since $dep(s)^+$ is a consistent subset of $dep(r)^+$ in r. $\Box$

The *FD relation equivalence problem* is the problem of deciding, given a complete relation s over R and a set of FDs F over R, whether $F^+ = dep(s)^+$ [GL90]. It was shown in [EG95] that the FD relation equivalence problem is as difficult as the problem of finding whether a hypergraph $\mathcal{H}$ is equal to the family of all minimal transversals of another hypergraph $\mathcal{G}$, where in our context a *hypergraph* is a family of subsets of R. Thus, although the FD relation equivalence problem is in co-NP it is not known whether it is intractable or not.

Proposition 5.3 The problem of deciding whether a set of FDs G over R is a maximally consistent subset of $dep(r)$ in a relation r over R is in co-NP and is at least as hard as the FD relation equivalence problem.

Proof. A similar argument as in Lemma 4.3 is can be used to show that problem is in co-NP. To prove the second part of the result it is sufficient to note that when r is a complete relation, $G^+ = dep(r)^+$ if and only G is maximally consistent in r. $\Box$

We now show that the problem of deciding whether a set of FDs is a maximally consistent subset of another set of FDs in a relation is at least as hard as deciding whether a set of FDs is a maximally consistent subset of the set of FDs that can be inferred from a relation. Thus by Proposition 5.3 the latter problem is also at least as hard as the FD relation equivalence problem.

Proposition 5.4 The problem of deciding whether a set of FDs G over R is a maximally consistent subset of a set of FDs F over R in a relation r over R is at least as hard as the problem of deciding whether G is a maximally consistent subset of $dep(r)$ in r.

Proof. Given G and r let $F = \{\emptyset \rightarrow \text{A} \mid \text{A} \in \text{R}\}$, i.e. F^+ is the set of all possible FDs over R. It can now be verified that G is a maximally consistent subset of F in r if and only if G is a maximally consistent subset of $dep(r)$ in r. $\Box$

6 The semantics of logical implication

In this final section we show an interesting connection between Lien and Atzeni's axiomatisation of FDs in incomplete relations [Lie82, AM86] (see also [LL95]) and maximally consistent sets of FDs. In particular, we show that we can refine the set of maximally consistent subsets of a set of FDs F by choosing an appropriate relation $r \in \text{SAT}(F)$, so that an FD can be derived from F using Lien and Atzeni's axiom system if and only if for some $G \in \text{MAX\_CON}(r, F)$, the FD can be derived from G using Armstrong's axiom system.

Definition 6.1 (Logical implication) We say that F *logically implies* X $\rightarrow$ Y, written $F \models$ X $\rightarrow$ Y, whenever $\text{SAT}(F) \subseteq \text{SAT}(\{\text{X} \rightarrow \text{Y}\})$.

Lien [Lie82], and Atzeni and Morfuni [AM86] have shown that the inference rules: reflexivity, augmentation, decomposition and union, are sound and complete for FDs holding in incomplete relations with respect to logical implication defined above; we call this axiom system, Lien and Atzeni's axiom system. That is, by dropping the pseudo-transitivity rule from Armstrong's axiom system and adding the decomposition and union rules, we obtain Lien and Atzeni's axiom system. We denote the closure of a set of FDs F over R with respect to Lien and Atzeni's axiom system by F^*, noting that $F^* \subseteq F^+$. For the sake of completeness we detail the inference rules mentioned above, where $F \vdash$ X $\rightarrow$ Y below reads: X $\rightarrow$ Y can be derived from F.

1. *Reflexivity*: if Y $\subseteq$ X $\subseteq$ R, then $F \vdash$ X $\rightarrow$ Y.
2. *Augmentation*: if $F \vdash$ X $\rightarrow$ Y and W $\subseteq$ R, then $F \vdash$ XW $\rightarrow$ YW.
3. *Pseudo-transitivity*: if $F \vdash$ X $\rightarrow$ Y and WY $\rightarrow$ Z, then $F \vdash$ XW $\rightarrow$ Z.
4. *Union*: if $F \vdash$ X $\rightarrow$ Y and $F \vdash$ X $\rightarrow$ Z, then $F \vdash$ X $\rightarrow$ YZ.
5. *Decomposition*: if $F \vdash$ X $\rightarrow$ YZ, then $F \vdash$ X $\rightarrow$ Y and $F \vdash$ X $\rightarrow$ Z.

We now present a semantics which is equivalent to logical implication for FDs in incomplete relations, in terms of maximally consistent sets.

Definition 6.2 (Finest set of maximally consistent subsets of FDs) Let F be a set of FDs over R and r_1 and r_2 be relations over R such that

$r_1, r_2 \in \mathrm{SAT}(F)$. Moreover, let M_1 and M_2 be the families of maximally consistent subsets of F in r_1 and r_2, respectively. Then, M_1 is *finer* than M_2 if

$$\forall G_1 \in M_1, \exists G_2 \in M_2 \text{ such that } G_1^+ \subseteq G_2^+.$$

A family M of sets of FDs over R is said to be a *finest* set of maximally consistent subsets of F, if for some $r \in \mathrm{SAT}(F)$ M is the set of maximally consistent subsets of F in r, and there does not exist $r' \in \mathrm{SAT}(F)$ such that M' is finer than M but M is not finer than M', where M' is the set of maximally consistent subsets of F in r'.

Prior to showing that there is always a unique finest set of maximally consistent subsets of a set of FDs we given an example.

Example 6.1 For the set of FDs $F = \{\mathrm{A} \rightarrow \mathrm{B},\ \mathrm{B} \rightarrow \mathrm{C}\}$ of Example 4.1, $M = \{\{\mathrm{AC} \rightarrow \mathrm{B},\ \mathrm{B} \rightarrow \mathrm{C}\}, \{\mathrm{A} \rightarrow \mathrm{B}\}\}$ is the unique finest set of maximally consistent subsets of F. We observe that M coincides with the set of maximally consistent subsets of F in the well known relation r_1 of Table 2, as was shown in Example 4.1, since $r_1 \not\models\!\approx \mathrm{A} \rightarrow \mathrm{C}$.

Lemma 6.1 Given a set of FDs F over R there exists a unique finest set of maximally consistent subsets of F, up to the covers chosen for the FDs sets in the family.

Proof. We construct a specific relation $r \in \mathrm{SAT}(F)$ and show that the family M of maximally consistent subsets of F in r, is the unique finest set of maximally consistent subsets of F.

Let us say that the satisfaction of a set of FDs F', with $(F')^+ \subseteq F^+$, is *not additive* if for some relation $r' \in \mathrm{SAT}(F)$, F' is inconsistent in r'. Now, we construct a relation r, by taking the union of all relations r' as above, one for each non-additive set of FDs F', with $(F')^+ \subseteq F^+$. We ensure that r is in $\mathrm{SAT}(F)$ by assuming, without loss of generality, that the non-null values in each such r' are pairwise disjoint. By the construction of r it follows that for all $G \in M$, G is consistent in all relations in $\mathrm{SAT}(F)$. Therefore, by Definition 4.1 M is finer than any other family M' of maximally consistent subsets of F in some relation in $\mathrm{SAT}(F)$. Moreover, M is unique, since by Definition 6.2 if there is any set M' that is finer than M then by the construction of r, M are M' are identical. $\square$

The next lemma shows the relationship between the closure of a set of FDs, say F^*, under Lien and Atzeni's axiom system and the individual closures of sets of FDs in the finest set of maximally consistent subsets of F.

Lemma 6.2 If M is the finest set of maximally consistent subsets of F, then $\bigcup_{G \in M} G^+ = F^*$.

Proof. Firstly, we show that $\bigcup_{G \in M} G^+ \subseteq F^*$. Assume to the contrary that for some $G \in M$ there exists an FD X $\rightarrow$ Y $\in G^+ - F^*$. By Definition 6.1 there exists a relation $r \in$ SAT(F) that violates X $\rightarrow$ Y in all relations in POSS(r). This implies that G is inconsistent in r which contradicts a property of the finest set of maximally consistent subsets of F established in Lemma 6.1.

Secondly, we show that $F^* \subseteq \bigcup_{G \in M} G^+$. It is sufficient to establish that if X $\rightarrow$ Y $\in F^*$, then X $\rightarrow$ Y $\in G^+$ for some $G \in M$. Now, since X $\rightarrow$ Y $\in F^*$, it follows that $\forall$ A $\in$ Y$-$X there is some W $\rightarrow$ Z $\in F$ such that W $\subseteq$ X and A $\in$ Z [AM86]. Let $H = \{$W $\rightarrow$ A $\mid$ A $\in$ Y$-$X and W $\rightarrow$ Z $\in F$ satisfies W $\subseteq$ X and A $\in$ Z$\}$ be a set of FDs consisting of $|$Y$-$X$|$ FDs implying that X $\rightarrow$ Y $\in H^*$. It can now be verified that for all relations $r \in$ SAT(F), we have that H is consistent in r (cf. [LL97]). The result follows, since all sets of FDs in M are maximally consistent subsets of F in the relation constructed in Lemma 6.1, and thus for some $G \in M$, $H^+ \subseteq G^+$. $\square$

The next theorem, which is an immediate consequence of Lemma 6.2 and the soundness and completeness of Lien and Atzeni's axiom system, states that the finest set of maximally consistent subsets of a set of FDs captures the notion of logical implication for FDs in incomplete relations.

Theorem 6.3 $F \models$ X $\rightarrow$ Y if and only if X $\rightarrow$ Y $\in G^+$ for some $G \in M$, where M is the finest set of maximally consistent subsets of F. $\square$

7 Concluding Remarks

Given a set of FDs F over R and an incomplete relation r over R we have investigated an approach that considers the maximally consistent subsets of F in r to be the appropriate semantics. We have shown that computational issues relating to finding maximally consistent subsets are, in general, intractable. In addition, we have shown how maximally consistent subsets should be taken into account when confronted with the dependency inference problem for FDs in incomplete relations. Finally, we have shown a connection between Lien and Atzeni's axiom system for FDs in incomplete relations and maximally consistent subsets. It is an open research problem to extend the formalism developed in this paper to a wider class of integrity constraints.

In our solution to the problem of recovering from inconsistency we change our belief regarding the set of data dependencies that should hold in a given relation and do not doubt the credibility of the relation itself. A complementary approach to the one taken in this paper, that is worth investigating when the data may contain errors, is given a set of FDs F over R and a relation r over R find the set of all relations that are *closest* to r in some well-defined sense such that F is consistent in each one of these relations (see [CL98]).

Acknowledgements. The authors would like to thank the referee for his constructive criticism and helpful remarks.

References

[AA93] P. Atzeni and V. De Antonellis. *Relational Database Theory*. Benjamin/Cummings, Redwood City, Ca., 1993.

[AM86] P. Atzeni and N.M. Morfuni. Functional dependencies and constraints on null values in database relations. *Information and Control*, 70:1–31, 1986.

[Arm74] W.W. Armstrong. Dependency structures of data base relationships. In *Proceedings of the IFIP Congress*, pages 580–583, Stockholm, 1974.

[BS95] R. Bagai and R. Sunderraman. A paraconsistent relational data model. *International Journal of Computer Mathematics*, 55:39–55, 1995.

[CL98] E. Collopy and M. Levene. Evolving example relations to satisfy functional dependencies. In *Proceedings of the Third Biennial World Conference on Integrated Design and Process Technology - Issues and Applications of Database Technology*, pages 440–447, Berlin, 1998.

[EG95] T. Eiter and G. Gottlob. Identifying the minimal transversals of a hypergraph and related problems. *SIAM Journal on Computing*, 24:1278–1304, 1995.

[EGH95] M. Elvang-Gøransson and A. Hunter. Argumentative logics : Reasoning with classically inconsistent information. *Data & Knowledge Engineering*, 16:125–145, 1995.

[Gär88] P. Gärdenfors. *Knowledge in flux: Modelling the dynamics of epistemic states.* MIT Press, Cambridge, MA, 1988.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness.* W.H. Freeman, New York, 1979.

[GL90] G. Gottlob and L. Libkin. Investigations on Armstrong relations, dependency inference and excluded functional dependencies. *Acta Cybernetica*, 4:385–402, 1990.

[Hon82] P. Honeyman. Testing satisfaction of functional dependencies. *Journal of the ACM*, 29:668–677, 1982.

[Jaś69] S. Jaśkowski. Propositional calculus for contradictory deductive systems. *Studia Logica*, 24:143–157, 1969.

[Kar72] R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum, New York, 1972.

[KdC78] J. Kotas and N.C.A. da Costa. On the problem of Jaškowski and the logics of Lukasiewicz. In A.I. Arrunda, N.C.A. da Costa, and R. Chuaqui, editors, *Mathematical Logic: Proceedings of the first Brazilian Conference*, pages 127–139. Marcel Dekker, New York, 1978.

[Lev95] M. Levene. A lattice view of functional dependencies in incomplete relations. *Acta Cybernetica*, 12:181–207, 1995.

[Lie82] Y.E. Lien. On the equivalence of database models. *Journal of the ACM*, 29:333–362, 1982.

[LL95] M. Levene and G. Loizou. A correspondence between variable relations and three-valued propositional logic. *International Journal of Computer Mathematics*, 55:29–38, 1995.

[LL97] M. Levene and G. Loizou. The additivity problem for functional dependencies in incomplete relations. *Acta Informatica*, 34:135–149, 1997.

[MR92] H. Mannila and K.-J. Räihä. *The Design of Relational Databases*. Addison-Wesley, Reading, MA, 1992.

[MR94] H. Mannila and K.-J. Räihä. Algorithms for inferring functional dependencies from relations. *Data & Knowledge Engineering*, 12:83–99, 1994.

[Ull88] J.D. Ullman. *Principles of Database and Knowledge-Base Systems*, volume 1. Computer Science Press, Rockville, MD, 1988.