On the Role of the Complementation Rule for Data Dependencies over Incomplete Relations

Flavio Ferrarotti[1], Sven Hartmann[2], and Sebastian Link[1]

[1] School of Information Management, Victoria University of Wellington, New Zealand
[2] Institut für Informatik, Technische Universität Clausthal, Germany

Abstract. Recently, an axiomatization for functional dependencies (FDs) and multivalued dependencies (MVDs) has been established where arbitrary attributes can be specified as NOT NULL. That is, the information stored over such attributes must not be incomplete. The axiomatization subsumes previous axiomatizations of FDs and MVDs where every attribute is declared to be NOT NULL, and where no attribute is declared to be NOT NULL. We establish axiomatizations which underpin formally the intuition that the complementation rule is a mere means of database normalization. The results unburden the existing theory of the strong assumption that all attributes are known at the time when the dependencies are specified. The findings extend and unify previous results for the special cases above.

1 Introduction

A database system manages a collection of persistent information in a shared, reliable, effective and efficient way. Most commercial database systems are still founded on *the relational model of data* [10]. Data administrators utilize various classes of data dependencies to restrict the relations in the database to those considered meaningful to the application at hand. According to [12] functional dependencies (FDs) capture around two-thirds, and multivalued dependencies (MVDs) around one-quarter of all uni-relational dependencies (those defined over a single relation schema) that arise in practice. In particular, MVDs are frequently exhibited in database applications [37], e.g. after denormalization or in views [1]. While research on this topic has been extensive, only very recently a theory has been established that can reason about FDs and MVDs exhibited by relations that satisfy arbitrary NOT NULL constraints [21].

Example 1. Consider a table SUPPLIES with column headers *A(rticle)*, *S(upplier)*, *L(ocation)* and *C(ost)*. The table collects information about suppliers that deliver articles from a location at a certain cost.

```
CREATE TABLE Supplies
    (Article CHAR[20],
    Supplier VARCHAR NOT NULL,
    Location VARCHAR NOT NULL,
    Cost CHAR[8]);
```

A. Dawar and R. de Queiroz (Eds.): WoLLIC 2010, LNAI 6188, pp. 136–147, 2010.

Suppose the database management system enforces the following constraints: The FD $A \rightarrow S$ says that for every article there is a most one supplier, the FD $AL \rightarrow C$ says that the costs are determined by the article and the location, and the MVD $S \twoheadrightarrow AC$ says that the supplier determines the article and cost pairs independently of the location. Do the following meaningful constraints also need to be enforced explicitly, or are they already enforced implicitly: i) the MVD $A \twoheadrightarrow L$ and ii) the FD $A \rightarrow C$? □

Indeed, the declaration of *Supplier* and *Location* as NOT NULL guarantees that both $A \twoheadrightarrow L$ and $A \rightarrow C$ are implied by $A \rightarrow S$; $AL \rightarrow C$ and $S \twoheadrightarrow AC$. However, reasoning about FDs and MVDs in the presence of an arbitrary null-free subschema (NFS), i.e. the set of attributes declared NOT NULL, is subtle. For example, if S is not declared NOT NULL, then neither the FD nor the MVD is implied. Consequently, the opportunity to specify an arbitrary NFS provides the data administrator with a flexible mechanism to control the expressiveness of the consequence relation. Dedicated tools for reasoning about FDs and MVDs in the presence of arbitrary NFSs have been established [21]. The set

$$\mathfrak{D} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{U}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{C}^R_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}\}$$

of inference rules from Table 1 is a finite axiomatization [21].

Example 2. Let $R = ASLC$, $R_s = SL$, $\Sigma = \{A \rightarrow S; AL \rightarrow C; S \twoheadrightarrow AC\}$ as in Example 1. The inference

$$\dfrac{\dfrac{\dfrac{\dfrac{A \rightarrow S}{\mathcal{I}_{\mathrm{FM}}: A \twoheadrightarrow S} \quad \dfrac{S \twoheadrightarrow AC}{\mathcal{C}^R_{\mathrm{M}}: S \twoheadrightarrow L}}{\mathcal{T}_{\mathrm{M}}: A \twoheadrightarrow L} \quad \dfrac{\dfrac{}{\mathcal{R}_{\mathrm{F}}: A \rightarrow A}}{\mathcal{I}_{\mathrm{FM}}: A \twoheadrightarrow A}}{\mathcal{U}_{\mathrm{M}}: A \twoheadrightarrow AL} \quad AL \rightarrow C}{\mathcal{T}_{\mathrm{FM}}: A \rightarrow C}$$

shows that $A \twoheadrightarrow L$ and $A \rightarrow C$ can be inferred from Σ by $\mathfrak{D}$. Since $\mathfrak{D}$ is sound, in particular, it follows that both dependencies are implied by Σ. □

The inference in Example 2 can be criticized in two different aspects. When inferring MVDs, then applications of the R-complementation rule $\mathcal{C}^R_M$ should be restricted to the very last step of the inference (if necessary at all). The ability to have inferences with this property for all implied MVDs would establish an axiomatization that appropriately reflects the database normalization process. Moreover, for every implied FD there should be an inference with no applications of the R-complementation rule $\mathcal{C}^R_M$ at all. The desirability of these two features has already been motivated and axiomatizations with these features have been established for the special cases where every attributes is NOT NULL [8,9,29] and where every attribute is NULL [30]. In this paper, we will show that the axiomatization $\mathfrak{D}$ has neither of these features. Subsequently, we will establish a finite axiomatization with both features. The results provide a unifying framework for all previous findings on these issues.

Example 3. Suppose the four attributes Article, Supplier, Location and Cost only constitute the fragment of a view that we are currently aware of. That is, the underlying schema information is undetermined. Therefore, applications of the MVD complementation rule are not sound in this setting since the underlying universe is no longer known. For example, while the MVD $S \twoheadrightarrow AC$ implies the MVD $S \twoheadrightarrow L$ over the schema SUPPLIES, $S \twoheadrightarrow AC$ does not imply $S \twoheadrightarrow L$ when the schema is undetermined. Consequently, if $S \twoheadrightarrow L$ is perceived as a meaningful semantic constraint that must be enforced by the DBMS, then we need to specify this MVD explicitly as well. □

As a second contribution of this paper we establish a finite axiomatization for the combined class of FDs, MVDs and arbitrary null-free subschema over undetermined universes. Again, this result subsumes all previous findings on this subject, in particular the special case where every attribute is NOT NULL [8,9,29] and the special case where every attribute is NULL [30].

Organization. We summarize previous work in Section 2. The basic definitions are given in Section 3. In Section 4 we establish an axiomatization of FDs and MVDs in the presence of an arbitrary NFS that enjoys both of the features. We establish an axiomatization over undetermined universes in Section 5. We conclude in Section 6.

2 Related Work

Data dependencies have been studied thoroughly in the relational model of data, cf. [1]. Applications comprise almost the full range of database topics, e.g. normalization, requirements engineering and schema validation, data mining, database security, view maintenance and query optimization. They have received considerable attention in other data models as well . New application areas involve data cleaning, data transformations, consistent query answering, data exchange and data integration.

FDs capture around two-thirds and MVDs around one-quarter of all uni-relational dependencies that arise in applications [12,37]. For total relations, Armstrong [3] established the first axiomatization for FDs. Beeri, Fagin, and Howard extended this axiomatization to the combined class of FDs and MVDs [6]. Biskup [8], Link [29] and Biskup/Link [9] studied notions of FD and MVD implication where the underlying set of attributes is not fixed. In the same papers, axiomatizations were presented that clarify the role of the R-complementation rule as a mere means of database normalization [8,9,29]. In general, axiomatizations can be applied by designers and administrators to validate the specification of explicit knowledge, to design and fine-tune databases or to optimize queries. An axiomatization ensures that all opportunities of utilizing implicit knowledge have been exploited. An analysis of the completeness argument can provide invaluable hints for finding algorithms that efficiently decide the implication problem.

One of the most important extensions of Codd's basic relational model [10] is incomplete information [11,23,26]. This is mainly due to the high demand for the correct handling of such information in real-world applications. Approaches to deal with incomplete information comprise incomplete relations, or-relations or fuzzy relations. In this paper we focus on incomplete relations. In the literature many kinds of null values have been proposed; for example, "missing" or "value unknown at present" [15], "non-existence" [31], "inapplicable" [15], "no information" [38] and "open" [14]. Most of the previous work on data dependencies is based on Zaniolo's no-information interpretation. This interpretation is valid for most database instances that occur in practice, since SQL allows only one unmarked null value. Consequently, the no information interpretation can model missing as well as incomplete information. Only recently, the set $\mathfrak{D} = \{\mathcal{R}_{\rm F}, \mathcal{D}_{\rm F}, \mathcal{U}_{\rm F}, \mathcal{U}_{\rm M}, \mathcal{T}_{\rm M}, \mathcal{C}^R_{\rm M}, \mathcal{I}_{\rm FM}, \mathcal{T}_{\rm FM}\}$ was shown to form an axiomatization for the combined class of FDs and MVDs in the presence of an arbitrary null-free subschema R_s [21], cf. Table 1. Moreover, it was shown [21] that the implication of FDs and MVDs in the presence of an arbitrary NFS R_s is equivalent to that of a fragment in Cadoli and Schaerf's R_s-3 logics [32]. The theory has unified previously orthogonal frameworks. For example, Beeri, Fagin and Howard's axiomatization [6] is covered when every attribute is NOT NULL, i.e. when $R_s = R$. Lien's axiomatization [27] is subsumed as the special case where every attribute is NULL, i.e., when $R_s = \emptyset$. Finally, Atzeni and Morfuni's axiomatization [4] $\mathfrak{AM} = \{\mathcal{R}_{\rm F}, \mathcal{D}_{\rm F}, \mathcal{U}_{\rm F}, \mathcal{T}_{\rm F}\}$ for FDs in the presence of an arbitrary NFS R_s is also subsumed. Link [30] presented an axiomatization for the class of MVDs that clarifies the role of the R-complementation rule, but only for the special case where every attribute is NOT NULL.

3 Preliminaries

We summarize the basic notions of data dependencies over partial relations.

Let $\mathfrak{A} = \{A_1, A_2, \ldots\}$ be a (countably) infinite set of distinct symbols, called attributes (column names). A *relation schema* is a finite non-empty subset R of $\mathfrak{A}$. Each attribute A of a relation schema R is associated with an infinite domain $dom(A)$ which represents the possible values that can occur in column A. To encompass incomplete information every column may have a null value, denoted by $\mathtt{ni} \in dom(A)$. The intention of $\mathtt{ni}$ is to mean "no information". This interpretation can model missing as well as incomplete information [4,38].

For attribute sets X and Y we may write XY for $X \cup Y$. If $X = \{A_1, \ldots, A_m\}$, then we may write $A_1 \cdots A_m$ for X. In particular, we may write simply A to represent the singleton $\{A\}$. A *tuple* over R (R-tuple or simply tuple, if R is understood) is a function $t : R \to \bigcup_{A \in R} dom(A)$ with $t(A) \in dom(A)$ for all $A \in R$. The null value occurrence $t(A) = \mathtt{ni}$ associated with an attribute A in a tuple t means that no information is available about the attribute A for the tuple t. For $X \subseteq R$ let $t[X]$ denote the restriction of the tuple t over R to X. A (partial) *relation* r over R is a finite set of tuples over R. Let t_1 and t_2 be two tuples over R. It is said that t_1 *subsumes* t_2 if for every attribute $A \in R$, $t_1[A] = t_2[A]$ or $t_2[A] = \mathtt{ni}$ holds. In consistency with previous work [4,27,38], the following restriction will be imposed, unless stated otherwise: No relation shall contain two tuples t_1 and t_2 such that t_1 subsumes t_2. With no null values present this means that no duplicate tuples occur.

For a tuple t over R and a set $X \subseteq R$, t is said to be X-total if for all $A \in X$, $t[A] \neq \mathtt{ni}$. Similar, a relation r over R is said to be X-total, if every tuple t of r is X-total. A relation r over R is said to be a *total relation* if it is R-total.

We recall projection and join operations [4,27]. Let r be some relation over R. Let X be some subset of R. The *projection* $r[X]$ of r on X is the set of tuples t for which (i) there is some $t_1 \in r$ such that $t = t_1[X]$ and (ii) there is no $t_2 \in r$ such that $t_2[X]$ subsumes t and $t_2[X] \neq t$. For $Y \subseteq X$, the *Y-total projection* $r_Y[X]$ of r on X is $r_Y[X] = \{t \in r[X] \mid t \text{ is } Y\text{-total}\}$. Given an X-total relation r over R and an X-total relation s over S such that $X = R \cap S$ the *natural join* $r \bowtie s$ of r and s is the relation over $R \cup S$ which contains those tuples t for which there are tuples $t_1 \in r$ and $t_2 \in s$ with $t_1 = t[R]$ and $t_2 = t[S]$ [4,27].

Functional dependencies are important for the relational [5,7,10] and other data models [2,16,17,18,19,20,22,24,25,28,33,34,35,36]. According to Lien [27], a *functional dependency with nulls* (FD) over R is a statement $X \to Y$ where $X, Y \subseteq R$. The FD $X \to Y$ over R is satisfied by a relation r over R ($\models_r X \to Y$) if and only if for all $t_1, t_2 \in r$ the following holds: if t_1 and t_2 are X-total and $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$. For total relations the FD definition reduces to the standard definition of a functional dependency [1], and so is a sound generalization. It is also consistent with the no-information interpretation [4,27].

In fact, tuples with nulls in attributes in X cannot cause a violation of the FD $X \to Y$: the nulls mean that no information is available about those attributes. Two X-total tuples t_1, t_2 where $t_1[X] = t_2[X]$ and t_2 is A-total while t_1 is not, violate any FD $X \to Y$ with $A \in Y$: t_1 indicates that no information is available about the value for A associated with $t_1[X]$, while t_2 indicates that the value for A associated with $t_2[X] = t_1[X]$ does exist. Hence, it violates the natural requirement of an FD that if the values for X are the same for two tuples, both tuples must contain the same information for the attributes in Y.

According to Lien [27], a *multivalued dependency with nulls* (MVD) over R is a statement $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. The MVD $X \twoheadrightarrow Y$ over R is satisfied by a relation r over R ($\models_r X \twoheadrightarrow Y$) if and only if for all $t_1, t_2 \in r$ the following holds: if t_1 and t_2 are X-total and $t_1[X] = t_2[X]$, then there is some $t \in r$ such that $t[XY] = t_1[XY]$ and $t[X(R-Y)] = t_2[X(R-Y)]$. Informally, the relation r satisfies $X \twoheadrightarrow Y$ when every X-total value determines the set of values on Y independently of the set of values on $R-Y$. It has been shown that $\models_r X \twoheadrightarrow Y$ if and only if $r_X[R] = r_X[XY] \bowtie r_X[X(R-Y)]$ [27]. Again, the MVD definition is a sound generalization of the standard definition over total relations [13].

Following Atzeni and Morfuni [4], a *null-free subschema* (NFS) over the relation schema R is a an expression R_s where $R_s \subseteq R$. The NFS R_s over R is satisfied by a relation r over R ($\models_r R_s$) if and only if r is R_s-total. SQL allows the specification of attributes as NOT NULL, cf. Example 1. Hence, the set of attributes declared NOT NULL forms the single NFS over the underlying relation schema.

For a set Σ of constraints over some relation schema R, we say that a relation r over R *satisfies* Σ ($\models_r \Sigma$) if r satisfies every $\sigma \in \Sigma$. If for some $\sigma \in \Sigma$ the relation r does not satisfy σ we say that r violates σ (and violates Σ) and write $\not\models_r \sigma$ ($\not\models_r \Sigma$). We will consider different classes $\mathcal{C}$ of constraints over a single relation schema, e.g. FDs and MVDs.

In schema design data dependencies are normally specified as semantic constraints on the relations intended to be instances of the schema.

During the design process or the lifetime of a database one usually needs to determine further dependencies which are implied by the given ones. Let R be a relation schema, let $R_s \subseteq R$ denote an NFS over R, and let $\Sigma \cup \{\varphi\}$ be a set of data dependencies over R in the class $\mathcal{C}$. We say that Σ *R-implies* φ in the presence of R_s ($\Sigma \models^R_{R_s} \varphi$) if every relation r over R that satisfies Σ and R_s also satisfies φ. If Σ does not R-imply φ in the presence of R_s we may also write $\Sigma \not\models^R_{R_s} \varphi$.

For a set Σ of data dependencies in $\mathcal{C}$ over a relation schema R and an NFS R_s over R, let $\Sigma^*_{(R,R_s)} = \{\varphi \in \mathcal{C} \mid \Sigma \models^R_{R_s} \varphi\}$ be its *semantic closure*. In order to determine the logical consequences of a set of FDs and MVDs with respect to R-implication one can utilise a syntactic approach by applying inference rules, e.g. those in Table 1. These inference rules have the form

$$\frac{\text{premise}}{\text{conclusion}}\,\text{condition},$$

and inference rules without any premises are called axioms. An inference rule is called sound for the R-implication of dependencies in the presence of an NFS, if whenever the set of dependencies in the premise of the rule and the NFS are satisfied by some relation over R and the dependencies and NFS satisfy the conditions of the rule, then the relation also satisfies the dependency in the conclusion of the rule. For a finite set $\Sigma \cup \{\varphi\}$ of dependencies and a set $\mathfrak{R}$ of inference rules let $\Sigma \vdash_{\mathfrak{R}} \varphi$ denote the *inference* of φ from Σ by $\mathfrak{R}$. That is, there is some sequence $\gamma = [\sigma_1, \ldots, \sigma_n]$ of dependencies such that $\sigma_n = \varphi$ and for every σ_i is an element of Σ or results from an application of an inference rule in $\mathfrak{R}$ to some dependencies in $\{\sigma_1, \ldots, \sigma_{i-1}\}$. For a finite set Σ of dependencies in $\mathcal{C}$, let $\Sigma^+_{\mathfrak{R}} = \{\varphi \mid \Sigma \vdash_{\mathfrak{R}} \varphi\}$ be its *syntactic closure* under inferences by $\mathfrak{R}$. A set $\mathfrak{R}$ of inference rules is said to be *sound* (*complete*) for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS if for every relation schema R, for every NFS R_s over R and for every set Σ of dependencies in $\mathcal{C}$ over R we have $\Sigma^+_{\mathfrak{R}} \subseteq \Sigma^*_{(R,R_s)}$ ($\Sigma^*_{(R,R_s)} \subseteq \Sigma^+_{\mathfrak{R}}$). The (finite) set $\mathfrak{R}$ is said to be a (finite) *axiomatization* for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS if $\mathfrak{R}$ is both sound and complete for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS.

4 Appropriate Reasoning

The goal of this section is to establish an axiomatization for the R-implication of FDs and MVDs in the presence of an NFS that enjoys the features described in the introduction. For this purpose we assume that sets $\mathfrak{R}$ of inference rules do not contain rules that are dependent on the underlying relation schema R with the exception of the R-complementation rule $\mathcal{C}_{\mathrm{M}}^{R}$. First we extend the notion of an *appropriate inference system* [9] to the presence of an arbitrary NFS.

Definition 1. *Let $\mathfrak{R}$ denote a set of inference rules that is complete for the R-implication of FDs and MVDs in the presence of an NFS.*

$\mathfrak{R}$ *is said to be* complementary *for the R-implication of FDs and MVDs if for every relation schema R, for every NFS R_s over R, for every set Σ of FDs and MVDs over R, and for every MVD φ over R such that φ is R-implied by Σ in the presence of R_s there is an inference of φ from Σ by $\mathfrak{R}$ in which the R-complementation rule $\mathcal{C}_M^R$ is applied at most once and if it is applied, then it is applied only in the very last step of the inference.*

$\mathfrak{R}$ *is said to be* adequate *for the R-implication of FDs and MVDs if for every relation schema R, for every NFS R_s over R, for every set Σ of FDs and MVDs over R, and for every FD φ over R such that φ is R-implied by Σ in the presence of R_s there is an inference of φ from Σ by $\mathfrak{R}$ in which the R-complementation rule $\mathcal{C}_M^R$ is not applied at all.*

$\mathfrak{R}$ *is said to be* appropriate *for the R-implication of FDs and MVDs in the presence of an NFS if $\mathfrak{R}$ is complementary and adequate.* □

The next result illustrates that the properties of complementarity and adequacy cannot be taken for granted.

Theorem 1. $\mathfrak{D}$ *is neither complementary nor adequate for the R-implication of FDs and MVDs in the presence of an NFS.* □

An immediate question is whether there exist any axiomatizations that are complementary and/or adequate. Before we can give an affirmative answer to this question, we introduce additional inference rules that we will require to identify such axiomatizations.

Lemma 1. *The additive null transitivity rule $\mathcal{T}_M^*$, null subset rule $\mathcal{S}_M$ and mixed null subset rule $\mathcal{S}_{FM}$ are sound for the R-implication of FDs and MVDs in the presence of an NFS.* □

We are now ready to present our first main result. For this purpose let

$$\mathfrak{U} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{U}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}^*, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{S}_{\mathrm{FM}}\}$$

and let $\mathfrak{F} = \mathfrak{U} \cup \{\mathcal{C}_{\mathrm{M}}^{R}\}$. The completeness of $\mathfrak{D}$ implies the completeness of $\mathfrak{F}$, and Lemma 1 shows that $\mathfrak{F}$ is an axiomatization for the R-implication of FDs and MVDs in the presence of an NFS. We will now show that $\mathfrak{F}$ is appropriate. The proof is constructive in the sense that it can be utilized to transform any inference that does not enjoy the features into an inference that does.

Theorem 2. *Let Σ be a set of FDs and MVDs over relation schema R, and $R_s \subseteq R$. For every inference γ from Σ by the system $\mathfrak{D}$ there is an inference ξ from Σ by the system $\mathfrak{F}$ with the following properties:*

Table 1. Inference rules for FDs and MVDs in the presence of an NFS R_s

$$\frac{}{XY \rightarrow Y}$$
(reflexivity, $\mathcal{R}_{\mathrm{F}}$)

$$\frac{X \rightarrow YZ}{X \rightarrow Y}$$
(decomposition, $\mathcal{D}_{\mathrm{F}}$)

$$\frac{X \rightarrow Y; X \rightarrow Z}{X \rightarrow YZ}$$
(FD union, $\mathcal{U}_{\mathrm{F}}$)

$$\frac{X \rightarrow Y; Y \rightarrow Z}{X \rightarrow Z}\, Y \subseteq XR_s$$
(null transitivity, $\mathcal{T}_{\mathrm{F}}$)

$$\frac{X \twoheadrightarrow Y; X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$
(MVD union, $\mathcal{U}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow W; Y \twoheadrightarrow Z}{X \twoheadrightarrow ZW}\, Y \subseteq X(W \cap R_s)$$
(additive null transitivity, $\mathcal{T}^*_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow W; Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - W}\, Y \subseteq X(W \cap R_s)$$
(null pseudo-transitivity, $\mathcal{T}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow W; Y \twoheadrightarrow Z}{X \twoheadrightarrow Z \cap W}\, Y \subseteq XR_s; (Y - X) \cap W = \emptyset$$
(null subset, $\mathcal{S}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$
(R-complementation, $\mathcal{C}^R_{\mathrm{M}}$)

$$\frac{X \rightarrow Y}{X \twoheadrightarrow Y}$$
(implication, $\mathcal{I}_{\mathrm{FM}}$)

$$\frac{X \twoheadrightarrow W; Y \rightarrow Z}{X \rightarrow Z \cap W}\, Y \subseteq XR_s; (Y - X) \cap W = \emptyset$$
(null mixed subset, $\mathcal{S}_{\mathrm{FM}}$)

$$\frac{X \twoheadrightarrow W; Y \rightarrow Z}{X \rightarrow Z - W}\, Y \subseteq X(W \cap R_s)$$
(null mixed pseudo-transitivity, $\mathcal{T}_{\mathrm{FM}}$)

1. *if γ infers an MVD, then*
   - *γ and ξ infer the same MVD,*
   - *in ξ the R-complementation rule is applied at most once, and*
   - *if the R-complementation rule is applied in ξ, then it is applied as the last rule.*
2. *if γ infers an FD, then*
   - *γ and ξ infer the same FD, and*
   - *in ξ the R-complementation rule is not applied at all.* □

As an example of Theorem 2 we illustrate how the inappropriate inferences by the system $\mathfrak{D}$ from Example 2 can be replaced by appropriate inferences by the system $\mathfrak{F}$.

Example 4. Let $R = ASLC$, $R_s = SL$, $\Sigma = \{A \rightarrow S; AL \rightarrow C; S \twoheadrightarrow AC\}$ as in Example 2. First we show an inference of the MVD $A \twoheadrightarrow L$ from Σ and R_s that utilizes the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ only in the last step.

$$\cfrac{\mathcal{T}^*_{\mathrm{M}}: \cfrac{\mathcal{I}_{\mathrm{FM}}: \cfrac{A \to S}{A \twoheadrightarrow S} \quad S \twoheadrightarrow AC}{A \twoheadrightarrow ACS}}{\mathcal{C}^R_{\mathrm{M}}: \quad A \twoheadrightarrow L}$$

Next we show an inference of the FD $A \to C$ from Σ and R_s that does not require any application of the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$.

$$\cfrac{\mathcal{T}^*_{\mathrm{M}}: \cfrac{\mathcal{I}_{\mathrm{FM}}: \cfrac{A \to S}{A \twoheadrightarrow S} \quad S \twoheadrightarrow AC}{A \twoheadrightarrow ACS} \qquad AL \to C}{\mathcal{S}_{\mathrm{FM}}: \quad A \to C}$$

For the application of the null mixed subset rule $\mathcal{S}_{\mathrm{FM}}$ note that $AL \subseteq ALS$ and $(AL - A) \cap ACS = \emptyset$ hold. In particular, the example showcases applications of the additive null transitivity rule $\mathcal{T}^*_{\mathrm{M}}$ and the null mixed subset rule $\mathcal{S}_{\mathrm{FM}}$. □

Corollary 1. *$\mathfrak{F}$ is an appropriate finite axiomatization for the R-implication of FDs and MVDs in the presence of an NFS.* □

Among others Theorem 2 shows that $\mathfrak{U}$ is nearly complete for the R-implication of FDs and MVDs in the presence of an NFS. Indeed, $\mathfrak{U}$ enables us to infer every R-implied FD. Moreover, for every R-implied MVD $X \twoheadrightarrow Y$ the system $\mathfrak{U}$ enables us to infer $X \twoheadrightarrow Y$ itself or $X \twoheadrightarrow R - Y$.

Corollary 2. *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and MVDs over the relation schema R. Then*

- *If φ denotes an FD, then: $\varphi \in \Sigma^+_{\mathfrak{F}}$ if and only if $\varphi \in \Sigma^+_{\mathfrak{U}}$.*
- *If φ denotes the MVD $X \twoheadrightarrow Y$, then: $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{F}}$ if and only if $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{U}}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{U}}$.* □

Another interpretation of Corollary 2 is the following: if $\mathfrak{U}$ is utilized to infer FDs, then the underlying universe does not need to be fixed at all; and if $\mathfrak{U}$ is utilized to infer MVDs, then the fixing of a universe can be deferred until the very last step of the inference.

5 Undetermined Universes

The system $\mathfrak{U}$ is almost complete for the R-implication of FDs and MVDs in the presence of an NFS. We show now that if we do not fix a relation schema R, then $\mathfrak{U}$ is actually complete for the corresponding notion of implication.

FDs, MVDs and NFSs are syntactical expressions as before, but their attribute sets are finite subsets of our countably infinite set $\mathfrak{A}$. Let $Dom(r)$ denote the domain of a relation r, i.e., the set of attributes over which the relation is defined. For an FD or MVD σ let $lhs(\sigma)$ and $rhs(\sigma)$ denote the attribute sets on the left-hand side and right-hand side, respectively. That is, $lhs(\sigma) = X$ and $rhs(\sigma) = Y$

if σ denotes the MVD $X \twoheadrightarrow Y$ or the FD $X \rightarrow Y$. Let $Attr(\sigma)$ denote the set of attributes that occur in σ, i.e., $Attr(\sigma) = lhs(\sigma) \cup rhs(\sigma)$. A relation r is said to satisfy the FD $X \rightarrow Y$ if $XY \subseteq Dom(r)$ and for all tuples $t_1, t_2 \in r$ the following holds: if $t_1[X] = t_2[X]$ and t_1 is X-total, then $t_1[Y] = t_2[Y]$. A relation r is said to satisfy the MVD $X \twoheadrightarrow Y$ if $Attr(\sigma) \subseteq Dom(r)$ and for all tuples $t_1, t_2 \in r$ the following holds: if $t_1[X] = t_2[X]$ and t_1 is X-total, then there is some $t \in r$ such that $t[XY] = t_1[XY]$ and $t[X(Dom(r) - Y)] = t_2[X(Dom(r) - Y)]$. Finally, a relation r satisfies the NFS R_s if $R_s \subseteq Dom(r)$ and r is R_s-total.

Definition 2. *Let $\Sigma \cup \{\varphi\}$ be a set of FDs and MVDs and R_s an NFS. We say that Σ* implies *φ in the presence of R_s if and only if every relation r satisfies the following condition: if $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq Dom(r)$ and r satisfies all $\sigma \in \Sigma$ and R_s, then r also satisfies φ.* □

The notions of *soundness* and *completeness* are simply adapted to the context of undetermined universes by dropping the reference to the underlying relation schema R from the corresponding notions in the context of fixed universes.

Let $\Sigma \cup \{\varphi\}$ be a set of FDs and MVDs, R_s an NFS, and let R be some relation schema such that $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq R$ holds. Based on the definition of an MVD and FD, respectively, the following hold:

1. If φ denotes an MVD, then Σ R-implies φ in the presence of R_s whenever Σ implies φ in the presence of R_s, but not necessarily vice versa.
2. If φ denotes an FD, then Σ R-implies φ in the presence of R_s if and only if Σ implies φ in the presence of R_s.

Next we illustrate that R-implication of an MVD does not necessarily entail the implication of the MVD.

Example 5. The MVD $S \twoheadrightarrow A, C$ SUPPLIES-implies the MVD $S \twoheadrightarrow L$ in the presence of $R_s = \emptyset$, but $S \twoheadrightarrow A, C$ does not imply $S \twoheadrightarrow L$ in the presence of R_s:

Supplier	Article	Cost	Location	Quantity
Taratua&Co	Kea	`ni`	Gisborne	2
Taratua&Co	Kea	`ni`	Wellington	3

.

□

We are now able to state our second main result of this paper.

Theorem 3. *The set $\mathfrak{U}$ is a finite axiomatization for the implication of FDs and MVDs in the presence of an NFS over undetermined universes.* □

6 Conclusion

We have established two finite axiomatizations of functional and multivalued dependencies over attribute sets in which arbitrarily many attributes can be declared NOT NULL. The axiomatizations capture the notion of semantic implication over fixed and undetermined universes, respectively. Together, they provide strong formal evidence for the intuition that the complementation rule is a mere means of database normalization. The results generalize several previous findings on the subject.

Acknowledgement

This research is supported by the Marsden fund council from Government funding, administered by the Royal Society of New Zealand. The second author is supported by a research grant of the Alfried Krupp von Bohlen and Halbach foundation, administered by the German Scholars organization.

References

1. Abiteboul, S., Hull, R., Vianu, V.: Foundations of Databases. Addison-Wesley, Reading (1995)
2. Arenas, M., Libkin, L.: A normal form for XML documents. ACM Trans. Database Syst. 29(1), 195–232 (2004)
3. Armstrong, W.W.: Dependency structures of database relationships. Information Processing 74, 580–583 (1974)
4. Atzeni, P., Morfuni, N.: Functional dependencies and constraints on null values in database relations. Information and Control 70(1), 1–31 (1986)
5. Beeri, C., Bernstein, P.: Computational problems related to the design of normal form relational schemas. ACM Trans. Database Syst. 4(1), 30–59 (1979)
6. Beeri, C., Fagin, R., Howard, J.H.: A complete axiomatization for functional and multivalued dependencies in database relations. In: SIGMOD Conference, pp. 47–61. ACM, New York (1977)
7. Bernstein, P.: Synthesizing third normal form relations from functional dependencies. ACM Trans. Database Syst. 1(4), 277–298 (1976)
8. Biskup, J.: Inferences of multivalued dependencies in fixed and undetermined universes. Theor. Comput. Sci. 10(1), 93–106 (1980)
9. Biskup, J., Link, S.: Appropriate reasoning about data dependencies in fixed and undetermined universes. In: FoIKS Conference, pp. 58–77 (2006)
10. Codd, E.F.: A relational model of data for large shared data banks. ACM Commun. 13(6), 377–387 (1970)
11. Codd, E.F.: Extending the database relational model to capture more meaning. ACM Trans. Database Syst. 4(4), 397–434 (1979)
12. Delobel, C., Adiba, M.: Relational database systems. North Holland, Amsterdam (1985)
13. Fagin, R.: Multivalued dependencies and a new normal form for relational databases. ACM Trans. Database Syst. 2(3), 262–278 (1977)
14. Gottlob, G., Zicari, R.: Closed world databases opened through null values. In: VLDB Conference, pp. 50–61. IEEE Computer Society, Los Alamitos (1988)
15. Grant, J.: Null values in a relational data base. Inf. Process. Lett. 6(5), 156–157 (1977)
16. Hartmann, S., Kirchberg, M., Link, S.: A subgraph-based approach towards functional dependencies for XML. In: SCI Conference, pp. 200–205 (2003)
17. Hartmann, S., Koehler, H., Link, S., Trinh, T., Wang, J.: On the notion of an XML key. In: Schewe, K.-D., Thalheim, B. (eds.) SDKB 2008. LNCS, vol. 4925, pp. 114–123. Springer, Heidelberg (2008)
18. Hartmann, S., Link, S.: Characterising nested database dependencies by fragments of propositional logic. Ann. Pure Appl. Logic 152(1-3), 84–106 (2008)
19. Hartmann, S., Link, S.: Efficient reasoning about a robust XML key fragment. ACM Trans. Database Syst. 34(2) (2009)

20. Hartmann, S., Link, S.: Numerical constraints on XML data. Inf. Comput. 208(5), 521–544 (2010)
21. Hartmann, S., Link, S.: When data dependencies over SQL tables meet the Logics of Paradox and S-3. In: PODS Conference (2010)
22. Hartmann, S., Link, S., Schewe, K.-D.: Weak functional dependencies in higher-order data models. In: Seipel, D., Turull-Torres, J.M.a. (eds.) FoIKS 2004. LNCS, vol. 2942, pp. 134–154. Springer, Heidelberg (2004)
23. Imielinski, T., Lipski Jr, W.: Incomplete information in relational databases. J. ACM 31(4), 761–791 (1984)
24. Langeveldt, W., Link, S.: Empirical evidence for the usefulness of Armstrong relations in the acquisition of meaningful functional dependencies. Inf. Syst. 35(3), 352–374 (2010)
25. Levene, M., Loizou, G.: Axiomatisation of functional dependencies in incomplete relations. Theor. Comput. Sci. 206(1-2), 283–300 (1998)
26. Levene, M., Loizou, G.: Database design for incomplete relations. ACM Trans. Database Syst. 24(1), 80–125 (1999)
27. Lien, E.: On the equivalence of database models. J. ACM 29(2), 333–362 (1982)
28. Link, S.: Consistency enforcement in databases. In: Bertossi, L., Katona, G.O.H., Schewe, K.-D., Thalheim, B. (eds.) Semantics in Databases 2001. LNCS, vol. 2582, pp. 122–143. Springer, Heidelberg (2003)
29. Link, S.: Charting the completeness frontier of inference systems for multivalued dependencies. Acta Inf. 45(7-8), 565–591 (2008)
30. Link, S.: On the implication of multivalued dependencies in partial database relations. Int. J. Found. Comput. Sci. 19(3), 691–715 (2008)
31. Mikinouchi, A.: A consideration on normal form of not-necessarily-normalised relation in the relational data model. In: VLDB Conference, pp. 447–453 (1977)
32. Schaerf, M., Cadoli, M.: Tractable reasoning via approximation. Artif. Intell. 74, 249–310 (1995)
33. Toman, D., Weddell, G.: On keys and functional dependencies as first-class citizens in description logics. J. Autom. Reasoning 40(2-3), 117–132 (2008)
34. Vincent, M., Liu, J., Liu, C.: Strong functional dependencies and their application to normal forms in XML. ACM Trans. Database Syst. 29(3), 445–462 (2004)
35. Weddell, G.: Reasoning about functional dependencies generalized for semantic data models. ACM Trans. Database Syst. 17(1), 32–64 (1992)
36. Wijsen, J.: Temporal FDs on complex objects. ACM Trans. Database Syst. 24(1), 127–176 (1999)
37. Wu, M.: The practical need for fourth normal form. In: ACM SIGCSE Conference, pp. 19–23 (1992)
38. Zaniolo, C.: Database relations with null values. J. Comput. Syst. Sci. 28(1), 142–166 (1984)