The Additivity Problem for Functional Dependencies in Incomplete Relations *

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

George Loizou
Department of Computer Science
Birkbeck College London
University of London
Malet Street, London, WC1E 7HX, UK
Email: george@dcs.bbk.ac.uk

Abstract

Incomplete relations are relations which contain null values, whose meaning is "value is at present unknown". A functional dependency (FD) is weakly satisfied in an incomplete relation if there exists a possible world of this relation in which the FD is satisfied in the standard way. Additivity is the property of equivalence of weak satisfaction of a set of FDs, say F, in an incomplete relation with the individual weak satisfaction of each member of F in the said relation. It is well known that satisfaction of FDs is not additive.

The problem that arises is: under what conditions is weak satisfaction of FDs additive. We solve this problem by introducing a syntactic subclass of FDs, called *monodependent* FDs, which informally means that for each attribute, say A, there is a unique FD that functionally determines A, and in addition only trivial cycles involving A arise between any two FDs one of which functionally determines A. We show that weak satisfaction of FDs is additive if and only if the set F of FDs is monodependent and that monodependence can be checked in time polynomial in the size of F.

1 Introduction

In order to handle incomplete information, Codd [7] suggested the addition to the database domains of an unmarked null value, whose meaning is "value at present unknown", which we denote by *unk*. We call such relations, whose tuples may contain the null value *unk*, *incomplete relations*.

Functional Dependencies (or simply FDs) are by far the most common integrity constraints in the real world [18, 3] and the notion of a key (derived from a given set of FDs) [7] is fundamental to the relational model. A sound and complete axiom system for FDs was first given in [1] (see also [18, 3]) and is known as *Armstrong's axiom system*. When considering the satisfaction of FDs in incomplete relations the transitivity rule is no longer sound. Thus, Lien [15] and Atzeni and Morfuni [2] extended FDs so as to deal with missing information, suggesting the interpretation of "inapplicable" or "nonexistent", and "no information", respectively, of the unmarked null value. A sound and complete axiom system for FDs holding in incomplete relations was obtained by dropping the transitivity rule and adding the union and decomposition rules [15, 2]; we shall refer to this axiom system as *Lien and Atzeni's axiom system*.

Let r be an incomplete relation over a relation schema R and let F be a set of FDs over R. Furthermore, let POSS(r) denote the set of *possible worlds* relative to r, i. e. the set of all relations that emanate from all possible substitutions of occurrences of *unk* in r by non-null values in the database domains. We say that an incomplete relation r *weakly satisfies* F (or simply satisfies F when no ambiguity arises), written r $\approx$ F, if $\exists$ s $\in$ POSS(r) such that s satisfies F on using the standard definition of an FD [18, 3].

*To appear in Acta Informatica

Weak satisfaction is not *additive* [3], i. e. it may be the case that $\forall$ X $\rightarrow$ Y $\in$ F, r $\approx\!\!\!|$ X $\rightarrow$ Y, but r $\not\approx$ F. For example, the well known incomplete relation [2], say r_1, shown in Table 1, is such that $r_1 \models_{\approx}$ A $\rightarrow$ B and $r_1 \models_{\approx}$ B $\rightarrow$ C but $r_1 \not\models_{\approx}$ {A $\rightarrow$ B, B $\rightarrow$ C}. This is due to the fact that $\forall s_1 \in$ POSS(r_1), $s_1 \not\models_{\approx}$ A $\rightarrow$ C, which is inferred from A $\rightarrow$ B and B $\rightarrow$ C by the transitivity rule of Armstrong's axiom system. Another example is the incomplete relation, say r_2, shown in Table 2, where $r_2 \models_{\approx}$ A $\rightarrow$ C and $r_2 \models_{\approx}$ B $\rightarrow$ C, but $r_2 \not\models_{\approx}$ {A $\rightarrow$ C, B $\rightarrow$ C}. This is due to the fact that $\forall s_2 \in$ POSS(r_2), the C-value of the second tuple is either 1 or 0 and therefore either $s_2 \not\models_{\approx}$ A $\rightarrow$ C or $s_2 \not\models_{\approx}$ B $\rightarrow$ C. In this case two incomparable sets of attributes, A and B, functionally determine a common attribute C. It is an interesting fact that it is also the case that $r_2 \not\models_{\approx}$ {A $\rightarrow$ B, B $\rightarrow$ C} and thus the second relation is also a counterexample for the first set of FDs. Finally, another relevant example is the incomplete relation, say r_3, shown in Table 3, where $r_3 \models_{\approx}$ B $\rightarrow$ A and $r_3 \models_{\approx}$ AC $\rightarrow$ B, but $r_3 \not\models_{\approx}$ {B $\rightarrow$ A, AC $\rightarrow$ B}. This is due to the fact that $\forall s_3 \in$ POSS(r_3), the A-value of the second tuple must be 0 due to B $\rightarrow$ A and therefore $s_3 \not\models_{\approx}$ AC $\rightarrow$ B.

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

Table 1: The counterexample relation r_1

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

Table 2: The counterexample relation r_2

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

Table 3: The counterexample relation r_3

The problem that is solved in this paper is to present a syntactic characterisation of when weak satisfaction is additive. This is important when dealing with incomplete information, since Lien and Atzeni's axiom system, which is sound and complete for FDs with respect to weak satisfaction of a single FD, does not, in general, cater for weak satisfaction of a set F of FDs.

We now briefly describe the solution to the additivity problem, where F is a set of FDs over a relation schema R. Informally, F is *monodependent* if for each attribute A there is a unique FD that functionally determines A, and in addition only trivial cycles involving A arise between any two FDs one of which functionally determines A. An example of a set of FDs where an attribute C is not uniquely determined is {A $\rightarrow$ C, B $\rightarrow$ C}, and an example of a set of FDs where a non-trivial cycle arises between two FDs is {B $\rightarrow$ A, AC $\rightarrow$ B}. It follows that these two sets of FDs are not monodependent.

We show that we can check whether F is monodependent in time polynomial in the size of F. We then solve the additivity problem by showing that weak satisfaction is additive if and only if F is monodependent. Thus, the most general class of sets of FDs for which additivity holds is the class of monodependent sets of FDs.

The layout of the rest of the paper is as follows. In Section 2 we formalise incomplete relations and define a partial order in the set of tuples of such relations. In Section 3 we define the notion of FDs and their satisfaction in the context of incomplete relations. In Section 4 we present our solution to the additivity problem in the form of monodependent sets of FDs. Finally, in Section 5 we give our concluding remarks.

2 Relations that model incomplete information

In this section we extend relation schemas and relations so as to model incomplete information.

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, T, i. e. S $\cup$ T, simply by ST. We will refer to the cardinality of some standard encoding [10] of S as the *size* of S.

Definition 2.1 (Relation schema and relation) A *relation schema* R is a finite set of attributes which we denote by schema(R); we denote the cardinality of R by type(R).

We assume a countably infinite domain of constants, *Dom*, containing two distinguished constants *unk* and *inc*, denoting the null values "unknown" and "inconsistent", respectively.

A type(R)-tuple (or simply a tuple whenever type(R) is understood from the context) is a total mapping from schema(R) into *Dom* such that $\forall A_i \in$ schema(R), t$(A_i) \in$ *Dom*. A relation over R is a finite set of type(R)-tuples.

From now on we *let R be a relation schema and r be a relation over R.* In addition, we let REL(R) denote the countably infinite set of relations over R.

We note that we have actually included two types of null value in our formalism: *unk* and *inc*. The inclusion of *unk* was motivated in the introduction and the inclusion of *inc* is motivated by the fact that it allows us to easily detect unwanted inconsistency; the latter also facilitates the construction of proofs later on.

Example 1 In Table 4 we show a relation, say r, over a relation schema, say R, where type(R) = 4 and schema(R) = {STUD, DEPT, HEAD, COURSE}. The semantics of R are: a STUDent belongs to one DEParTment, and takes one or more COURSEs. In addition, a department has one HEAD and each course is given by one department. We note that if we insert the tuple, <Hanna, *unk*, History, *inc*>, into r then the History department would have an inconsistent head.

STUD	COURSE	DEPT	HEAD
Iris	Databases	Computing	Dan
Iris	Set_Theory	Computing	*unk*
Reuven	Set_Theory	*unk*	*unk*
Naomi	Programming	Maths	Annette
Naomi	*unk*	Maths	*unk*
Eli	Logic	*unk*	Brian

Table 4: The students departments relation

Definition 2.2 (Projection) The *projection* of a type(R)-tuple t onto a set of attributes Y $\subseteq$ schema(R), denoted by t[Y] (also called the Y-value of t), is the restriction of t to Y. The projection of a relation r over R onto Y, denoted π_Y(r), is defined by π_Y(r) = {t[Y] | t $\in$ r}.

Definition 2.3 (Complete and consistent relations) A type(R)-tuple is said to be *complete* if $\forall A_i \in$ schema(R), t$[A_i] \neq$ *unk* and t$[A_i] \neq$ *inc*, i. e. t$[A_i]$ is a non-null value, otherwise t is said to be *incomplete*; t is said to be *inconsistent* if $\exists A_i \in$ schema(R), such that t$[A_i] =$ *inc*, otherwise t is said to be *consistent*.

A relation r over R is said to be *complete* if $\forall t \in r$, t is complete, otherwise r is said to be *incomplete* (when no confusion arises we use relation to mean incomplete relation). A relation r over R is said to be *inconsistent* if $\exists t \in r$ such that t is inconsistent, otherwise r is said to be *consistent*.

We let COMPLETE(R) denote the countably infinite set of all complete relations over R.

Definition 2.4 (Less informative constants and tuples) Let r be a relation over R. We define a partial order in *Dom*, denoted by $\sqsubseteq$, as follows:

$$u \sqsubseteq v \text{ if and only if } u = v \text{ or } u = unk \text{ or } v = inc, \text{ where } u, v \in Dom.$$

We extend $\sqsubseteq$ to be a partial order in the set of type(R)-tuples as follows: if t_1 and t_2 are type(R)-tuples, t_1 is *less informative than* t_2 (or equivalently t_2 is *more informative than* t_1), written $t_1 \sqsubseteq t_2$, if $\forall A_i \in$ schema(R), $t_1[A_i] \sqsubseteq t_2[A_i]$.

Two type(R)-tuples t_1 and t_2 are *information-wise equivalent*, i. e. $t_1 \sqsubseteq t_2$ and $t_2 \sqsubseteq t_1$, if and only if $t_1 = t_2$. We observe that the set of all type(R)-tuples is a *complete lattice*, with no infinite chains [8], whose bottom element is <*unk*, ..., *unk*> and whose top element is <*inc*, ..., *inc*>.

We next define the *join* operator [8] of this complete lattice of tuples.

Definition 2.5 (The join operator) Let r be a relation over R. We define the *join* operator, denoted by $\sqcup$, as a mapping from an ordered pair (v_1, v_2) in $Dom \times Dom$ to a single value in Dom as follows: $v_1 \sqcup v_2$ is the least upper bound of v_1 and v_2 with respect to $\sqsubseteq$. We extend $\sqcup$ to be a mapping from an ordered pair, (t_1, t_2), of type(R)-tuples to a single type(R)-tuple as follows: $t_1 \sqcup t_2 =$ t, where t is a type(R)-tuple and $\forall A_i \in$ schema(R), t$[A_i] = t_1[A_i] \sqcup t_2[A_i]$.

It can easily be verified that $t_1 \sqcup t_2$ returns the least upper bound of t_1 and t_2, namely the join operator realises the lattice theoretic join.

Definition 2.6 (The set of possible worlds of a relation) The set of all *possible worlds* relative to a relation r over R, denoted by POSS(r), is defined by

$$POSS(r) = \{s \mid s \text{ is a relation over R and there exists a total and onto mapping } f : r \to s \text{ such that } \forall t \in r, t \sqsubseteq f(t) \text{ and } f(t) \text{ is complete}\}.$$

Proposition 2.1 A relation r over R is inconsistent if and only if POSS(r) = $\emptyset$. $\square$

Hereafter we assume that *unless otherwise stated relations are consistent.*

3 Functional dependencies in relations which may be incomplete

We now define Functional Dependencies (FDs) and their satisfaction in the context of incomplete relations.

Definition 3.1 (Functional dependency) A *functional dependency* over R (or simply an FD) is a statement of the form X $\to$ Y, where X, Y $\subseteq$ schema(R). (For the semantics of FDs see Definition 3.4.)

We call an FD of the form X $\to$ Y, where Y $\subseteq$ X, a *trivial* FD. Two non-trivial FDs of the forms X $\to$ A and Y $\to$ A are said to be *incomparable* if X and Y are incomparable. Two non-trivial FDs of the forms XB $\to$ A and YA $\to$ B are said to be *cyclic*.

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 sets of FDs over R. From now on we will assume that *F is a set of FDs over R.*

Definition 3.2 (Armstrong's axiom system) The closure of F with respect to Armstrong's axiom system [1] (see also [18] and [3]), denoted by F^+, is the smallest set of FDs that contains F and satisfies the following five conditions:

FD1 Reflexivity: if Y $\subseteq$ X $\subseteq$ schema(R), then X $\to$ Y $\in F^+$.

FD2 Augmentation: if X $\to$ Y $\in F^+$ and W $\subseteq$ schema(R), then XW $\to$ YW $\in F^+$.

FD3 Pseudo-transitivity: if X $\rightarrow$ Y $\in F^+$ and WY $\rightarrow$ Z $\in F^+$, then XW $\rightarrow$ Z $\in F^+$.

FD4 Union: if X $\rightarrow$ Y $\in F^+$ and X $\rightarrow$ Z $\in F^+$, then X $\rightarrow$ YZ $\in F^+$.

FD5 Decomposition: if X $\rightarrow$ YZ $\in F^+$, then X $\rightarrow$ Y $\in F^+$ and X $\rightarrow$ Z $\in F^+$.

We observe that FD4 and FD5 are derivable from FD1, FD2 and FD3 and that when W = $\emptyset$, then FD3 is called the *transitivity rule*. The closure of F with respect to Lien and Atzeni's axiom system [15, 2], denoted by F^*, is the smallest set of FDs that contains F and satisfies FD1, FD2, FD4 and FD5.

A set of FDs G over R is a *cover* of a set of FDs F over R if $G^+ = F^+$. The closure of a set of attributes, X $\subseteq$ schema(R), with respect to Armstrong's axiom system and F, denoted as X_F^+ (or simply X^+ whenever F is understood from the context), is given by $X^+ = \cup$ {Y | X $\rightarrow$ Y $\in F^+$}. We note that X $\rightarrow X^+ \in F^+$ and that X^+ can be computed in linear time in the size of F [4]. In the sequel we use the equivalent statements A $\in X^+$ and X $\rightarrow$ A $\in F^+$, interchangeably.

Definition 3.3 (Reduced and canonical sets of FDs) An FD X $\rightarrow$ Y $\in F^+$ is *reduced* [4] if there does not exist a set of attributes W $\subset$ X such that W $\rightarrow$ Y $\in F^+$. A set of FDs F is *reduced* if all the FDs in F are reduced.

A set of FDs F is *canonical* if it is reduced and the right-hand sides of all the FDs in F are singletons.

We note that reduced and canonical covers G of a set of FDs F can be obtained in polynomial time in the size of F [4].

Definition 3.4 (Satisfaction of an FD) An FD X $\rightarrow$ Y is *weakly* satisfied (or simply satisfied whenever no ambiguity arises) in a relation r over R, denoted by r $\models_\approx$ X $\rightarrow$ Y, if and only if $\exists s \in$ POSS(r) such that $\forall t_1, t_2 \in$ s, if t_1[X] = t_2[X] then t_1[Y] = t_2[Y].

We note that the definition of weak satisfaction of an FD in a relation reduces to the standard definition of the satisfaction of an FD when the relation is complete [18] (in that case there exists exactly one s $\in$ POSS(r)). Furthermore, it follows from the above definition that r must be consistent in order to satisfy an FD. The reason for this condition is that we assume that only consistent relations are stored in the database. A more liberal definition would only insist that π_{XY}(r) is consistent in order that X $\rightarrow$ Y be satisfied.

The following lemma, which gives a syntactic characterisation of satisfaction of an FD, follows from Definition 3.4 (cf. Lemma 6.2 in [3]).

Lemma 3.1 Let r be a consistent relation over R and X, Y $\subseteq$ schema(R). Then r $\models_\approx$ X $\rightarrow$ Y if and only if $\forall t_1, t_2 \in$ r, if t_1[X] and t_2[X] are complete and t_1[X] = t_2[X], then t_1[Y] $\sqcup$ t_2[Y] is consistent. $\square$

Example 2 Let r be the relation shown in Table 4 and let F = {STUD $\rightarrow$ DEPT, COURSE $\rightarrow$ DEPT, DEPT $\rightarrow$ HEAD}. It can be verified that r $\models_\approx$ STUD $\rightarrow$ DEPT, r $\models_\approx$ COURSE $\rightarrow$ DEPT and r $\models_\approx$ DEPT $\rightarrow$ HEAD are all satisfied.

The next lemma follows directly from Lemma 3.1.

Lemma 3.2 The following statements, where r is a relation over R and X, Y $\subseteq$ schema(R), are true:

1. If Y $\subseteq$ X, then $\forall$ r $\in$ REL(R), if r is consistent, then r $\models_\approx$ X $\rightarrow$ Y.

2. r $\models_\approx$ X $\rightarrow$ Y if and only if r $\models_\approx$ X $\rightarrow$ Y$-$X.

3. If r $\models_\approx$ X $\rightarrow$ Y, then r $\models_\approx$ XW $\rightarrow$ YW, where W $\subseteq$ schema(R).

4. r $\models_\approx$ X $\rightarrow$ YZ if and only if r $\models_\approx$ X $\rightarrow$ Y and r $\models_\approx$ X $\rightarrow$ Z. $\square$

As a result of the above lemma we assume without loss of generality that *F does not contain any non-trivial FDs and that the right-hand sides of all the FDs in F are singletons.* This assumption will allow us to simplify the proofs of the results in Section 4.

We observe that Lemma 3.2 shows that satisfaction is closed under reflexivity (FD1), augmentation (FD2), union (FD4) and decomposition (FD5). On the other hand, as was demonstrated in Table 1, satisfaction is not closed under pseudo-transitivity (FD3).

It was shown in [1] (see also [18, 3]) that Armstrong's axiom system is sound and complete for complete relations. On the other hand, when considering incomplete relations Armstrong's axiom system is no longer sound, since as was just noted the pseudo-transitivity rule (and thus also the transitivity rule) is no longer sound in this case. A sound and complete axiom system for FDs holding in incomplete relations was obtained in [15, 2] by dropping the pseudo-transitivity rule from Armstrong's axiom system, thus obtaining the inference rules FD1, FD2, FD4 and FD5.

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.

Definition 3.5 (Satisfaction of a set of FDs) A set of FDs F over R is *weakly* satisfied (or simply satisfied whenever no ambiguity arises) in a relation r over R, denoted by r $\models_\approx$ F, if and only if $\exists s \in$ POSS(r) such that $\forall$ X $\rightarrow$ Y $\in$ F, s $\models_\approx$ X $\rightarrow$ Y.

The following lemma follows from the above definition and the fact that Armstrong's axiom system is sound and complete for complete relations.

Lemma 3.3 *Let r be a relation over R and F be a set of FDs over R. Then $r \models_\approx F$ if and only if $r \models_\approx F^+$.* $\square$

Definition 3.6 (Additive satisfaction) Following [3] we will say that satisfaction, is *additive* with respect to a class of relations, say **RC**, and a class of sets of FDs, say **FC**, whenever: $\forall$ r $\in$ **RC**, $\forall$ F $\in$ **FC**, r $\models_\approx$ F, if and only if there exists a reduced cover G of F such that $\forall$ X $\rightarrow$ Y $\in$ G, r $\models_\approx$ X $\rightarrow$ Y.

We note that additivity corresponds to the rule for introducing a conjunction on the right-hand side of a sequent in the sequent calculus [9]. The next example shows that in Definition 3.6 we cannot relax the condition that G is a reduced cover.

Example 3 Consider the relation r_1 shown in Table 1 and let F = {A $\rightarrow$ B, AB $\rightarrow$ C}. It can easily be verified that $r_1 \models_\approx$ A $\rightarrow$ B and $r_1 \models_\approx$ AB $\rightarrow$ C but $r_1 \not\models_\approx$ F. On the other hand, if we let G = {A $\rightarrow$ B, A $\rightarrow$ C}, i.e. G is a reduced cover of F, then $r_1 \not\models_\approx$ A $\rightarrow$ C.

An immediate consequence of the above definition is that satisfaction is additive with respect to COMPLETE(R) and FD(R). It is well known that satisfaction is *not* additive with respect to REL(R) and FD(R) due to the fact that transitivity is no longer sound for weak satisfaction [15, 2] (see Table 1). We also note, as we have shown in Tables 2 and 3, the fact that satisfaction is not additive with respect to REL(R) and FD(R) is not necessarily due to the lack of transitivity. The following proposition summarises these two facts.

Proposition 3.4 The following statements are true:

1. Satisfaction is additive with respect to COMPLETE(R) and FD(R).
2. Satisfaction is *not* additive with respect to REL(R) and FD(R). $\square$

The lack of additivity for satisfaction with respect to REL(R) and FD(R) gives rise to the problem that Lien and Atzeni's axiom system, which is sound and complete with respect to satisfaction of a single FD, does not, in general, cater for satisfaction of a set F of FDs. In the sequel we refer to this lack of additivity for satisfaction with respect to REL(R) and FD(R) as *the additivity problem.*

We close this section by mentioning that the time complexity of deciding whether r $\mid\!\approx$ F is polynomial in the sizes of r and F. A polynomial-time algorithm for deciding whether r $\mid\!\approx$ F, designated by CHASE(r, F), can be derived directly from Theorem 6.4 in [3] on using Theorem 3 in [11]. The algorithm is analogous to the standard chase procedure for FDs [11] 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), where F is canonical, which given the inputs r and F returns a relation whose tuples are more informative than the tuples of r, is presented in Algorithm 1.

Algorithm 1 (CHASE(r, F))

1. **begin**
2.   Tmp := r;
3.   i := 1;
4.   **for each** A $\in$ schema(R) **do**
5.   **for each** t $\in$ Tmp such that t[A] = *unk* **do**
6.   t[A] := $\perp_i$;
7.   i := i + 1;
8.   **end for**
9.   **end for**
10.   **while** $\exists t_1, t_2 \in$ Tmp **and** $\exists$ X $\rightarrow$ A $\in$ F such that t_1[X] = t_2[X] **and** t_1[A] $\neq$ t_2[A] **do**
11. t_1[A] := t_1[A] $\sqcup$ t_2[A];
12. t_2[A] := t_1[A] $\sqcup$ t_2[A];
13.   **end while**
14.   **return** Tmp;
15. **end.**

The following proposition states an important property of the chase procedure [3, 11], which will be useful in the next section. (Note that Dum $\subseteq$ *Dom*.)

Proposition 3.5 Let r be a consistent relation over R and F be a set of FDs over R. Then r $\mid\!\approx$ F if and only if CHASE(r, F) is consistent, or equivalently, CHASE(r, F) $\in$ POSS(r). $\square$

It can be verified that CHASE(r, F) is unique only up to the order in which the constants in Dum are assigned to the null attribute values in r. This is due to the fact that the for loops beginning at lines 4 and 5 of Algorithm 1 do not specify the order in which these statements are to be executed.

Example 4 Let r be the relation shown in Table 4 and let F = {STUD → DEPT, COURSE → DEPT, DEPT → HEAD}. CHASE(r, F) is shown in Table 5, where during the execution of CHASE(r, F) the occurrence of *unk* in the fifth tuple of r was replaced by $\perp_1$ and the occurrence *unk* in the sixth tuple of r was replaced by $\perp_2$. It can be verified that CHASE(r, F) is consistent and that r $\models_{\approx}$ F holds.

STUD	COURSE	DEPT	HEAD
Iris	Databases	Computing	Dan
Iris	Set_Theory	Computing	Dan
Reuven	Set_Theory	Computing	Dan
Naomi	Programming	Maths	Annette
Naomi	$\perp_1$	Maths	Annette
Eli	Logic	$\perp_2$	Brian

Table 5: The relation CHASE(r, F)

4 Monodependent sets of functional dependencies

In this section we solve the additivity problem for weak satisfaction of a set F of FDs. Intuitively, a set of FDs F over R is monodependent if for each attribute A ∈ schema(R) there exists a unique FD that functionally determines A and, in addition, only trivial cycles involving A, such as XA → A, arise between any two FDs one of which functionally determines A. We show that we can check whether F is monodependent in time polynomial in the size of F. We then solve the additivity problem by showing that weak satisfaction is additive if and only if F is monodependent. We believe that monodependent sets of FDs arise naturally in the real world, since they *avoid* ambiguity in the representation of the semantics of a set of FDs.

Definition 4.1 (A monodependent set of FDs) A set of FDs F, over R, is a *monodependent* set of FDs over R (or simply monodependent, whenever R is understood from the context) if ∀ A ∈ schema(R), the following two conditions are true:

1. Whenever there exist incomparable FDs, X → A, Y → A ∈ F^+, then X ∩ Y → A ∈ F^+.
2. Whenever there exist cyclic FDs, XB → A, YA → B ∈ F^+, then either Y → B ∈ F^+ or (X ∩ Y)A → B ∈ F^+.

An immediate consequence of the above definition is that if G is a cover of a set of FDs F over R, then F is monodependent if and only if G is monodependent.

We observe that the two defining conditions of monodependent sets of FDs correspond to the two defining properties of *conflict free* sets of *multi-valued dependencies* (MVDs) [17, 15, 5]. In particular, condition (1) corresponds to the *intersection property* and condition (2) corresponds to the *split-freedom property*. We further observe that the set of MVDs that are logically implied by a monodependent set of FDs may not be conflict free and thus monodependence is a weaker notion than conflict freedom. For example, let R = {A → B, B → A} be a set of FDs over R, with schema(R) = {A, B, C}. It can easily be verified that R is monodependent but that the set of MVDs logically implied by R are not conflict free.

Example 5 Examples of set of FDs that are *not* monodependent are: {A → B, B → C}, {A → C, B → C} and {B → A, AC → B}.

Examples of monodependent sets of FDs are: {A → B, C → D}, {BC → A, AC → B}, {B → A, A → B} and {BC → A, C → B}.

The next theorem shows that if F satisfies the intersection property, then the closure of F with respect to Armstrong's axiom system (i.e. F^+) is equal to the closure of F with respect to Lien and Atzeni's axiom system (i.e. F^*).

Theorem 4.1 If F is reduced and satisfies the intersection property then $F^+ = F^*$.

Proof. We assume without loss of generality that F is canonical, since, as noted after Definition 4.1, F is monodependent if and only if every cover of F is monodependent.

Obviously, $F^* \subseteq F^+$, since Lien and Atzeni's axiom system can be derived from Armstrong's axiom system. We conclude the proof by showing that $F^+ \subseteq F^*$; we use induction on the minimal number of times, say k, the pseudo-transitivity rule (FD3) is used in order to obtain F^+.

(*Basis*): If k = 0, then the result is immediate.

(*Induction*): Assume the result holds when the minimal number of times FD3 was used in obtaining F^+ is k, with k $\geq$ 0; we then need to prove that the result holds when the minimal number of times FD3 was used in obtaining F^+ is $k+1$. Suppose that the last time FD3 was used in the process of obtaining F^+ from F, the FD XW $\rightarrow$ Z was added to a state G of F^+, where X $\rightarrow$ Y $\in$ G, WY $\rightarrow$ Z $\in$ G, assuming that W $\cap$ Y = $\emptyset$.

Without loss of generality we can assume that Z = {A} is a singleton, since otherwise we can apply FD3 |Z| times, once for each A $\in$ Z, and then apply the union rule to obtain XW $\rightarrow$ Z $\in F^+$. Moreover, we can assume that A $\notin$ XW, since otherwise XW $\rightarrow$ A is a trivial FD and can be derived by using the reflexivity rule. In addition, we can assume that A $\notin$ Y, since otherwise the result follows by inductive hypothesis and the use of the decomposition rule if Y $\neq$ {A}.

Now, it is must be the case that X $\rightarrow$ Y is a non-trivial FD, since otherwise XW $\rightarrow$ A $\in F^+$ can be derived by using the augmentation and decomposition rules. Therefore, we have that Y $\not\subseteq$ X. Now, if it were the case that X and Y are incomparable, then W(Y $\cap$ X) $\rightarrow$ A $\in F^+$, since F satisfies the intersection property, with W(Y $\cap$ X) $\subset$ XW. The result follows, since by inductive hypothesis it must be the case that $\exists$ V $\subseteq$ W(Y $\cap$ X) such that V $\rightarrow$ U $\in$ G, with A $\in$ U, and therefore XW $\rightarrow$ A $\in F^+$ can be derived by using at some point in the derivation process the augmentation and decomposition rules.

Finally, if it were the case that X $\subset$ Y, then WY $\rightarrow$ A $\in F^+$ is not reduced. Now, suppose that WY $\rightarrow$ A can be reduced to V $\rightarrow$ A $\in F^+$. We claim that V = XW. Obviously XW $\subset$ V is not possible, since in this case V $\rightarrow$ A would not be reduced. Also, V and XW cannot be incomparable, since otherwise V $\rightarrow$ A would again not be reduced due to the fact that F satisfies the intersection property. So, assume that V $\subset$ XW. In this case the result follows, since by inductive hypothesis $\exists V' \subseteq$ V such that $V' \rightarrow$ U $\in$ G, with A $\in$ U, and therefore XW $\rightarrow$ A $\in F^+$ can be derived by using at some point in the derivation process the augmentation and decomposition rules. Therefore, our claim that V = XW is proved and thus XW $\rightarrow$ A $\in F^+$ is reduced.

Now, since XW $\rightarrow$ A $\in F^+$, there must exist a reduced FD $V' \rightarrow$ A $\in$ F. If $V' \subseteq$ XW the result follows, since XW $\rightarrow$ A $\in F^+$ can then be derived by using at some point in the derivation process the augmentation and decomposition rules. Next consider the case when V' and XW are incomparable. Since F satisfies the intersection property $V' \cap$ XW $\rightarrow$ A $\in F^+$, with ($V' \cap$ XW) $\subset$ XW; the result follows by inductive hypothesis and the fact that XW $\rightarrow$ A $\in F^+$ can be derived by using at some point in the derivation process the augmentation and decomposition rules. So it only remains to consider the case when XW $\subset V'$. However, this contradicts the fact that $V' \rightarrow$ A is reduced. □

The converse of Theorem 4.1 is, in general, false. For example, let F = {A $\rightarrow$ C, B $\rightarrow$ C} be a set of FDs over R, with schema(R) = {A, B, C}. It can be easily verified that $F^+ = F^*$ but that F does not satisfy the intersection property, since $\emptyset \rightarrow$ C $\notin F^+$.

The next lemma gives an alternative characterisation of a monodependent set of FDs rephrased in terms of a canonical set of FDs.

Lemma 4.2 A set of FDs F over R is monodependent if and only if $\forall$ A $\in$ schema(R), the following two conditions are true, where G is a canonical cover of F:

1. There exists at most one FD in G of the form X $\rightarrow$ A, and if X $\rightarrow$ A $\in$ G, then $\forall$ B $\in$ X, A $\notin$ (schema(R) $-$ AB)$^+$.
2. Whenever X $\rightarrow$ A, Y $\rightarrow$ B $\in$ G, then either A $\notin$ Y or Y $\subseteq$ (X $\cap$ Y)A.

Proof. If: Assume that the above two conditions are true and that G is a canonical cover of F. For the first part of the definition of monodependence assume that there exist incomparable FDs, X $\rightarrow$ A, Y $\rightarrow$ A $\in F^+$. We need to show that X $\cap$ Y $\rightarrow$ A $\in F^+$. Let W $\rightarrow$ A $\in$ G be the single FD in G that functionally determines A.

Assume that W $\not\subseteq$ X and thus W $\neq \emptyset$. Thus $\exists$ B $\in$ W such that B $\notin$ X. Therefore we deduce that A $\in$ (schema(R) $-$ AB)$^+$, since X $\subseteq$ schema(R) $-$ AB. A contradiction has arisen implying that W $\subseteq$ X. Similarly, we can deduce that W $\subseteq$ Y. Since W $\subseteq$ X and W $\subseteq$ Y, it follows that W $\subseteq$ XY; since W $\rightarrow$ A $\in$ G, this implies that X $\cap$ Y $\rightarrow$ A $\in F^+$.

For the second part of the definition of monodependence assume that there exist cyclic FDs, XB $\rightarrow$ A, YA $\rightarrow$ B $\in F^+$. We need to show that either Y $\rightarrow$ B $\in F^+$ or (X $\cap$ Y)A $\rightarrow$ B $\in F^+$. Let W $\rightarrow$ A, Z $\rightarrow$ B $\in$ G be the two FDs in G that functionally determine A and B, respectively.

Assume that W $\not\subseteq$ XB and thus W $\neq \emptyset$. Thus $\exists$ C $\in$ W such that C $\notin$ XB. Therefore we deduce that A $\in$ (schema(R) $-$ AC)$^+$, since XB $\subseteq$ schema(R) $-$ AC. A contradiction has arisen to the first condition of this lemma implying that W $\subseteq$ XB. Similarly, Z $\subseteq$ YA is implied, since A $\neq$ B. The result follows, since either of the following two assertions *are* true. Firstly, A $\notin$ Z implies that Z $\subseteq$ Y and thus Y $\rightarrow$ B $\in F^+$. Secondly, Z $\subseteq$ (W $\cap$ Z)A implies that (W $\cap$ Z)A $\rightarrow$ B $\in F^+$ is a non-trivial FD. Therefore, (XB $\cap$ YA)A $\rightarrow$ B $\in F^+$, since (W $\cap$ Z)A $\subseteq$ (XB $\cap$ YA)A, and thus (X $\cap$ Y)A $\rightarrow$ B $\in F^+$, since A $\notin$ X and B $\notin$ Y.

Only if: Assume that F is monodependent. For the first condition above, it follows that there exists at most one FD in G of the form X $\rightarrow$ A, since G is canonical. Assume that X $\rightarrow$ A $\in$ G. Next, if X = schema(R) $-$ A or X = $\emptyset$, then trivially $\forall$ B $\in$ X, A $\notin$ (schema(R) $-$ AB)$^+$. So, assume that $\emptyset \subset$ X $\subset$ schema(R) $-$ A. Next, let W = schema(R) $-$ AB be a set of attributes such that B $\in$ X and assume to the contrary that A $\in W^+$. It follows that X and W are incomparable and thus W $\cap$ X $\subset$ X. However, W $\cap$ X $\rightarrow$ A $\notin$ G, since G is canonical. This contradicts the fact that F is monodependent. Therefore $\forall$ B $\in$ X, A $\notin$ (schema(R) $-$ AB)$^+$.

For the second condition above, assume that X $\rightarrow$ A, Y $\rightarrow$ B $\in$ G but that both A $\in$ Y and Y $\not\subseteq$ (X $\cap$ Y)A. Assume that B $\notin$ X. Then we can derive X(Y $-$ A) $\rightarrow$ B $\in F^+$ by using the pseudo-transitivity rule. Now, Y $\not\subseteq$ X(Y $-$ A), since A $\notin$ X. Furthermore, X(Y $-$ A) $\not\subseteq$ Y, since X(Y $-$ A) $\subset$ Y would contradict the fact that G is canonical. Therefore, Y and X(Y $-$ A) are incomparable and thus by the first part of the definition of monodependence Y $\cap$ (X(Y $-$ A)) $\rightarrow$ B $\in F^+$. This leads to a contradiction of G being canonical, since Y $\cap$ (X(Y $-$ A)) $\subset$ Y.

Assume that B $\in$ X. A contradiction to F being monodependent arises as follows. The assumption that A $\in$ Y implies that Y $-$ A $\rightarrow$ B $\notin F^+$, since G is canonical. Now, Y $\not\subseteq$ (X $\cap$ Y)A implies that (X $\cap$ Y)A $\subset$ Y. Therefore, it is also true that (X $\cap$ Y)A $\rightarrow$ B $\notin F^+$, since G is canonical. The result follows, since either A $\notin$ Y or Y $\subseteq$ (X $\cap$ Y)A must hold. $\square$

The following theorem utilises the previous lemma.

Theorem 4.3 *Monodependence of a set of FDs F over R can be checked in time polynomial in the size of F.*

Proof. It was shown in [16] that a canonical cover G of a set of FDs F over R can be obtained in time polynomial in the size of F. The result follows by Lemma 4.2. $\square$

The following lemma shows that monodependence implies additivity.

Lemma 4.4 Let F be a set of monodependent FDs over R. Then $\forall r \in$ REL(R), if $\forall$ X $\rightarrow$ Y $\in$ F, r $\models_{\approx}$ X $\rightarrow$ Y, then r $\models_{\approx}$ F.

Proof. Assume without loss of generality that F is a canonical set of FDs, noting that a canonical cover is reduced. We show that if $\exists$ r $\in$ REL(R) such that $\forall$ X $\rightarrow$ Y $\in$ F, r $\models_{\approx}$ X $\rightarrow$ Y, but r $\not\models_{\approx}$ F, then F cannot be monodependent as assumed. Let r be such a relation. By Proposition 3.5 it follows that CHASE(r, F) is inconsistent. We call an execution of lines 11 and 12 in Algorithm 1 a *chase step* and we say that the chase step *applies* the FD X $\rightarrow$ A $\in$ F to the tuples t_1 and t_2. We conclude the result by induction on the minimal number of chase steps, k, required to show that CHASE(r, F) is inconsistent.

(*Basis*): At least two chase steps are needed to show that CHASE(r, F) is inconsistent, since by assumption $\forall$ X $\rightarrow$ Y $\in$ F, r $\models_{\approx}$ X $\rightarrow$ Y. Thus consider the case when k = 2. Suppose the second chase step applies the FD X $\rightarrow$ A $\in$ F to the tuples t_1 and t_2. Then after the first chase step is applied t_1[X] = t_2[X] but t_1[A] $\sqcup$ t_2[A] is inconsistent. Without loss of generality assume that the first chase step applies the FD Y $\rightarrow$ B $\in$ F to the tuples t_2 and t_3 in r. There are two cases to consider.

Case 1: B = A. In this case the first chase step applies the FD Y $\rightarrow$ A to t_2 and t_3 resulting in t_1[A] $\sqcup$ t_2[A] being inconsistent. It follows that t_1[A] $\sqcup$ t_3[A] is inconsistent in r, since otherwise t_1[A] $\sqcup$ t_2[A] is inconsistent in r implying that one chase step is sufficient to show that CHASE(r, F) is inconsistent. Now, if X = Y, then one chase step which applies the FD X $\rightarrow$ A to the tuples t_1 and t_3 in r is sufficient to show that CHASE(r, F) is inconsistent, thus X $\neq$ Y. The result that F is not monodependent now follows by the assumption that F is canonical, since X $\rightarrow$ A and Y $\rightarrow$ A must be incomparable FDs and X $\cap$ Y $\rightarrow$ A $\notin$ F^+.

Case 2: B $\neq$ A. In this case the first chase step applies Y $\rightarrow$ B to t_2 and t_3 in r resulting in t_1[B] = t_2[B]. Furthermore, t_1[A] $\sqcup$ t_2[A] is inconsistent in r, since k = 2. It follows that B $\in$ X, otherwise one chase step would suffice to show that CHASE(r, F) is inconsistent.

Now suppose that A $\notin$ Y. Then we can derive (X $-$ B)Y $\rightarrow$ A $\in$ F^+ by pseudo-transitivity, where X $\not\subseteq$ (X $-$ B)Y, since B $\notin$ Y. Now, if (X $-$ B)Y $\not\subseteq$ X, then (X $-$ B)Y $\rightarrow$ A and X $\rightarrow$ A are incomparable. Furthermore, (X $-$ B) $\subset$ X and therefore (X $-$ B) $\rightarrow$ A $\notin$ F^+, since F is canonical. Thus, there exists a canonical set of FDs G such that W $\rightarrow$ A, X $\rightarrow$ A $\in$ G, where W $\subseteq$ (X $-$ B)Y. The result that F is not monodependent follows from Case 1 by replacing F with G. Therefore, we assume that (X $-$ B)Y $\subseteq$ X, implying that (X $-$ B)Y $\subset$ X, since B $\notin$ Y. This contradicts the fact that F is canonical, since (X $-$ B)Y $\rightarrow$ A $\in$ F^+.

We therefore suppose that A $\in$ Y. The result that F is not monodependent now follows, since Y $\rightarrow$ B and X $\rightarrow$ A are cyclic FDs and in addition the following two assertions are true.

Firstly, (X $-$ B) $\rightarrow$ A $\notin$ F^+ due to the fact that by assumption F is canonical. Secondly, t_1[B] = t_3[B] in r, since Y $\rightarrow$ B was applied to t_2 and t_3 and k = 2. Moreover, t_1[(X $\cap$ Y)] = t_3[(X $\cap$ Y)] in r, since both t_1[(X $\cap$ Y)] = t_2[(X $\cap$ Y)] in r and t_2[(X $\cap$ Y)] = t_3[(X $\cap$ Y)] in r. Furthermore, t_2[A] = t_3[A] in r, since A $\in$ Y, and therefore t_1[A] $\sqcup$ t_3[A] is inconsistent in r. Thus, if (X $\cap$ Y)B $\rightarrow$ A $\in$ F^+, then due to the assumption that F is canonical one chase step which applies the FD (X $\cap$ Y)B $\rightarrow$ A to t_1 and t_2 would suffice to show that CHASE(r, F) is inconsistent. It therefore follows that (X $\cap$ Y)B $\rightarrow$ A $\notin$ F^+ which is equivalent to ((X $-$ B) $\cap$ (Y $-$ A))B $\rightarrow$ A $\notin$ F^+, since B $\notin$ Y and A $\notin$ X. (Recall that (X $-$ B) $\rightarrow$ A $\notin$ F^+.) The result that F is not monodependent follows.

(*Induction*): Assume the result holds when the minimal number of chase steps required to show that CHASE(r, F) is inconsistent is k, with k $\geq$ 2; we then need to prove that the result holds when the minimal number of chase steps required to show that CHASE(r, F) is inconsistent is k+1. Suppose the last chase step applies the FD X $\rightarrow$ A $\in$ F to the tuples t_1 and t_2 in the penultimate state of r during the execution of CHASE(r, F). Then after the penultimate chase step is applied t_1[X] = t_2[X] but t_1[A] $\sqcup$ t_2[A] is inconsistent. Without loss of generality assume that the penultimate chase step applies the FD Y $\rightarrow$ B $\in$ F to the tuples t_2 and t_3 in the state prior to the penultimate state of r during the execution of CHASE(r, F).

The result follows by an argument similar to the basis step noting that if k or less steps are sufficient to show that CHASE(r, F) is inconsistent, then the result follows by inductive hypothesis. □

The following lemma shows that additivity implies monodependence.

Lemma 4.5 Let F be a set of FDs over R and assume that $\forall r \in$ REL(R), if $\forall$ X $\rightarrow$ Y $\in$ F, r $\models\approx$ X $\rightarrow$ Y, then r $\models\approx$ F. Then F is monodependent.

Proof. We show that if F is not monodependent, then $\exists$ r $\in$ REL(R) such that $\forall$ X $\rightarrow$ Y $\in$ F, r $\models\approx$ X $\rightarrow$ Y but r $\not\models\approx$ F. There are two cases to consider.

Case 1: There exist incomparable FDs X $\rightarrow$ A, Y $\rightarrow$ A $\in F^+$ but X $\cap$ Y $\rightarrow$ A $\notin F^+$. Let Rest = schema(R) $-$ XYA and let r $\in$ REL(R) be the relation shown in Table 6.

It can be verified that $\forall$ V $\rightarrow$ T $\in$ F, r $\models\approx$ V $\rightarrow$ T, since the only FDs that are not weakly satisfied in r are of the form W $\rightarrow$ A, where W $\subseteq$ X $\cap$ Y, which is justified by Lemma 3.2 due to the assumption that X $\cap$ Y $\rightarrow$ A $\notin F^+$. The result now follows due to the fact that r $\not\models\approx$ F, since it can easily be verified that $\forall$ s $\in$ POSS(r), either s $\not\models\approx$ X $\rightarrow$ A or s $\not\models\approx$ Y $\rightarrow$ A.

X $\cap$ Y	X $-$ Y	Y $-$ X	A	Rest
$0 \ldots 0$	$0 \ldots 0$	$unk \ldots unk$	0	$unk \ldots unk$
$0 \ldots 0$	$0 \ldots 0$	$0 \ldots 0$	unk	$unk \ldots unk$
$0 \ldots 0$	$unk \ldots unk$	$0 \ldots 0$	1	$unk \ldots unk$

Table 6: The relation pertaining to Case 1

Case 2: There exist cyclic FDs, XB $\rightarrow$ A, YA $\rightarrow$ B $\in F^+$, however Y $\rightarrow$ B $\notin F^+$ and (X $\cap$ Y)A $\rightarrow$ B $\notin F^+$. Let Rest = schema(R) $-$ XYAB and let r $\in$ REL(R) be the relation shown in Table 7.

The only FDs that are not weakly satisfied in r are of the form W $\rightarrow$ B, where W $\subseteq$ Y, or of the form W $\rightarrow$ B, where W $\subseteq$ (X $\cap$ Y)A. These violations are justified by Lemma 3.2 due to the assumption that Y $\rightarrow$ B $\notin F^+$ and (X $\cap$ Y)A $\rightarrow$ B $\notin F^+$. The result now follows due to the fact that r $\not\models\approx$ F, since it can easily be verified that $\forall$ s $\in$ POSS(r), s $\not\models\approx$ YA $\rightarrow$ B. □

X $\cap$ Y	X $-$ Y	Y $-$ X	A	B	Rest
$0 \ldots 0$	$0 \ldots 0$	$unk \ldots unk$	0	0	$unk \ldots unk$
$0 \ldots 0$	$0 \ldots 0$	$0 \ldots 0$	unk	0	$unk \ldots unk$
$0 \ldots 0$	$unk \ldots unk$	$0 \ldots 0$	0	1	$unk \ldots unk$

Table 7: The relation pertaining to Case 2

The following theorem summarises Lemmas 4.4 and 4.5.

Theorem 4.6 Weak satisfaction is additive with respect to REL(R) and a class of sets of FDs **FC** if and only if all the sets of FDs in **FC** are monodependent. □

5 Concluding Remarks

We have solved the additivity problem by showing in Theorem 4.6 that the largest class of sets FDs for which additivity holds with respect to the set of all relations over a fixed schema is the class of monodependent sets of FDs. Furthermore, by Theorem 4.3 monodependence of a set of FDs F can be checked in time polynomial in the size of the F.

Weak satisfaction corresponds to *possibility*, that is the existence of a possible relation that satisfies a set of FDs. It is also possible to define *strong satisfaction* which corresponds to *necessity*, that is to say the situation when all possible relations satisfy a set of FDs. Strong satisfaction has the advantage over its weak counterpart, since in this case the satisfaction of FDs need not be rechecked after each update of a null value to a non-null value in an incomplete relation. On the other hand, strong satisfaction is, in general, stricter that weak satisfaction. A sound and complete axiom system for FDs which caters for both strong and weak satisfaction and a polynomial time algorithm for the implication problem thereof can be found in [13].

It would be an interesting research topic to extend the results presented herein to *or-sets* [12], i.e. allowing, instead of any occurrence of *unk*, a finite set of possible values, one of which is the true value. Another interesting research topic, which was taken up in [14], is an extension of the formalism presented herein to deal with inclusion dependencies [6] in the presence of incomplete information.

Acknowlegements. The authors would like to thank the referees (a multiset of five) for their constructive criticism and helpful remarks.

References

[1] Armstrong, W.W.: Dependency structures of data base relationships. *Proceedings of the IFIP Congress*, Stockholm, pp. 580-583 (1974)

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

[3] Atzeni, P. and De Antonellis, V.: Relational database theory. Redwood City, CA: Benjamin/Cummings, 1993

[4] Beeri, C., Bernstein, P.A.: Computational problems related to the design of normal form relational schemas. *ACM Transactions on Database Systems* **4**, 30-59 (1979)

[5] Beeri, C., Fagin, R., Maier, D., Yannakakis, M.: On the desirability of acyclic database schemes. *Journal of the ACM* **30**, 479-513 (1983)

[6] Casanova, M.A., Fagin, R., Papadimitriou, C.H.: Inclusion dependencies and their interaction with functional dependencies. *Journal of Computer and System Sciences* **28**, 29-59 (1984)

[7] Codd, E.F.: Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems* **4**, 397-434 (1979)

[8] Davey, B.A., Priestly, H.A.: Introduction to lattices and order. Cambridge, U.K.: Cambridge University Press, 1990

[9] Fitting, M.: First-order logic and automated theorem proving. Berlin: Springer-Verlag, 1990

[10] Garey, M.R., Johnson, D.S.: Computers and intractability: A guide to the theory of NP-completeness. New York: W.H. Freeman, 1979

[11] Honeyman, P.: Testing satisfaction of functional dependencies. *Journal of the ACM* **29**, 668-677 (1982)

[12] Imielinski, T.: Incomplete information in logical databases. *IEEE Data Engineering* **12**, 29-40 (1989)

[13] Levene, M., Loizou, G.: Axiomatisation of functional dependencies in incomplete relations. Research Note RN/93/87, Department of Computer Science, University College London, 1993

[14] Levene, M., Loizou, G.: Null inclusion dependencies in relational databases. Research Note RN/94/26, Department of Computer Science, University College London, 1994

[15] Lien, Y.E.: On the equivalence of database models. *Journal of the ACM* **29**, 333-362 (1982)

[16] Maier, D.: Minimum covers in the relational database model. *Journal of the ACM* **27**, 664-674 (1980)

[17] Sciore, E.: Real world mvd's. *Proceedings of ACM SIGMOD Conference on Management of Data*, Ann Arbor, pp. 121-132 (1981)

[18] Ullman, J.D.: Principles of database and knowledge-base systems, Volume I. Rockville, MD, Computer Science Press, 1988