Acta Informatica (2008) 45:565–591
DOI 10.1007/s00236-008-0080-5

ORIGINAL ARTICLE

Charting the completeness frontier of inference systems for multivalued dependencies

Sebastian Link

Received: 4 April 2007 / Revised: 18 June 2008 / Published online: 23 October 2008

Abstract The implication of multivalued dependencies in relational databases has originally been defined in the context of some fixed finite universe. While axiomatisability and implication problems have been intensely studied with respect to this notion almost no research has been devoted towards the alternative notion of implication in which the underlying universe of attributes is left undetermined. Based on a set of common inference rules we establish all axiomatisations in undetermined universes, and all axiomatisations in fixed universes that indicate the role of the complementation rule as a means of database normalisation. This characterises the expressiveness of several incomplete sets of inference rules. We also establish relationships between axiomatisations in fixed and undetermined universes, and study the time complexity of the implication problem in undetermined universes. The results of this paper establish a foundation for reasoning about multivalued dependencies without the assumption of a fixed underlying universe.

1 Introduction

Relational databases still form the core of most database management systems, even after more than three decades following their introduction in [12]. The relational model organises data into a collection of relations. These structures permit the storage of inconsistent data, inconsistent in a semantic sense. Since this is not acceptable additional assertions, called dependencies, are formulated that every database is compelled to obey. There are many different classes of dependencies which can be utilised for improving the representation of the target database [17,41,44].

This research is supported by the Marsden fund council from Government funding, administered by the Royal Society of New Zealand.
A preliminary version of this paper was presented at the 4th International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2006), Budapest, Hungary.

S. Link (✉)
School of Information Management, Victoria University of Wellington, Wellington, New Zealand
e-mail: sebastian.link@vuw.ac.nz

Previous work. Multivalued dependencies (MVDs) [14,16,49] are an important class of dependencies. Informally, a relation r over the universe R of attributes satisfies the MVD $X \twoheadrightarrow Y$ whenever the value on X determines the set of values on Y independently from the set of values on $R - Y$. This suggests that the universe R is overloaded in the sense that it carries two independent facts XY and $X(R - Y)$. Indeed, the relation r exhibits the MVD $X \twoheadrightarrow Y$ precisely when r is decomposable into its projections $r[XY]$ and $r[X(R - Y)]$ without loss of information, i.e., when r is equal to the natural join $r[XY] \bowtie r[X(R - Y)]$ [16], cf. Example 1.1. Recently, extensions of multivalued dependencies have been found very useful for various design problems in advanced data models such as the nested relational data model [18], fuzzy databases [39], temporal databases [27], the Entity-Relationship model [42], data models that support nested lists [22,24,25] and XML [38,45,46].

Example 1.1 Consider the relation schema Dvd = {Title, Actor, Feature}. For each title, the set of actors is independent of the set of features that are available on the DVD. We therefore specify the MVD

$$\text{Title} \twoheadrightarrow \text{Actor}$$

on Dvd. The following table represents a typical relation r over Dvd

Title	Actor	Feature
Sleepy Hollow	Johnny Depp	Trailer
Sleepy Hollow	Johnny Depp	Director's commentary
Sleepy Hollow	Johnny Depp	Photo Gallery
Sleepy Hollow	Christina Ricci	Trailer
Sleepy Hollow	Christina Ricci	Director's commentary
Sleepy Hollow	Christina Ricci	Photo Gallery

that satisfies *Title* $\twoheadrightarrow$ *Actor*. Hence, r can be split into $r_1 = r$[Title, Actor]

Title	Actor
Sleepy Hollow	Johnny Depp
Sleepy Hollow	Christina Ricci

and $r_2 = r$[Title, Feature]

Title	Feature
Sleepy Hollow	Trailer
Sleepy Hollow	Director's commentary
Sleepy Hollow	Photo Gallery

and the original relation r is indeed equal to the natural join $r_1 \bowtie r_2$. □

The characteristic of decomposing a relation without loss of information is fundamental to relational database design, in particular 4NF [16]. Consequently, a lot of research has been devoted to studying the implication problem of multivalued dependencies. The implication problem is to decide whether an arbitrary MVD φ is implicitly specified (or logically implied) by an arbitrary set Σ of explicitly specified MVDs, i.e., whether φ is satisfied by every database that satisfies all the dependencies in Σ. The implication problem is crucial, e.g., for data modelling, database design, query optimisation and consistent query answering. The existence of a sound and complete set of inference rules (i.e. an axiomatisation) for the implication of data dependencies can form the basis of an enumeration algorithm that lists all logical consequences. Such an enumeration is very desirable in practice, as all of the implicit knowledge is mechanically derived from the explicit knowledge. Alternatively, an axiomatisation may also enable one to develop an algorithm which decides the implication of dependencies efficiently. This complements the enumeration algorithm by a further reasoning capability that can make efficient, but only partial decisions whether some dependency is implicitly specified or not. In contrast, the enumeration algorithm lists all of the implicitly specified dependencies without the additional burden of guessing any implied dependency.

The classical notion of an MVD [16] is dependent on the underlying universe R. This dependence is reflected syntactically by the R-complementation rule which enables us to conclude that every relation that satisfies the MVD $X \twoheadrightarrow Y$ will also satisfy the MVD $X \twoheadrightarrow R - Y$. In Example 1.1 for instance, the MVD $\mathit{Title} \twoheadrightarrow \mathit{Feature}$ can be derived by a single application of the DVD-complementation rule from the MVD $\mathit{Title} \twoheadrightarrow \mathit{Title}, \mathit{Actor}$. In all axiomatisations of MVDs, the R-complementation rule (or a slight variation of it) is special in the sense that it is the only inference rule in that axiomatisation which is dependent on R. This dependence on the underlying universe imposes an additional constraint on solving the implication problem: the underlying universe has to be fixed before any attempt can be made to infer any implicit knowledge. This restriction distinguishes MVDs from other dependencies, e.g. functional dependencies whose satisfaction does not depend on the underlying universe. For instance, the well-known synthesis approach towards achieving the third normal form is only possible because this restriction does not hold [7]. In fact, one of the open problems in relational database design is a generalisation of the synthesis approach to multivalued dependencies. This problem, however, appears to be difficult to address when the underlying set of attributes is assumed to be fixed.

These interesting issues have led to an alternative notion of semantic implication in which the underlying universe is left undetermined [11]. In the same paper Biskup establishes an axiomatisation $\mathfrak{S}_0$ for MVDs in undetermined universes. If $\mathfrak{S}_0^{\mathcal{C}}$ results from adding the R-complementation rule to $\mathfrak{S}_0$, then $\mathfrak{S}_0^{\mathcal{C}}$ becomes an axiomatisation for MVDs in fixed universes. In fact, every inference of an MVD by $\mathfrak{S}_0^{\mathcal{C}}$ can be turned into an inference of the same MVD in which the R-complementation rule is applied at most once, and if it is applied, then in the last step of the inference ($\mathfrak{S}_0^{\mathcal{C}}$ is said to be R-complementary). This indicates that the R-complementation rule simply reflects a part of the decomposition process, and does not necessarily infer semantically meaningful consequences.

Example 1.2 Consider the universe DVD of Example 1.1 together with the MVD $\mathit{Title} \twoheadrightarrow \mathit{Title}, \mathit{Actor}$. In the fixed universe DVD the MVD $\mathit{Title} \twoheadrightarrow \mathit{Feature}$ is implied by the MVD $\mathit{Title} \twoheadrightarrow \mathit{Title}, \mathit{Actor}$. However, in the universe {Title,Actor,Feature,Language} the MVD $\mathit{Title} \twoheadrightarrow \mathit{Feature}$ is not implied by the MVD $\mathit{Title} \twoheadrightarrow \mathit{Title}, \mathit{Actor}$ as the following example shows:

Title	Actor	Feature	Language
Miyamoto Musashi	T. Mifune	Trailer	English
Miyamoto Musashi	T. Mifune	Subtitles	Japanese

Consequently, the MVD $\mathit{Title} \twoheadrightarrow \mathit{Feature}$ is a meaningful consequence in the universe DVD but not a consequence in any other universes. □

Interestingly, research has not been continued in this direction but focused on the original notion of implication in fixed universes. Since research on MVDs seems to experience a recent revival in the context of other data models [18,22,24,25,33,38,42,45,46] it seems desirable to further extend the knowledge on relational MVDs. An advancement of such knowledge may simplify the quest of finding suitable and comprehensible extensions of MVDs to currently popular data models. Furthermore, the results of this paper show that fixing a universe of attributes is not an assumption that is necessary for MVDs, and e.g., for data modelling, database design or query optimisation based on MVDs. Our findings show that even for the original notion of implication in fixed universes this assumption becomes almost negligible.

Contributions. In this paper we will further study the notions of MVD implication in both fixed and undetermined universes [11,16]. In summary, we will establish axiomatisations for both notions of implication, strong connections between the two notions, and an algorithm that efficiently solves the implication problem for MVDs in undetermined universes. More specifically, the findings of this article can be summarised as follows:

- Based on a set $\mathfrak{S}_U$ of eight common sound inference rules we will identify all minimal complete subsets of $\mathfrak{S}_U$ with respect to implication in undetermined universes. Here, minimality refers to the fact that none of the rules can be omitted without losing completeness. Essentially, it turns out that apart from the set $\mathfrak{S}_0$ there are two other subsets of $\mathfrak{S}_U$ that are also minimal. The same problem was studied by Mendelzon [34] with respect to implication in fixed universes. On one hand this contribution provides a complete picture of what inference rules are necessary to gain all the implicitly specified MVDs. On the other hand it provides data administrators with different choices to enumerate implicit knowledge.
- It is formally shown that implication in undetermined universes is captured by a sound set $\mathfrak{S}$ of inference rules precisely if for all relation schemata R the set $\mathfrak{S}$ extended by the R-complementation rule is both complete and complementary for the implication of MVDs on R. In particular, we identify all axiomatisations of MVDs in fixed universes that are minimal with respect to both completeness and complementarity (i.e. no proper subset is both complete and complementary). This contribution establishes a strong connection between the notions in fixed and undetermined universes.
- These findings are not just of theoretical interest. In practice one does not necessarily want to generate all consequences of a given set of MVDs but only some of them. Such a task can be accomplished by using incomplete sets of inference rules. The results above show that the axiomatisations in undetermined universes are nearly complete for fixed universes, and can be used to fix the universe in the very last step of any inference.
- We also investigate the notion of minimality for complete inference systems in fixed and undetermined universes. In particular, we study weaker versions of the reflexivity axiom which, in combination with other inference rules, still suffice to gain completeness. Moreover, we introduce the novel notion of strong minimality. Intuitively, a set of inference rules is strongly minimal if the omission of any of its inference rules results in the inability to infer all *trivial* MVDs. Therefore, we define and syntactically characterise trivial MVDs in undetermined universes. This reveals further differences to the notion of MVDs in fixed universes. The notion of strong minimality enables us to further classify previously established inference systems for MVDs in fixed universes. In sharp contrast, none of our minimal MVD axiomatisations in undetermined universes is strongly minimal.
- Subsequently, we study the time-complexity of the implication problem in undetermined universes. Herein, the classical notion of a dependency basis of an attribute set with respect to a set of MVDs can be rephrased [4]. It turns out that the dependency basis of X with respect to Σ in a fixed universe deviates from the dependency basis of X with respect to Σ in undetermined universes by at most one set. This set, however, can be described by the notion of the Σ-scope of an attribute set X, which is the union of all those attribute sets

Y such that $X \twoheadrightarrow Y$ is implied by Σ. The Σ-scope itself can be computed in time linear in the total number of attributes that occur in Σ. The computation is similar to computing the closure of an attribute set with respect to a set of functional dependencies [5]. As a first consequence the currently best known upper bound for deciding MVD implication in fixed universes [19] also applies to undetermined universes. A further consequence is the fact that the existence of a linear-time algorithm for deciding MVD implication in fixed universes implies the same for undetermined universes, and vice versa.

Organisation. The paper is structured as follows. Section 2 summarises notions from the relational model of data. Section 3 identifies all minimal axiomatisations in $\mathfrak{S}_U$ for the implication of MVDs in undetermined universes. The notion of minimality is explored in Sect. 4. A rather general result is proven in Sect. 5 which shows an equivalence between axiomatisations of MVDs in fixed universes and those in undetermined ones. Finally, the implication problem of MVDs in undetermined universes is studied in Sect. 6. The paper concludes with discussing related work and some open problems in Sect. 7.

2 Multivalued dependencies in relational databases

In this section we will fix notions and notations fundamental to the relational model of data. In particular, we will summarise the notion of MVD implication in both fixed and undetermined universes. This will provide the basis for the development of our results in the following sections.

Let $\mathfrak{A} = \{A_1, A_2, \dots\}$ be a (countably) infinite set of attributes. A *relation schema* is a finite set $R = \{A_1, \dots, A_n\}$ of distinct symbols, called *attributes*, which represent column names of a relation. Each attribute A_i of a relation schema is associated an infinite domain $\mathrm{dom}(A_i)$ which represents the set of possible values that can occur in the column named A_i. If X and Y denote sets of attributes, then we may write XY for $X \cup Y$. If $X = \{A_1, \dots, A_m\}$, then we may write $A_1, \dots, A_m$ for X. In particular, we may write simply A to represent the singleton $\{A\}$. A *tuple* over $R = \{A_1, \dots, A_n\}$ (R-tuple or simply tuple, if R is understood) is a function $t : R \to \bigcup_{i=1}^{n} \mathrm{dom}(A_i)$ with $t(A_i) \in \mathrm{dom}(A_i)$ for $i = 1, \dots, n$. For $X \subseteq R$ let $t[X]$ denote the restriction of the tuple t over R on X, and $\mathrm{dom}(X) = \prod_{A \in X} \mathrm{dom}(A)$ the Cartesian product of the domains of attributes in X. A *relation* r over R is a finite set of tuples over R. If a relation r is given without reference to the relation schema R over which it is defined, then we denote R also by $Attr(r)$, i.e., the set of attributes over which r is defined. Let $r[X] = \{t[X] \mid t \in r\}$ denote the *projection* of the relation r over R on $X \subseteq R$. For $X, Y \subseteq R$, $r_1 \subseteq \mathrm{dom}(X)$ and $r_2 \subseteq \mathrm{dom}(Y)$ let $r_1 \bowtie r_2 = \{t \in \mathrm{dom}(XY) \mid \exists t_1 \in r_1, t_2 \in r_2 \text{ with } t[X] = t_1[X] \text{ and } t[Y] = t_2[Y]\}$ denote the *natural join* of r_1 and r_2. Note that the 0-ary relation $\{()\}$ is the projection $r[\emptyset]$ of a non-empty relation r on $\emptyset$ as well as left and right identity of the natural join operator.

2.1 MVDs in fixed universes

Functional dependencies (FDs) between sets of attributes have always played a central role in the study of relational databases [5,7,8,12,13], and seem to be central for the study of database design in other data models as well [1,21,22,29,40,47,48]. The notion of a functional dependency is well-understood and the semantic interaction between these dependencies has been syntactically captured by Armstrong's well-known axioms [2,3]. A *functional dependency* (FD) [13] on the relation schema R is an expression $X \to Y$ where $X, Y \subseteq R$. A relation r over R *satisfies* the FD $X \to Y$, denoted by $\models_r X \to Y$, if and only if every pair of tuples in r that agrees on each of the attributes in X also agrees on the attributes in Y. That is, $\models_r X \rightarrow Y$ if and only if $t_1[Y] = t_2[Y]$ whenever $t_1[X] = t_2[X]$ holds for any $t_1, t_2 \in r$.

FDs are incapable of modelling many important properties that database users have in mind. Multivalued dependencies (MVDs) provide a more general notion and offer a response to the shortcomings of FDs. A *multivalued dependency* (MVD) [14,16,49] on R is an expression $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. A relation r over R *satisfies* the MVD $X \twoheadrightarrow Y$, denoted by $\models_r X \twoheadrightarrow Y$, if and only if for all $t_1, t_2 \in r$ with $t_1[X] = t_2[X]$ there is some $t \in r$ with $t[XY] = t_1[XY]$ and $t[X(R - Y)] = t_2[X(R - Y)]$. Informally, the relation r satisfies $X \twoheadrightarrow Y$ when the value on X determines the set of values on Y independently from the set of values on $R - Y$. This actually suggests that the relation schema R is overloaded in the sense that it carries two independent facts XY and $X(R - Y)$. More precisely, it is shown in [16] that MVDs "provide a necessary and sufficient condition for a relation to be decomposable into two of its projections without loss of information (in the sense that the original relation is guaranteed to be the join of the two projections)". This means that $\models_r X \twoheadrightarrow Y$ if and only if $r = r[XY] \bowtie r[X(R - Y)]$. This characteristic of MVDs is fundamental to relational database design and 4NF [16]. A lot of research has therefore been devoted to studying the behaviour of these dependencies.

For the design of a relational database schema dependencies are normally specified as semantic constraints on the relations which are intended to be instances of the schema. During the design process one usually needs to determine further dependencies which are logically implied by the given ones. In order to emphasise the dependence of implication from the underlying relation schema R we refer to R-implication.

Definition 2.1 Let R be a relation schema, $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ a set of MVDs and $X \twoheadrightarrow Y$ an MVD on R, i.e., $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R$. Then Σ R-implies $X \twoheadrightarrow Y$ if and only if each relation r over R that satisfies all MVDs in Σ also satisfies $X \twoheadrightarrow Y$. $\square$

In order to determine all logical consequences of a set of MVDs one can use the following set of inference rules for the R-implication of multivalued dependencies [35].

$$\begin{array}{ccc}
\dfrac{}{X \twoheadrightarrow A}\, A \in X & \dfrac{}{X \twoheadrightarrow Y}\, Y \subseteq X & \dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\, V \subseteq U \\
\text{(membership, } \mathcal{M}) & \text{(reflexivity, } \mathcal{R}) & \text{(augmentation, } \mathcal{A}) \\[2ex]
\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} & \dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y} & \dfrac{}{\emptyset \twoheadrightarrow R} \\
\text{(pseudo-transitivity, } \mathcal{T}) & (R\text{-complementation, } \mathcal{C}_R) & (R\text{-axiom, } \mathcal{C}.1) \\[2ex]
\dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ} & \dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} & \dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z} \\
\text{(union, } \mathcal{U}) & \text{(difference, } \mathcal{D}) & \text{(intersection, } \mathcal{I})
\end{array}$$

These inference rules have the form

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

and inference rules without a premise are called *axioms*. Notice that the membership axiom $\mathcal{M}$, reflexivity axiom $\mathcal{R}$ and augmentation rule $\mathcal{A}$ can also be rewritten into this form, e.g. by

$$\frac{}{XA \twoheadrightarrow A} \qquad \frac{}{XY \twoheadrightarrow Y} \qquad \frac{X \twoheadrightarrow Y}{XUV \twoheadrightarrow YV},$$

respectively. However, as the literature prefers the first notation we will follow this standard.

The set $\mathfrak{F} = \{\mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{C}_R, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$ of inference rules is both R-sound and R-complete for the R-implication of MVDs, for all relation schema R [6,10,35].

Let $\Sigma \cup \{\varphi\}$ be a set of MVDs on the relation schema R. Furthermore, we use $\mathfrak{S}$ to denote a set of inference rules. Unless stated otherwise, we consider only those sets of inference rules in which the R-complementation rule $\mathcal{C}_R$ is the only inference rule that is dependent on R. Let $\Sigma \vdash_{\mathfrak{S}} \varphi$ denote the inference of φ from Σ with respect to $\mathfrak{S}$. Let $\Sigma^+_{\mathfrak{S}} = \{\varphi \mid \Sigma \vdash_{\mathfrak{S}} \varphi\}$ denote the *syntactic hull* of Σ under inference using only rules from $\mathfrak{S}$. An inference rule is called R*-sound* if the set of dependencies in the premise of the rule R-implies the dependency in the conclusion. It is well-known that all the rules above are R-sound for all R [35]. The set $\mathfrak{S}$ is called R*-sound* for the R-implication of MVDs if and only if for every set Σ of MVDs on the relation schema R we have $\Sigma^+_{\mathfrak{S}} \subseteq \Sigma^*_R = \{\varphi \mid \Sigma\ R\text{-implies}\ \varphi\}$. The set $\mathfrak{S}$ is called R*-complete* for the R-implication of MVDs if and only if for every set Σ of MVDs on R we have $\Sigma^*_R \subseteq \Sigma^+_{\mathfrak{S}}$.

An interesting question is now whether all the rules of a certain set of inference rules are really necessary to capture the R-implication of MVDs for every relation schema R. More precisely, an inference rule $\mathfrak{R}$ is said to be *independent* of the set $\mathfrak{S}$ if and only if there is some relation schema R and some finite set $\Sigma \cup \{\varphi\}$ of MVDs on R such that $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathfrak{R}\}}$. In this case we say that Σ and φ are *witnesses* to the independence of $\mathfrak{R}$ from $\mathfrak{S}$. Let $\mathfrak{S}$ be a set of inference rules that is R-complete for the R-implication of MVDs for all relation schemata R. Then $\mathfrak{S}$ is said to be *minimal* for the R-implication of MVDs if and only if every inference rule $\mathfrak{R} \in \mathfrak{S}$ is independent of $\mathfrak{S} - \{\mathfrak{R}\}$. This means that no proper subset of $\mathfrak{S}$ is still R-complete for the R-implication of MVDs on all relation schemata R. It was shown by Mendelzon [34] that $\mathfrak{M} = \{\mathcal{R}, \mathcal{C}_R, \mathcal{T}\}$ forms such a minimal set for the R-implication of MVDs. The R-complementation rule $\mathcal{C}_R$ plays a special role as it is the only rule which depends on the underlying relation schema R. In the same paper, Mendelzon further motivates the study of the independence of inference rules and comments in more detail on the special role of the R-complementation rule.

2.2 MVDs in undetermined universes

Consider the classical example [16] in which the MVD *Employee* ↠ *Child* is specified, i.e., the set of children is completely determined by an employee, independently from the rest of the information in any schema. If the relation schema R consists of the attributes *Employee*, *Child* and *Salary*, then we may infer the MVD *Employee* ↠ *Salary* by means of the complementation rule. However, if the underlying relation schema R consists of the four attributes *Employee*, *Child*, *Salary* and *Year*, then the MVD *Employee* ↠ *Salary* is no longer R-implied. Note the fundamental difference of the MVDs

$$\textit{Employee} \twoheadrightarrow \textit{Child} \quad \text{and} \quad \textit{Employee} \twoheadrightarrow \textit{Salary}.$$

The first MVD has been specified to establish the relationship of employees and their children as a fact due to a set-valued correspondence. The second MVD does not necessarily correspond to any semantic information, but simply results from the context in which *Employee* and *Child* are considered. If the context changes, the MVD disappears.

If we argue this way, cf. Examples 1.2 and 2.1, then consequences which are dependent on the underlying relation schema are in fact no consequences. This implies, however, that the notion of R-implication is not suitable.

Biskup introduced the following notion of implication [11]. An MVD is a syntactic expression $X \twoheadrightarrow Y$ with $X, Y \subseteq \mathfrak{A}$. The MVD $X \twoheadrightarrow Y$ is satisfied by some relation r if and only if $X \cup Y \subseteq Attr(r)$ and $r = r[XY] \bowtie r[X \cup (Attr(r) - Y)]$.

Definition 2.2 The set $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ of MVDs *implies* the single MVD $X \twoheadrightarrow Y$ if and only if for each relation r with $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq Attr(r)$ the MVD $X \twoheadrightarrow Y$ is satisfied by r whenever r already satisfies all MVDs in Σ. ☐

In this definition, the underlying relation schema is left undetermined. The only requirement is that the MVDs must apply to the relations. If $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R$, then it follows immediately that $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ R-implies $X \twoheadrightarrow Y$ whenever Σ implies $X \twoheadrightarrow Y$. The converse, however, is false [11] as the following example demonstrates.

Example 2.1 For R ={*Employee*, *Child*, *Salary*} and Σ ={*Employee* $\twoheadrightarrow$ *Child*} we have that Σ R-implies *Employee* $\twoheadrightarrow$ *Salary*. However, Σ does not imply *Employee* $\twoheadrightarrow$ *Salary*. Consider for instance the following relation r with domain {*Employee*, *Child*, *Salary*, *Year*}.

Employee	*Child*	*Salary*	*Year*
Homer	Bart	4000	2004
Homer	Lisa	3500	2005
Homer	Bart	3500	2005
Homer	Lisa	4000	2004

The two relations r[*Employee, Child*] and r[*Employee, Salary, Year*]

Employee	*Child*
Homer	Bart
Homer	Lisa

Employee	*Salary*	*Year*
Homer	4000	2004
Homer	3500	2005

show that r satisfies the MVD *Employee* $\twoheadrightarrow$ *Child*. However, the two relations r[*Employee*, *Salary*] and r[*Employee, Child, Year*]

Employee	*Salary*
Homer	4000
Homer	3500

Employee	*Child*	*Year*
Homer	Bart	2004
Homer	Bart	2005
Homer	Lisa	2004
Homer	Lisa	2005

indicate that r does not satisfy *Employee* $\twoheadrightarrow$ *Salary*. Consequently, Σ does not imply *Employee* $\twoheadrightarrow$ *Salary*. ☐

The notions of *soundness, completeness, independence* and *minimality* 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.

While the singletons $\mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{U}, \mathcal{D}, \mathcal{I}$ are all sound, the R-complementation rule $\mathcal{C}_R$ and R-axiom $\mathcal{C}.1$ are R-sound, but not sound [11]. In fact, the main result of [11] shows that the following set $\mathfrak{S}_0$ of inference rules

$$\frac{}{\emptyset \twoheadrightarrow \emptyset} \qquad \frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\, V \subseteq U \qquad \frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$

(empty-set-axiom, $\mathcal{R}_\emptyset$) (augmentation, $\mathcal{A}$) (pseudo-transitivity, $\mathcal{T}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow YZ} \qquad \frac{X \twoheadrightarrow Y, W \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}\, Y \cap W = \emptyset$$

(additive transitivity, $\mathcal{T}^*$) (subset, $\mathcal{S}$)

is sound and complete for the implication of MVDs. The major proof argument shows that every inference of an MVD $X \twoheadrightarrow Y$ using the set $\mathfrak{S}_0^{\mathcal{C}} = \{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{S}, \mathcal{C}_R\}$ can be turned into an inference of $X \twoheadrightarrow Y$ which applies the R-complementation rule $\mathcal{C}_R$ at most once, and if it is applied, then it is applied in the last step (a set of inference rules with this property is said to be *R-complementary*). This shows that

$$X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}} \quad \text{iff} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0} \text{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}_0} \tag{2.1}$$

where $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \cup Y \cup \bigcup_{i=1}^k (X_i \cup Y_i) \subseteq R$.

3 Minimal axiomatisations in undetermined universes

In this section, all minimal complete subsets of

$$\mathfrak{S}_U = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$$

will be revealed. That is, a subset of $\mathfrak{S}_U$ is complete if and only if it is a superset of at least one of these minimal sets. On one hand this contribution provides a complete picture of what inference rules are necessary to gain all the implicitly specified MVDs in undetermined universes. On the other hand it provides data administrators with different choices to enumerate all implicit knowledge. The same problem was studied by Mendelzon [34] with respect to implication in fixed universes.

The set $\mathfrak{S}_1 = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*\}$ results from the $\mathfrak{S}_0$ by replacing the empty-set-axiom $\mathcal{R}_\emptyset$ by the reflexivity axiom $\mathcal{R}$. It follows immediately that $\mathfrak{S}_1$ is complete for the implication of MVDs.

3.1 The system $\mathfrak{S}_2$

We will start with some independence proofs. These proofs have been checked using simple GNU Pascal programs (which offer good set arithmetic) that compute the closure $\Sigma^+_{\mathfrak{S}}$ as the least fixpoint of a given set Σ of MVDs under the application of inference rules from a given set $\mathfrak{S}$ (neglecting trivial MVDs).

Lemma 3.1 *Let* $\mathfrak{S} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$. *The reflexivity axiom* $\mathcal{R}$ *is independent of* $\mathfrak{S}$.

Proof $\mathcal{R}$ is the only axiom. If $\Sigma = \emptyset$ and $\varphi = \emptyset \twoheadrightarrow \emptyset$, then $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{R}\}}$. □

The next result shows that the subset rule $\mathcal{S}$ is independent of

$$\{\mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{U}, \mathcal{D}, \mathcal{I}\}.$$

This strengthens the result of Biskup who has shown [11,Theorem 2] that the subset rule is independent of $\{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*\}$.

Lemma 3.2 *Let* $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$. *The subset rule* $\mathcal{S}$ *is independent of* $\mathfrak{S}$.

Proof Let $\Sigma = \{\emptyset \twoheadrightarrow AB, C \twoheadrightarrow BC\}$, and $\varphi = \emptyset \twoheadrightarrow B$. We will use tables to represent the closure $\Sigma^+_{\mathfrak{S}}$. An MVD $X \twoheadrightarrow Y$ belongs to $\Sigma^+_{\mathfrak{S}}$ if and only if the entry at row labelled X and column labelled Y is a cross ×. Note that we omit other trivial MVDs $X \twoheadrightarrow Y$ with $Y \subseteq X$ and where XY contains any attribute that does not occur in any of the MVDs in

Σ. The following table may be generated as follows. First we enter those MVDs that are in Σ. Then we apply the reflexivity axiom $\mathcal{R}$ as often as possible. Subsequently, we apply the difference rule $\mathcal{D}$ to enter $C \twoheadrightarrow B$. Next, the augmentation rule $\mathcal{A}$ is used several times. This gains the AB-column as well as the BC-column below. This is followed by several applications of the difference rule $\mathcal{D}$ again. Finally, the union rule $\mathcal{U}$ can be applied to obtain the table below. Further applications of any rules in $\mathfrak{S}$ do not result in any new MVDs. In particular, φ cannot be inferred from Σ since the only non-trivial MVD at our disposal is $\emptyset \twoheadrightarrow AB$ and the inferable MVDs with left-hand side AB are all trivial, i.e., applications of the pseudo-transitivity rule $\mathcal{T}$ always result in the inference of $\emptyset \twoheadrightarrow \emptyset$.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×				×			
A	×	×	×		×			
B	×	×	×		×			
C	×	×	×	×	×	×	×	×
AB	×	×	×		×			
AC	×	×	×	×	×	×	×	×
BC	×	×	×	×	×	×	×	×
ABC	×	×	×	×	×	×	×	×

Since $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{S}\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{S}$ from $\mathfrak{S}$. □

Lemmas 3.1 and 3.2 show that any complete subset of $\mathfrak{S}_U$ must include both reflexivity axiom $\mathcal{R}$ and subset rule $\mathcal{S}$. The following result has been proven in [22, Lemma 2].

Lemma 3.3 *The augmentation rule $\mathcal{A}$ is derivable from $\{\mathcal{R}, \mathcal{T}, \mathcal{U}\}$.*

Lemma 3.4 *The additive transitivity rule $\mathcal{T}^*$ is derivable from the pseudo-transitivity rule $\mathcal{T}$ and the union rule $\mathcal{U}$.*

Proof

$$\mathcal{U}: \dfrac{\mathcal{T}: \dfrac{X \twoheadrightarrow Y \quad Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \quad X \twoheadrightarrow Y}{X \twoheadrightarrow Y \cup Z}$$

This concludes the proof. □

Recall that $\mathfrak{S}_0 = \{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{S}\}$ is complete. Applying Lemmas 3.3 and 3.4 to $\mathfrak{S}_0$ results in the first new complete set.

Theorem 3.1 *The set $\mathfrak{S}_2 = \{\mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{U}\}$, consisting of reflexivity axiom, subset rule, pseudo-transitivity rule and union rule, is sound and complete for the implication of MVDs.*

Proof It suffices to show that the empty-set-axiom $\mathcal{R}_\emptyset$, augmentation rule $\mathcal{A}$ and additive transitivity rule $\mathcal{T}^*$ are derivable from $\mathfrak{S}_2$. $\mathcal{R}_\emptyset$ is a very weak form of the reflexivity axiom $\mathcal{R}$. Lemmas 3.3 and 3.4 show that $\mathcal{A}$ and $\mathcal{T}^*$ are derivable from $\mathfrak{S}_2$ as well. This concludes the proof. □

3.2 The system $\mathfrak{S}_3$

Lemma 3.5 *The pseudo-transitivity rule $\mathcal{T}$ can be derived from the additive pseudo-transitivity rule $\mathcal{T}^*$ and the difference rule $\mathcal{D}$.*

Proof Recall that $(Y \cup Z) - Y = Z - Y$.

$$\mathcal{D}:\ \dfrac{X \twoheadrightarrow Y \qquad \mathcal{T}^*:\ \dfrac{X \twoheadrightarrow Y \qquad Y \twoheadrightarrow Z}{X \twoheadrightarrow Y \cup Z}}{X \twoheadrightarrow \underbrace{(Y \cup Z) - Y}_{=Z-Y}}$$

This concludes the proof. □

Lemma 3.6 *The union rule $\mathcal{U}$ is derivable from the reflexivity axiom $\mathcal{R}$, the additive pseudo-transitivity rule $\mathcal{T}^*$ and the difference rule $\mathcal{D}$.*

Proof Note that $YZ = XYZ - (X - YZ)$.

$$\mathcal{D}:\ \dfrac{\mathcal{T}^*:\ \dfrac{\mathcal{T}^*:\ \dfrac{\mathcal{R}:\ \dfrac{}{X \twoheadrightarrow X} \quad X \twoheadrightarrow Y}{X \twoheadrightarrow XY} \qquad \mathcal{T}^*:\ \dfrac{\mathcal{R}:\ \dfrac{}{XY \twoheadrightarrow X} \quad X \twoheadrightarrow Z}{XY \twoheadrightarrow XZ}}{X \twoheadrightarrow XYZ} \qquad \mathcal{R}:\ \dfrac{}{X \twoheadrightarrow (X - (YZ))}}{X \twoheadrightarrow \underbrace{XYZ - (X - (YZ))}_{=YZ}}$$

This concludes the proof. □

The next theorem shows that one can achieve completeness without using the pseudo-transitivity rule nor the union rule.

Theorem 3.2 *The set $\mathfrak{S}_3 = \{\mathcal{R}, \mathcal{S}, \mathcal{T}^*, \mathcal{D}\}$, consisting of reflexivity axiom, subset rule, additive pseudo-transitivity rule and difference rule, is sound and complete for the implication of MVDs.*

Proof The completeness of $\mathfrak{S}_3$ follows from Theorem 3.1, Lemmas 3.5 and 3.6. □

3.3 All axiomatisations of MVDs in undetermined Universes

We will now show that $\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3$ are the only minimal complete subsets of $\mathfrak{S}_U$.

Lemma 3.7 *Let $\mathfrak{S} = \{\mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{T}^*, \mathcal{I}\}$. The union rule $\mathcal{U}$ is independent of $\mathfrak{S}$.*

Proof Let $\Sigma = \{AB \twoheadrightarrow BC, AB \twoheadrightarrow CD\}$ and $\varphi = AB \twoheadrightarrow BCD$. The closure $\Sigma^+_{\mathfrak{S}}$ may be generated as follows. First, we enter the MVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. Then, we apply the intersection rule $\mathcal{I}$ to enter $AB \twoheadrightarrow C$. Subsequently, we apply the additive transitivity rule $\mathcal{T}^*$ three times to gain the ABC-column below and to infer $AB \twoheadrightarrow ABCD$. Next, the pseudo-transitivity rule $\mathcal{T}$ is applied three times to gain the CD column and the C-column below. We then apply the additive transitivity rule $\mathcal{T}^*$ twice to gain the $ABCD$-column below. Then, we apply the pseudo-transitivity rule $\mathcal{T}$ to $AB \twoheadrightarrow ABC$ and $ABC \twoheadrightarrow ABCD$ to infer $AB \twoheadrightarrow D$, and again to $ABC \twoheadrightarrow ABC$ and

$ABC \twoheadrightarrow ABCD$ to infer $ABC \twoheadrightarrow D$. Another two applications of the additive transitivity rule $\mathcal{T}^*$ yield the ABD-column below. Further applications do not result in new MVDs.

	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	$ABCD$
A	×														
B		×													
C			×												
D				×											
AB	×	×	×	×	×			×		×	×	×			×
AC	×		×			×									
AD	×			×			×								
BC		×	×					×							
BD		×		×					×						
CD			×	×						×					
ABC	×	×	×	×	×	×		×		×	×	×			×
ABD	×	×	×	×	×		×		×	×	×	×			×
ACD	×		×	×		×	×			×			×		
BCD		×	×	×				×	×	×				×	

Since $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{U}\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{U}$ from $\mathfrak{S}$. □

For the next lemma note that $\mathfrak{S}_3 = \{\mathcal{R}, \mathcal{S}, \mathcal{T}^*, \mathcal{D}\}$ is complete, i.e., the pseudo-transitivity rule $\mathcal{T}$ is derivable from any of its supersets, in particular.

Lemma 3.8 *The pseudo-transitivity rule $\mathcal{T}$ is independent of both*

1. $\mathfrak{S} = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}^*, \mathcal{U}, \mathcal{I}\}$, *and*
2. $\mathfrak{S}' = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$.

Proof For the first statement let $\Sigma = \{A \twoheadrightarrow B, B \twoheadrightarrow C\}$ and $\varphi = A \twoheadrightarrow C$. The closure $\Sigma^+_{\mathfrak{S}}$ may be generated as follows. First, we enter the MVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. Then, we apply the union rule $\mathcal{U}$ to derive $A \twoheadrightarrow AB$ and $B \twoheadrightarrow BC$. We apply the additive transitivity rule $\mathcal{T}^*$ to infer $A \twoheadrightarrow BC$ from the MVDs in Σ. Subsequently, five applications of the augmentation rule $\mathcal{A}$ result in the B, C, AB, and BC-columns below. Finally, four applications of the union rule $\mathcal{U}$ result in the table below. Further applications do not result in new MVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×		×		×	×
B	×		×	×			×	
C	×			×				
AB	×	×	×	×	×	×	×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{T}\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{T}$ from $\mathfrak{S}$.

For the second statement let $\Sigma = \{A \twoheadrightarrow B, B \twoheadrightarrow C\}$ and $\varphi = A \twoheadrightarrow C$. The closure $\Sigma^+_{\mathfrak{S}'}$ may be generated as follows. First, we enter the MVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. Then, we apply the augmentation rule $\mathcal{A}$ twice to infer

$AB \twoheadrightarrow C$ and $AC \twoheadrightarrow B$. Finally, the union rule $\mathcal{U}$ is applied until we end up with the table below. Further applications do not result in new MVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×		×			
B	×		×	×			×	
C	×			×				
AB	×	×	×	×	×	×	×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\varphi \notin \Sigma^+_{\mathfrak{S}'}$, but $\varphi \in \Sigma^+_{\mathfrak{S}' \cup \{\mathcal{T}\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{T}$ from $\mathfrak{S}'$. □

Lemma 3.9 *Let $\mathfrak{S} = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{D}, \mathcal{I}\}$. The additive transitivity rule $\mathcal{T}^*$ is independent of $\mathfrak{S}$.*

Proof Let $\Sigma = \{A \twoheadrightarrow B, B \twoheadrightarrow C\}$ and $\varphi = A \twoheadrightarrow BC$. The closure $\Sigma^+_{\mathfrak{S}}$ may be generated as follows. First, we enter the MVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. Then, we apply the pseudo-transitivity rule $\mathcal{T}$ to infer $A \twoheadrightarrow C$. Several applications of the augmentation rule $\mathcal{A}$ result in the table below. Further applications do not result in new MVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×	×	×	×		
B	×		×	×			×	
C	×			×				
AB	×	×	×	×	×	×	×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{T}^*\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{T}^*$ from $\mathfrak{S}$. □

Theorem 3.3 *The only minimal, sound and complete subsets of $\mathfrak{S}_U$ for the implication of MVDs are $\mathfrak{S}_1$, $\mathfrak{S}_2$ and $\mathfrak{S}_3$.*

Proof Lemmas 3.1 and 3.2 show that $\mathcal{R}$ and $\mathcal{S}$ must be part of any complete subset of $\mathfrak{S}_U$. For a complete subset $\mathfrak{S} \subseteq \mathfrak{S}_U$ every inference rule in $\mathfrak{S}_U - \mathfrak{S}$ must be derivable from $\mathfrak{S}$. This follows from the following line of reasoning. Suppose there is an inference rule $\mathfrak{R} \in \mathfrak{S}_U - \mathfrak{S}$ that is independent of $\mathfrak{S}$. That is, there is some relation schema R and some set $\Sigma \cup \{\varphi\}$ of MVDs on R such that $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathfrak{R}\}} - \Sigma^+_{\mathfrak{S}}$. In particular, $\Sigma^+_{\mathfrak{S}} \subset \Sigma^+_{\mathfrak{S} \cup \{\mathfrak{R}\}}$. Since all the rules in $\mathfrak{S}_U$ are sound for R-implication we derive the contradiction $\Sigma^*_R \subseteq \Sigma^+_{\mathfrak{S}} \subset \Sigma^+_{\mathfrak{S} \cup \{\mathfrak{R}\}} \subseteq \Sigma^*_R$.

We consider every subset $\mathfrak{S}$ of $\mathfrak{S}_U$ that includes at least $\mathcal{R}$ and $\mathcal{S}$. If $\mathfrak{S}$ is not a superset of $\mathfrak{S}_1$, $\mathfrak{S}_2$ or $\mathfrak{S}_3$, then at least one of Lemmas 3.7 or 3.8 or 3.9 shows that there is some inference rule in $\mathfrak{S}_U - \mathfrak{S}$ that is independent of $\mathfrak{S}$, i.e., $\mathfrak{S}$ cannot be complete.

It follows from Theorems 3.1 and 3.2 as well as [11] that $\mathfrak{S}_1$, $\mathfrak{S}_2$ and $\mathfrak{S}_3$ are complete. The minimality of all three sets follows from the fact that no proper subset of $\mathfrak{S}_1$, or $\mathfrak{S}_2$ or $\mathfrak{S}_3$, respectively, is still complete. □

4 Minimality and strong minimality

In the last section we have revealed all minimal axiomatisations of MVDs in undetermined universes. Consequently, a set of inference rules is an axiomatisation precisely if it is a superset of a minimal axiomatisation. This tells the data administrator precisely which rules together can be used to enumerate all implicit knowledge. So far, minimality has been defined as the property of those complete sets of inference rules for which none of the rules can be omitted without losing the ability to infer all dependencies. Strictly speaking, this is not the strongest form of minimality one may define. We will use this section to explore two different ways of pushing the notion of minimality. In the first case we give examples of axiomatisations in which the conditions, under which an inference rule is applicable, are weakened further. In the second case we introduce the new notion of strong minimality as the property of those complete sets of inference rules for which none of the rules can be omitted without losing the ability to infer all *trivial* dependencies. The notion of strong minimality enables us to further classify previously established inference systems for MVDs in fixed universes. In sharp contrast, none of our minimal MVD axiomatisations in undetermined universes is strongly minimal. Moreover, we characterise trivial MVDs in undetermined universes, and thereby reveal further differences to the notion of MVDs in fixed universes.

4.1 Weakening reflexivity

Although the set $\mathfrak{S}_1 = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*\}$ is minimal, the reflexivity axiom $\mathcal{R}$ can be replaced by the much weaker empty-set-axiom $\mathcal{R}_\emptyset$ resulting in the complete set $\mathfrak{S}_0$. This is because $\mathcal{R}$ is derivable from $\mathcal{R}_\emptyset$ and $\mathcal{A}$.

Instead of using the reflexivity axiom $\mathcal{R}$ within $\mathfrak{S}_2$ we may also use the empty-set-axiom $\mathcal{R}_\emptyset$ together with the membership-axiom $\mathcal{M}$.

Theorem 4.1 *The set $\mathfrak{S}_4 = \{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{S}, \mathcal{T}, \mathcal{U}\}$, consisting of empty-set-axiom, membership-axiom, subset rule, pseudo-transitivity rule and union rule, is sound and complete for the implication of MVDs.*

Proof We show that the reflexivity axiom $\mathcal{R}$ is derivable from $\{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{T}, \mathcal{U}\}$. If $X = \emptyset$, then the only instance of the reflexivity axiom is the empty-set-axiom $\mathcal{R}_\emptyset$. We may therefore assume that $X \neq \emptyset$. We proceed by induction on the number n of attributes in Y. If $n = 0$, then we obtain the following inference:

$$\begin{array}{rc} \mathcal{M}: & \overline{X \twoheadrightarrow A}^{A\in X} \quad \mathcal{M}: \overline{A \twoheadrightarrow A}^{A\in\{A\}} \\ \cline{2-2} \mathcal{T}: & X \twoheadrightarrow \emptyset \end{array} .$$

Suppose $Y = \{A_1, \ldots, A_n, A_{n+1}\}$. Note that $\{A_1, \ldots, A_n\} \subseteq X$ and $A_{n+1} \in X$ as $Y \subseteq X$. We then have the following inference

$$\begin{array}{rc} \mathcal{R}_{\text{hypothesis}}: & \overline{X \twoheadrightarrow \{A_1, \ldots, A_n\}}^{\{A_1,\ldots,A_n\}\subseteq X} \quad \mathcal{M}: \overline{X \twoheadrightarrow A_{n+1}}^{A_{n+1}\in X} \\ \cline{2-2} \mathcal{U}: & X \twoheadrightarrow Y \end{array} .$$

This completes the proof. □

We may also replace $\mathcal{R}$ in $\mathfrak{S}_3$ by $\mathcal{R}_\emptyset$ and $\mathcal{M}$.

Theorem 4.2 *The set $\mathfrak{S}_5 = \{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{T}^*, \mathcal{D}, \mathcal{U}, \mathcal{S}\}$, which consists of empty-set-axiom, membership axiom, additive pseudo-transitivity rule, difference rule, union rule, and subset rule, is sound and complete for the implication of MVDs.*

Proof Theorem 3.1 and Lemma 3.5 show that $\{\mathcal{R}, \mathcal{S}, \mathcal{T}^*, \mathcal{D}, \mathcal{U}\}$ is complete. The proof of Theorem 4.1 shows then that $\mathcal{R}$ can be replaced by $\mathcal{R}_\emptyset$ and $\mathcal{M}$ still maintaining completeness. However, while the union rule $\mathcal{U}$ is derivable from $\{\mathcal{R}, \mathcal{T}^*, \mathcal{D}\}$, it is independent of $\mathfrak{S} = \{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{T}^*, \mathcal{D}, \mathcal{S}\}$. Namely, let $\Sigma = \emptyset$ and $\varphi = AB \twoheadrightarrow AB$. The following table can be generated as follows. First, the empty-set-axiom $\mathcal{R}_\emptyset$ and the membership-axiom are applied as often as possible. Finally, three applications of the difference rule $\mathcal{D}$ yield the $\emptyset$-column. The MVDs are closed under applications of inference rules in $\mathfrak{S}$.

	$\emptyset$	A	B	AB
$\emptyset$	×			
A	×	×		
B	×		×	
AB	×	×	×	

Since $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{U}\}}$ we have found witnesses Σ and φ to the independence of $\mathcal{U}$ from $\mathfrak{S}$. □

4.2 Strong minimality

We will now study the following novel notion of minimality, which we first introduce in the context of undetermined universes. Intuitively, we define strong minimality as the property of those complete sets of inference rules for which none of the rules can be omitted without losing the ability to infer all *trivial* MVDs.

Definition 4.1 Let $\mathfrak{S}$ be a complete set of inference rules for the implication of MVDs. The set $\mathfrak{S}$ is said to be *strongly minimal* for the implication of MVDs if and only if for all inference rules $\mathfrak{R} \in \mathfrak{S}$ there is some trivial MVD φ such that $\varphi \notin \Sigma^+_{\mathfrak{S}-\{\mathfrak{R}\}}$ but $\varphi \in \Sigma^+_{\mathfrak{S}}$ where $\Sigma = \emptyset$. □

Definition 4.1 utilises the notion of a trivial MVD. An MVD $X \twoheadrightarrow Y$ is said to be *trivial* if and only if all relations r to which this MVDs applies satisfy $X \twoheadrightarrow Y$, i.e., for all relations r such that $XY \subseteq Attr(r)$ holds we have $\models_r X \twoheadrightarrow Y$. The following result characterises trivial MVDs in undetermined universes syntactically.

Lemma 4.1 *In undetermined universes, an MVD $X \twoheadrightarrow Y$ is trivial if and only if $Y \subseteq X$ holds.*

Proof Let r be an arbitrary relation to which $X \twoheadrightarrow Y$ applies, i.e., $XY \subseteq Attr(r)$. If $Y \subseteq X$, then $X(Attr(r) - Y) = Attr(r)$. Consequently, we have $r[XY] \bowtie r[X(Attr(r) - Y)] = r[XY] \bowtie r[Attr(r)] = r$, i.e., $\models_r X \twoheadrightarrow Y$. Since r was arbitrary, the MVD $X \twoheadrightarrow Y$ is trivial.

Let Y not be a subset of X, i.e., $X - Y \neq \emptyset$. Define the following two-tuple relation $r = \{t_1, t_2\}$ where $Attr(r)$ is the disjoint union of $X, Y - X$ and the attribute B, and where $t_1[A] = t_2[A]$ if and only if $A \in X$. It is not difficult to see that $r \neq r[XY] \bowtie r[X(Attr(r) - Y)]$, and since $XY \subseteq Attr(r)$ we conclude that $X \twoheadrightarrow Y$ is not trivial. □

Notice the difference to MVDs $X \twoheadrightarrow Y$ in fixed universes R which are trivial precisely if $X \subseteq Y$ or $XY = R$ holds. For fixed universes, a set $\mathfrak{S}$ that is R-complete for the R-implication of MVDs for all relation schemata R is said to be *strongly minimal* for the R-implication of MVDs if and only if for all inference rules $\mathfrak{R} \in \mathfrak{S}$ there is some relation schema R and some trivial MVD φ on R such that $\varphi \notin \Sigma^+_{\mathfrak{S}-\mathfrak{R}}$ but $\varphi \in \Sigma^+_{\mathfrak{S}}$ where $\Sigma = \emptyset$. This notion helps to provide some further insight into differences between various axiomatisations for MVDs proposed in the literature. Mendelzon's minimal set $\mathfrak{M} = \{\mathcal{R}, \mathcal{C}_R, \mathcal{T}\}$ for MVDs [34] is not strongly minimal.

Theorem 4.3 *The minimal set* $\mathfrak{M} = \{\mathcal{R}, \mathcal{C}_R, \mathcal{T}\}$ *is not strongly minimal.*

Proof The result follows immediately from the fact that the subset $\{\mathcal{R}, \mathcal{C}_R\}$ is sufficient to infer all trivial MVDs on R. Suppose that $X \twoheadrightarrow Y$ is an arbitrary trivial MVD defined on R. If $Y \subseteq X$, then $X \twoheadrightarrow Y$ can be inferred in a single step by using the reflexivity axiom $\mathcal{R}$. If $X \cup Y = R$, then $R - Y \subseteq X$ and we can infer $X \twoheadrightarrow (R - Y)$ by means of the reflexivity axiom $\mathcal{R}$. Subsequently, we use the complementation rule $\mathcal{C}_R$ to infer $X \twoheadrightarrow Y$ from $X \twoheadrightarrow (R - Y)$. Thus, every trivial MVD can already be inferred by $\{\mathcal{R}, \mathcal{C}_R\}$. □

A further complete set of inference rules that is minimal for the R-implication of MVDs is $\{\mathcal{C}.1, \mathcal{A}, \mathcal{T}\}$ [10].

Theorem 4.4 *The minimal set* $\{\mathcal{C}.1, \mathcal{A}, \mathcal{T}\}$ *is strongly minimal.*

Proof Consider first the system $\mathfrak{S} = \{\mathcal{A}, \mathcal{T}\}$, and let $R = \emptyset$, $\Sigma = \emptyset$ and $\varphi = \emptyset \twoheadrightarrow R$. Note that φ is trivial on R. It follows that $\varphi \notin \Sigma^+_{\mathfrak{S}} = \emptyset$ but $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{C}.1\}}$.

Next consider the system $\mathfrak{S} = \{\mathcal{C}.1, \mathcal{T}\}$, and let $R = A$, $\Sigma = \emptyset$ and $\varphi = \emptyset \twoheadrightarrow \emptyset$. Note that φ is trivial on R. It follows that $\varphi \notin \Sigma^+_{\mathfrak{S}} = \{\emptyset \twoheadrightarrow A\}$ but $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{A}\}}$.

Finally, consider the system $\mathfrak{S} = \{\mathcal{C}.1, \mathcal{A}\}$, and let $R = A$, $\Sigma = \emptyset$ and $\varphi = A \twoheadrightarrow A$. Note that φ is trivial on R. It follows that $\varphi \notin \Sigma^+_{\mathfrak{S}} = \{\emptyset \twoheadrightarrow A, A \twoheadrightarrow A\}$ but $\varphi \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{T}\}}$.

This shows that none of the inference rules of $\{\mathcal{C}.1, \mathcal{A}, \mathcal{T}\}$ can be omitted without losing the ability to infer all trivial MVDs. □

There are more strongly minimal axiomatisations for MVDs that have previously been proposed in the literature. In fact, the minimality proofs in [22] show that each of the three sets

$$\{\mathcal{C}.1, \mathcal{M}, \mathcal{T}, \mathcal{U}\}, \{\mathcal{C}.1, \mathcal{M}, \mathcal{T}, \mathcal{D}\}, \{\mathcal{C}.1, \mathcal{M}, \mathcal{T}, \mathcal{I}\}$$

is strongly minimal.

For undetermined universes, none of the sets $\mathfrak{S}_0, \ldots, \mathfrak{S}_5$ is strongly minimal. This fact is relatively easy to see since the reflexivity axiom $\mathcal{R}$ is either already part of those systems considered or can be inferred from a proper subsystem (see the proof of Theorem 4.1, and Lemma 3.5).

Theorem 4.5 *None of the minimal sets* $\mathfrak{S}_0, \mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \mathfrak{S}_4, \mathfrak{S}_5$ *is strongly minimal.* □

This warrants future research to find out whether there is any strongly minimal axiomatisation for MVDs in undetermined universes.

5 All complementary axiomatisations of MVDs in fixed universes

In this section we will reveal all complementary axiomatisations of MVDs in fixed universes. This provides the data administrator with various choices on the set of inference rules sufficient for deriving either all implicit knowledge after fixing a relation schema, or for deriving almost all implicit knowledge before such a fixation.

Recall that $\mathfrak{S}_0^{\mathcal{C}} = \mathfrak{S}_0 \cup \{\mathcal{C}_R\}$, i.e., we obtain the (minimal) complementary axiomatisation $\mathfrak{S}_0^{\mathcal{C}}$ in fixed universes from the (minimal) axiomatisation $\mathfrak{S}_0$ in undetermined universes by adding the complementation rule $\mathcal{C}_R$. We will show in this section that this correspondence is not just limited to this particular (minimal) set of inference rules, but holds in general. This result will then enable us to determine all complementary axiomatisations in fixed universes.

Theorem 5.1 *Let $\mathfrak{S}$ be a sound set of inference rules for the implication of MVDs. The set $\mathfrak{S}$ is complete for the implication of MVDs if and only if for all relation schemata R the set $\mathfrak{S}^{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is R-complete and R-complementary for the R-implication of MVDs.*

Proof We show first that if $\mathfrak{S}$ is complete for the implication of MVDs, then for each relation schema R the set $\mathfrak{S}^{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is both R-complete and R-complementary for the R-implication of MVDs.

Let R be arbitrary. We know that $\mathfrak{S}_0^{\mathcal{C}}$ is R-complete, i.e., $\Sigma_R^* \subseteq \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$. Moreover, $\mathfrak{S}$ and $\mathfrak{S}_0$ are both sound and complete, i.e., $\Sigma^+_{\mathfrak{S}_0} = \Sigma^* = \Sigma^+_{\mathfrak{S}}$. Let $X, Y \subseteq R$ and $X \twoheadrightarrow Y \in \Sigma_R^*$. Since $\mathfrak{S}_0^{\mathcal{C}}$ is R-complete it follows that $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$. Equation (2.1) shows that $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}_0}$ holds. Since $\Sigma^+_{\mathfrak{S}_0} = \Sigma^+_{\mathfrak{S}}$ this is equivalent to $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}}$. However, $Y = R - (R - Y)$ and therefore $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}^{\mathcal{C}}}$. This shows that $\Sigma_R^* \subseteq \Sigma^+_{\mathfrak{S}^{\mathcal{C}}}$, i.e., $\mathfrak{S}^{\mathcal{C}}$ is R-complete. Moreover, $\Sigma^+_{\mathfrak{S}^{\mathcal{C}}} = \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$ and Eq. (2.1) imply that

$$X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}^{\mathcal{C}}} \quad \text{if and only if} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}} \text{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}}$$

whenever $X, Y \subseteq R$ and Σ is a set of MVDs on R. That is, every inference of an MVD $X \twoheadrightarrow Y$ using $\mathfrak{S}^{\mathcal{C}}$ can be turned into an inference of $X \twoheadrightarrow Y$ in which the R-complementation rule $\mathcal{C}_R$ is applied at most once, and if it is applied, then as the last rule of the inference. Therefore, the set $\mathfrak{S}^{\mathcal{C}}$ is R-complete and R-complementary for the R-implication of MVDs on all relation schemata R.

It remains to show that $\mathfrak{S}$ is complete for the implication of MVDs whenever for all relation schemata R the set $\mathfrak{S}^{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is both R-complete and R-complementary for the R-implication of MVDs. We need to show that $\Sigma^* \subseteq \Sigma^+_{\mathfrak{S}}$ holds for every finite set Σ of MVDs. Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \dots, X_k \twoheadrightarrow Y_k\}$ and $X \twoheadrightarrow Y \in \Sigma^* = \Sigma^+_{\mathfrak{S}_0}$. Let $T := X \cup Y \cup \bigcup_{i=1}^k (X_i \cup Y_i)$ and R be some relation schema such that T is properly contained in R, i.e., $T \subset R$. Fact 1 in [11] shows that $X \twoheadrightarrow Y \in \Sigma_R^*$. The R-completeness of $\mathfrak{S}^{\mathcal{C}}$ implies further that $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}^{\mathcal{C}}}$. Since $\mathfrak{S}^{\mathcal{C}}$ is also R-complementary we must have $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}}$. Assume that $X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}}$. The soundness of $\mathfrak{S}$ implies that $X \twoheadrightarrow (R - Y) \in \Sigma^* = \Sigma^+_{\mathfrak{S}_0}$. Furthermore, the derivability of the union rule from $\mathfrak{S}_0$ implies that $X \twoheadrightarrow R \in \Sigma^+_{\mathfrak{S}_0}$. However, we obtain the contradiction $R \subseteq T \subset R$ by Lemma 5 of [11]. Consequently, $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}}$ must hold, and this shows the completeness of $\mathfrak{S}$. □

Corollary 5.1 *For all relation schemata R the sets $\mathfrak{S}_1^{\mathcal{C}}, \mathfrak{S}_2^{\mathcal{C}}, \mathfrak{S}_3^{\mathcal{C}}, \mathfrak{S}_4^{\mathcal{C}}, \mathfrak{S}_5^{\mathcal{C}}$ are R-sound, R-complete and R-complementary for the R-implication of MVDs.* □

If $\mathfrak{S}$ is a minimal, sound and complete set of inference rules for the implication of MVDs, then it is not necessarily true that the set $\mathfrak{S}^{\mathcal{C}}$ is minimal for the R-implication of MVDs. In fact, the next theorem indicates that complete and complementary sets $\mathfrak{S}^{\mathcal{C}}$ may not be minimal.

Theorem 5.2 *The complete sets $\mathfrak{S}_1^{\mathcal{C}}$, $\mathfrak{S}_2^{\mathcal{C}}$, $\mathfrak{S}_3^{\mathcal{C}}$, $\mathfrak{S}_4^{\mathcal{C}}$, and $\mathfrak{S}_5^{\mathcal{C}}$ are not minimal for the R-implication of MVDs.*

Proof $\mathfrak{S}_1^{\mathcal{C}} = \{\mathcal{R}, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{C}_R\}$ and $\mathfrak{S}_2^{\mathcal{C}} = \{\mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{U}, \mathcal{C}_R\}$ are not minimal as $\mathcal{S}$ is derivable from $\{\mathcal{C}_R, \mathcal{A}, \mathcal{T}\}$ according to [11,Theorem 1].

$\mathfrak{S}_3^{\mathcal{C}} = \{\mathcal{R}, \mathcal{S}, \mathcal{T}^*, \mathcal{D}, \mathcal{C}_R\}$ is not minimal as $\mathcal{A}$ is derivable from $\{\mathcal{R}, \mathcal{T}^*, \mathcal{D}\}$ (recall that the union rule $\mathcal{U}$ is derivable from $\{\mathcal{R}, \mathcal{T}^*, \mathcal{D}\}$, Lemma 3.6):

$$\mathcal{D}: \dfrac{\mathcal{U}: \dfrac{\mathcal{U}: \dfrac{\mathcal{R}: \overline{XU \twoheadrightarrow X}^{X\subseteq XU} \quad X \twoheadrightarrow Y}{XU \twoheadrightarrow XY} \quad \mathcal{R}: \overline{XU \twoheadrightarrow V}^{V\subseteq XU}}{XU \twoheadrightarrow XVY} \quad \mathcal{R}: \overline{XU \twoheadrightarrow (X-(VY))}}{XU \twoheadrightarrow \underbrace{XVY-(X-(VY))}_{=YV}}$$

and the subset rule $\mathcal{S}$ is derivable from $\{\mathcal{A}, \mathcal{T}^*, \mathcal{C}_R\}$:

$$\mathcal{C}_R: \dfrac{\mathcal{T}^*: \dfrac{\mathcal{C}_R: \dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow Y^{\mathcal{C}}} \quad \mathcal{C}_R: \dfrac{\mathcal{A}: \dfrac{W \twoheadrightarrow Z}{Y^{\mathcal{C}} \twoheadrightarrow Z}\, W\subseteq Y^{\mathcal{C}}}{Y^{\mathcal{C}} \twoheadrightarrow Z^{\mathcal{C}}}}{X \twoheadrightarrow Y^{\mathcal{C}} \cup Z^{\mathcal{C}}}}{X \twoheadrightarrow \underbrace{(Y^{\mathcal{C}} \cup Z^{\mathcal{C}})^{\mathcal{C}}}_{=Y\cap Z}} .$$

$\mathfrak{S}_4^{\mathcal{C}} = \{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{S}, \mathcal{T}, \mathcal{U}, \mathcal{C}_R\}$ is not minimal. The proof of Theorem 4.1 shows that $\mathcal{R}$ is derivable from $\{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{T}, \mathcal{U}\}$. Furthermore, the proof of Lemma 2 in [22] shows that $\mathcal{A}$ is derivable from $\{\mathcal{R}, \mathcal{T}, \mathcal{U}\}$. Finally, the subset rule $\mathcal{S}$ is derivable from $\{\mathcal{C}_R, \mathcal{A}, \mathcal{T}\}$ [11,Theorem 1].

$\mathfrak{S}_5^{\mathcal{C}} = \{\mathcal{R}_\emptyset, \mathcal{M}, \mathcal{T}^*, \mathcal{D}, \mathcal{U}, \mathcal{S}, \mathcal{C}_R\}$ is not minimal. The DeMorgan rules imply that the difference rule $\mathcal{D}$ is derivable from the R-complementation rule $\mathcal{C}_R$ and the union rule $\mathcal{U}$. □

In particular, in order to make the minimal set $\mathfrak{M}$ also complementary one may add the subset rule $\mathcal{S}$ and the union rule $\mathcal{U}$ in order to obtain $\mathfrak{S}_2^{\mathcal{C}}$.

Corollary 5.2 *Let $\mathfrak{S}$ be a sound set of inference rules for the implication of MVDs. The set $\mathfrak{S}$ is minimal and complete for the implication of MVDs if and only if for all relation schemata R the set $\mathfrak{S}^{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is R-complete and R-complementary for the R-implication of MVDs, and there is no inference rule $\mathfrak{R} \in \mathfrak{S}^{\mathcal{C}}$ such that for all relation schemata R the set $\mathfrak{S}^{\mathcal{C}} - \{\mathfrak{R}\}$ is still both R-complete and R-complementary for the R-implication of MVDs.* □

The last corollary helps us find all subsets of $\mathfrak{S}_U^{\mathcal{C}}$ that are complete and complementary for the R-implication of MVDs. The next lemma is an extension of [34, Lemma 1].

Lemma 5.1 *The R-complementation rule $\mathcal{C}_R$ is independent of $\mathfrak{S}_U$.*

Proof Let $R = A$, $\Sigma = \emptyset$ and $\varphi = \emptyset \twoheadrightarrow A$. Since $\varphi \notin \Sigma^+_{\mathfrak{S}_U}$, but $\varphi \in \Sigma^+_{\mathfrak{S}_U \cup \{\mathcal{C}_R\}}$ we have found witnesses R, Σ and φ to the independence of $\mathcal{C}_R$ from $\mathfrak{S}_U$. □

The next corollary is a consequence of Theorem 3.3, Corollary 5.2 and Lemma 5.1.

Corollary 5.3 *There are no proper subsets of $\mathfrak{S}_1^{\mathcal{C}}$, $\mathfrak{S}_2^{\mathcal{C}}$ and $\mathfrak{S}_3^{\mathcal{C}}$ which are both for all relation schemata R: R-complete and R-complementary for the R-implication of MVDs.* □

6 The implication problem

In the previous sections we have established axiomatisations of MVDs in undetermined universes, and complementary axiomatisations of MVDs in fixed universes. This provides us with various choices for mechanically inferring all implicitly specified knowledge. Quite often, we are not interested in all this implicit knowledge but want to determine whether some fixed MVD is implied by the set of explicitly specified MVDs. In this case, an enumeration procedure provided by an axiomatisation is too expensive as we are not looking for all implicit knowledge. Consequently, the implication problem of MVDs must be studied. This has been the subject of many papers where a fixed universe is assumed to be given. However, as a universe cannot always be assumed to be fixed in practice the implication problem should also be addressed for undetermined universes. In fact, the results in this section will show that the assumption of a fixed universe is not necessary for solving the implication problem efficiently.

The implication problem for MVDs in fixed universes is to decide whether for an arbitrary relation schema R and an arbitrary set $\Sigma \cup \{\varphi\}$ of MVDs on R, Σ R-implies φ in the sense of Definition 2.1. This problem has been well-studied [4,15,19,20,26,28,36,37,43]. The fundamental notion is that of a dependency basis for an attribute set $X \subseteq R$ with respect to a set Σ of MVDs [4]. Given some set $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ of MVDs with $\bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R$ the set $Dep_R(X) = \{Y \mid X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}\}$ consists of all those attribute sets $Y \subseteq R$ such that $X \twoheadrightarrow Y$ is derivable from Σ using some R-sound and R-complete set of inference rules for the R-implication of MVDs, in this case $\mathfrak{S}_0^{\mathcal{C}}$. The structure $(Dep_R(X), \subseteq, \cup, \cap, -, \emptyset, R)$ is a finite Boolean powerset algebra due to the derivability of union, intersection and difference rule from $\mathfrak{S}_0^{\mathcal{C}}$. Recall that an element $a \in P$ of a poset $(P, \sqsubseteq, 0)$ with least element 0 is called an *atom* of $(P, \sqsubseteq, 0)$ [9] if and only if $a \neq 0$ and every element $b \in P$ with $b \sqsubseteq a$ satisfies $b = 0$ or $b = a$. $(P, \sqsubseteq, 0)$ is called *atomic* if and only if for every element $b \in P$ with $b \neq 0$ there is an atom $a \in P$ with $a \sqsubseteq b$. In particular, every finite Boolean algebra is atomic. The set $DepB_R(X)$ of all atoms of $(Dep_R(X), \subseteq, \emptyset)$ is called the dependency basis of X with respect to Σ.

We will now study the implication problem for MVDs. The problem is to decide whether an arbitrary finite set Σ of MVDs implies a single MVD φ in the sense of Definition 2.2. Therefore, the following definition introduces the notion of a dependency basis for undetermined universes.

Definition 6.1 Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ be a set of MVDs, and $X \subseteq \mathfrak{A}$ some finite set of attributes. Let $Dep_U(X) = \{Y \mid X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}\}$ be the set of all attribute sets Y such that $X \twoheadrightarrow Y$ is derivable from Σ using $\mathfrak{S}_0$. The *dependency basis* $DepB_U(X)$ of X with respect to Σ is the set of all atoms of $(Dep_U(X), \subseteq, \emptyset)$. □

Note that $Dep_U(X)$ is invariant under different choices of sound and complete sets of inference rules for the implication of MVDs. More precisely, if $\mathfrak{S}_1$ and $\mathfrak{S}_2$ are two different sound and complete sets of inference rules for the implication of MVDs, then $\{Y \mid X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_1}\} = \{Y \mid X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_2}\}$. We will now introduce the notion of the Σ-scope for an attribute set X.

Definition 6.2 Let Σ be a finite set of MVDs and $X \subseteq \mathfrak{A}$ some finite set of attributes. The set $X^S = \bigcup Dep_U(X)$ is called the *scope of X with respect to* Σ (or short Σ-scope of X). □

If $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$, then $X^S \subseteq R_{\min} := X \cup \bigcup_{i=1}^{k}(X_i \cup Y_i)$ according to [11, Lemma 5]. It follows immediately that $X^S \in Dep_U(X)$ since the union rule is derivable from any complete set of inference rules for the implication of MVDs. Moreover, if

$Y \in Dep_U(X)$, then $Y \subseteq X^S$, i.e., X^S is the maximal element of $Dep_U(X)$ with respect to $\subseteq$.

In fact, union, intersection and difference rule are all derivable from any complete set of inference rules. Therefore $(Dep_U(X), \subseteq, \cup, \cap, -, \emptyset, X^S)$ is again a finite Boolean algebra, with top-element X^S. The existence and uniqueness of $DepB_U(X)$ follow from the fact that every finite Boolean algebra is atomic.

The notion of a dependency basis gains its importance from the fact that $X \twoheadrightarrow Y$ is logically R-implied by Σ if and only if Y is the union of some sets of $DepB_R(X)$ [4]. In order to solve the R-implication problem, it is therefore sufficient to find an algorithm for computing the dependency basis. The following theorem shows that the same is true for $DepB_U(X)$, i.e., for implication in undetermined universes.

Theorem 6.1 *Let Σ be a finite set of MVDs. Then $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ if and only if $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq DepB_U(X)$.*

Proof If $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$, then $Y \in Dep_U(X)$. That means $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq DepB_U(X)$ since $DepB_U(X)$ consists of all atoms of $(Dep_U(X), \subseteq, \emptyset)$.

Vice versa, if $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq DepB_U(X)$, then $Y \in Dep_U(X)$ according to the derivability of the union rule from $\mathfrak{S}_0$. It follows that $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ holds. □

Considerable effort has been devoted to finding fast algorithms that compute $DepB_R(X)$ given X and given Σ, see for instance [4,15,19,20,26,28,36,37,43]. Currently, the best upper bound for solving $\Sigma \models X \twoheadrightarrow Y$ is $\mathcal{O}((1 + \min\{s, \log p\}) \cdot n)$ from [19] where s denotes the number of dependencies in Σ, p the number of sets in $DepB_R(X)$ that have non-empty intersection with Y and n denotes the total number of occurrences of attributes in Σ.

We will show now that an extension of any algorithm that computes $DepB_R(X)$ for any R with $R_{\min} \subseteq R$ can be used to compute $DepB_U(X)$. The following theorem shows that $DepB_U(X)$ and $DepB_R(X)$ deviate in at most one element, namely $R - X^S$. Intuitively, that makes perfect sense since X^S is $\subseteq$-maximal among the attribute sets Y with $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ and, given that $X^S \subset R$, the R-complement $R - X^S$ of X^S is an atom of $(Dep_R(X), \subseteq, \emptyset)$.

Theorem 6.2 *Let Σ be a finite set of MVDs, $X \subseteq \mathfrak{A}$ some finite set of attributes, and R some relation schema with $R_{min} \subseteq R$. Then*

$$DepB_U(X) = \begin{cases} DepB_R(X), & \text{if } X^S = R \\ DepB_R(X) - \{R - X^S\}, & \text{if } X^S \subset R \end{cases}.$$

Proof Let $X^S = R$, i.e., in particular $X^S = R_{\min}$. We show that $Dep_U(X) = Dep_R(X)$ holds.

For $Y \in Dep_U(X)$ follows $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ and therefore also $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$ since $Y \subseteq X^S \subseteq R$. This, however, means that $Y \in Dep_R(X)$.

If $Y \in Dep_R(X)$, then $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$. Consequently,

$$X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0} \quad \text{or} \quad X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}_0}$$

by Eq. (2.1). Since $X^S = R$ we have in the latter case that $X \twoheadrightarrow (X^S - Y) \in \Sigma^+_{\mathfrak{S}_0}$. According to the derivability of the union rule from $\mathfrak{S}_0$ we have $X \twoheadrightarrow X^S \in \Sigma^+_{\mathfrak{S}_0}$, and due to the derivability of the difference rule from $\mathfrak{S}_0$ we also know that $X \twoheadrightarrow (X^S - (X^S - Y)) \in \Sigma^+_{\mathfrak{S}_0}$ holds. Since $(X^S - (X^S - Y)) = Y$, it follows that $Y \in Dep_U(X)$. Therefore, $Dep_U(X) = Dep_R(X)$ holds indeed and this shows that $DepB_U(X) = DepB_R(X)$ whenever $X^S = R$.

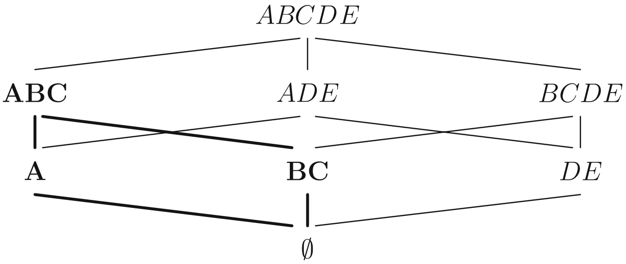

Fig. 1 Dependency basis in fixed and undetermined universe

Let now be $X^S \subset R$, i.e., $R - X^S \neq \emptyset$. We show first that $R - X^S \in DepB_R(X)$. First, $X^S \in Dep_R(X)$ by derivability of the union rule from $\mathfrak{S}_0^{\mathcal{C}}$. Then $R - X^S \in Dep_R(X)$ by application of the R-complementation rule from $\mathfrak{S}_0^{\mathcal{C}}$. Suppose there is some $Y \in Dep_R(X)$ with $Y \subseteq (R - X^S)$. We know again by equation (2.1) that $Y \in Dep_U(X)$ or $(R - Y) \in Dep_U(X)$ holds. In the first case we have $Y \subseteq X^S$ and therefore $Y \subseteq X^S \cap (R - X^S)$, i.e., $Y = \emptyset$. In the latter case we have $(R - Y) \subseteq X^S$, i.e., $(R - X^S) \subseteq Y$, i.e., $Y = (R - X^S)$. We have shown that every $Y \in Dep_R(X)$ with $Y \subseteq (R - X^S)$ satisfies $Y = \emptyset$ or $Y = R - X^S$. That means $R - X^S$ is an atom of $(Dep_R(X), \subseteq, \emptyset)$, i.e., $R - X^S \in DepB_R(X)$.

We show now that $DepB_U(X) \subseteq DepB_R(X)$. If $Y \in DepB_U(X)$, then $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0}$ in particular. Consequently, $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$. Suppose there is some Z with $\emptyset \neq Z \subset Y$ and $X \twoheadrightarrow Z \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$. Then either $X \twoheadrightarrow Z \in \Sigma^+_{\mathfrak{S}_0}$ which contradicts the assumption that Y is an atom of $(Dep_U(X), \subseteq, \emptyset)$, or $X \twoheadrightarrow (R - Z) \in \Sigma^+_{\mathfrak{S}_0}$. Consequently, $R - Z \subseteq X^S$ and since $R - Y \subseteq R - Z$ we have $R - Y \subseteq X^S$, too. However, from $Y \subseteq X^S$ follows then that $Y \cup (R - Y) \subseteq X^S$, i.e., $R = X^S$, a contradiction to our assumption. Therefore, $Y \in DepB_R(X)$. So far, we have shown that $DepB_U(X) \cup \{R - X^S\} \subseteq DepB_R(X)$.

As $R - X^S \notin DepB_U(X)$, it remains to show that

$$DepB_R(X) \subseteq DepB_U(X) \cup \{R - X^S\}$$

holds as well. Let Y be some atom of $(Dep_R(X), \subseteq, \emptyset)$. In particular, $Y \in Dep_U(X)$ or $(R - X^S) \subseteq Y$. In the first case, Y must also be an atom of $Dep_U(X)$ since $Dep_U(X) \subseteq Dep_R(X)$. In the second case assume that $Y \neq R - X^S$. Then we have $R - X^S \neq \emptyset$ and $R - X^S \subset Y$. However, $X \twoheadrightarrow (R - X^S) \in \Sigma^+_{\mathfrak{S}_0^{\mathcal{C}}}$ since $X \twoheadrightarrow X^S \in \Sigma^+_{\mathfrak{S}_0}$ and the R-complementation rule is in $\mathfrak{S}_0^{\mathcal{C}}$. This contradicts our assumption that Y is an atom of $Dep_R(X)$. Hence, $Y = R - X^S$. We have therefore shown that every atom of $Dep_R(X)$ is either an atom of $Dep_U(X)$ or equals $R - X^S$. □

The following example illustrates Theorem 6.2.

Example 6.1 Suppose $\Sigma = \{A \twoheadrightarrow BC\}$ and the underlying relation schema $R = \{A, B, C, D, E\}$. The Boolean algebra of right-hand sides Y of those MVDs $A \twoheadrightarrow Y$ which are R-implied by Σ is illustrated in Fig. 1. The dependency basis $DepB_R(A) = \{A, BC, DE\}$ consists of the atoms of the algebra. Since $A^S = ABC$ we obtain $DepB_U(A) = \{A, BC\}$ and the Boolean algebra of right-hand sides Y of those MVDs $A \twoheadrightarrow Y$ which are implied by Σ is printed boldly in Fig. 1. □

Theorem 6.2 provides us with a strategy for computing the dependency basis $DepB_U(X)$ of X with respect to $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$. In fact, one may use any algorithm for computing $DepB_{R_{\min}}(X)$ with respect to Σ. Having computed the Σ-scope X^S of X, one removes $R_{\min} - X^S$ from $DepB_{R_{\min}}(X)$ if and only if $X^S \subset R_{\min}$. Therefore, it remains to compute the Σ-scope X^S.

Recall from [5] the notion of a closure $\overline{X} = \bigcup\{Y \mid X \to Y \in \Sigma^+\}$ of an attribute set X with respect to a set Σ of functional dependencies. In order to compute $\overline{X}$ all FDs $Y \to Z \in \Sigma$ are inspected, and whenever $Y \subseteq \overline{X}$, then $\overline{X}$ is replaced by $\overline{X} \cup Z$. This process is repeated until no further attributes have been added to $\overline{X}$ after inspecting all FD in Σ. The correctness of this algorithm is due to the soundness of the following inference rule: if $Y \to Z \in \Sigma^+$ and $Y \subseteq X$, then $X \to XZ \in \Sigma^+$.

The definition of $\overline{X}$ is very similar to that of the Σ-scope X^S. In fact, if $Y \twoheadrightarrow Z \in \Sigma^+$ and $Y \subseteq X$, then we also have $X \twoheadrightarrow XZ \in \Sigma^+$ according to the soundness of the reflexivity rule, additive transitivity rule and union rule. The idea is therefore to compute the Σ-scope X^S of X by employing essentially the linear-time algorithm from [5] for computing the closure of X with respect to a set of FDs.

Algorithm 1 (Dependency Basis)
Input: $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$, and a finite set X of attributes
Output: $DepB_U(X)$ with respect to $\boldsymbol{\Sigma}$
Method:

```
VAR $R_{\min}, X^S_{\text{new}}, X^S_{\text{old}}, X^S_{\text{alg}}$: Set of attributes; MVDList: List of MVDs;

 (1)  $R_{\min} := X \cup \bigcup_{i=1}^{k} (X_i \cup Y_i)$;
 (2)  Use the Algorithm in [19] to compute $DepB_{R_{\min}}(X)$;
 (3)  $X^S_{\text{new}} := X$;
 (4)  MVDList:= List of MVDs in $\Sigma$;
 (5)  REPEAT
 (6)    $X^S_{\text{old}} := X^S_{\text{new}}$;
 (7)    Remove all attributes in $X^S_{\text{new}}$ from the LHS of
          all MVDs in MVDList;
 (8)    FOR all MVDs $\emptyset \twoheadrightarrow Y$ in MVDList LET $X^S_{\text{new}} := X^S_{\text{new}} \cup Y$;
 (9)  UNTIL $X^S_{\text{new}} = X^S_{\text{old}}$;
(10)  $X^S_{\text{alg}} := X^S_{\text{new}}$;
(11)  IF $X^S_{\text{alg}} = R_{\min}$ THEN RETURN($DepB_{R_{\min}}(X)$)
(12)                      ELSE RETURN($DepB_{R_{\min}}(X) - \{R_{\min} - X^S\}$);
```

The next theorem verifies essentially the correctness of Algorithm 1 in computing the Σ-closure X^S of X [i.e. the part between line (4) and line (10)].

Theorem 6.3 *Algorithm 1 computes $DepB_U(X)$ with respect to Σ in time $\mathcal{O}((1 + \min\{s, \log \bar{p}\}) \cdot n)$ where s denotes the number of dependencies in Σ, $\bar{p}$ the number of sets in $DepB_U(X)$ and n denotes the total number of occurrences of attributes in Σ.*

Proof If we show that the part between line (4) and line (10) of Algorithm 1 computes indeed the scope of X with respect to Σ, i.e. $X^S_{\text{alg}} = X^S$, then the correctness of Algorithm 1 follows from the correctness of the algorithm in [19] and Theorem 6.2.

In order to show that $X^S_{\text{alg}} = X^S$ we prove first that $X \twoheadrightarrow X^S_{\text{alg}} \in \Sigma^+$, i.e., $X^S_{\text{alg}} \subseteq X^S$ holds. We proceed by induction on the number j of runs through the REPEAT loop between line (5) and (9). If $j = 0$, then $X^S_{\text{new}} = X$ and $X \twoheadrightarrow X \in \Sigma^+$ by reflexivity. For $j > 0$ we assume that $X \twoheadrightarrow X^S_{\text{new}} \in \Sigma^+$ for X^S_{new} after the jth run through the REPEAT loop. Suppose there is some MVD $Z \twoheadrightarrow Y \in \Sigma$ such that $Z - X^S_{\text{new}} = \emptyset$, i.e., $Z \subseteq X^S_{\text{new}}$ (otherwise there is nothing to show). By derivability of the augmentation rule we infer $X^S_{\text{new}} \twoheadrightarrow Y \in \Sigma^+$. An application of the additive transitivity rule shows that $X \twoheadrightarrow X^S_{\text{new}} \cup Y \in \Sigma^+$. Therefore, $X \twoheadrightarrow X^S_{\text{new}} \in \Sigma^+$ holds also after the $j + 1$st run through the REPEAT loop since $X^S_{\text{new}} := X^S_{\text{new}} \cup Y$ in line (8) of Algorithm 1.

It remains to show that $X^S \subseteq X^S_{\text{alg}}$. Consider the chain $\Sigma = \Sigma_0 \subset \Sigma_1 \subset \cdots \subset \Sigma_k = \Sigma^+$ where Σ_{i+1} is generated from Σ_i by an application of a single inference rule from, say, the complete set $\mathfrak{S}_2$. We show by induction on i that

$$\text{if } U \twoheadrightarrow V \in \Sigma_i \quad \text{and} \quad U \subseteq X^S_{\text{alg}}, \text{ then also } V \subseteq X^S_{\text{alg}}. \tag{6.2}$$

For $i = k$ it follows that if $U \twoheadrightarrow V \in \Sigma^+$ and $U \subseteq X^S_{\text{alg}}$, then also $V \subseteq X^S_{\text{alg}}$. The proof concludes for $U = X$ and $V = X^S$ since $X \twoheadrightarrow X^S \in \Sigma^+$ by the derivability of the union rule. It remains to show (6.2). For $i = 0$ we have $U \twoheadrightarrow V \in \Sigma$ and $U \subseteq X^S_{\text{alg}}$ means that $U \subseteq X^S_{\text{new}}$ at some point in time. This implies that $\emptyset \twoheadrightarrow V$ occurs in the MVDList after line (7). Consequently, after line (8) we have $V \subseteq X^S_{\text{new}} \subseteq X^S_{\text{alg}}$ as well. If $i > 0$, then $\Sigma_{i+1} - \Sigma_i$ contains exactly one $U \twoheadrightarrow V$. There is nothing to show for the MVDs in Σ_i (hypothesis). Thus, it suffices to consider $U \twoheadrightarrow V$, and we distinguish between 4 different cases.

1. If $U \twoheadrightarrow V$ results from an application of the reflexivity axiom, then $V \subseteq U \subseteq X^S_{\text{alg}}$.
2. If $U \twoheadrightarrow V$ results from an application of the subset rule, then $U \twoheadrightarrow Z, Y \twoheadrightarrow W \in \Sigma_i$ with $Z \cap Y = \emptyset$ and $V = Z \cap W$. From $U \subseteq X^S_{\text{alg}}$ follows $Z \subseteq X^S_{\text{alg}}$ since $U \twoheadrightarrow Z \in \Sigma_i$. However, that means $V = Z \cap W \subseteq Z \subseteq X^S_{\text{alg}}$.
3. If $U \twoheadrightarrow V$ results from an application of the pseudo-transitivity rule, then $U \twoheadrightarrow W, W \twoheadrightarrow Z \in \Sigma_i$ and $V = Z - W$. From $U \subseteq X^S_{\text{alg}}$ follows $W \subseteq X^S_{\text{alg}}$ since $U \twoheadrightarrow W \in \Sigma_i$. Consequently, $W \subseteq X^S_{\text{alg}}$ implies $Z \subseteq X^S_{\text{alg}}$ since $W \twoheadrightarrow Z \in \Sigma_i$. That means, $V = Z - W \subseteq Z \subseteq X^S_{\text{alg}}$.
4. If $U \twoheadrightarrow V$ results from an application of the union rule, then $U \twoheadrightarrow W, U \twoheadrightarrow Z \in \Sigma_i$ and $V = W \cup Z$. From $U \subseteq X^S_{\text{alg}}$ follows $W \subseteq X^S_{\text{alg}}$ and $Z \subseteq X^S_{\text{alg}}$. Consequently, $V = W \cup Z \subseteq X^S_{\text{alg}}$.

The time complexity of Algorithm 1 is essentially the time complexity of the algorithm in [19]. The code between line (4) and (10) to compute the Σ-scope of X can be implemented in time $\mathcal{O}(n)$ [5]. Therefore, the time-complexity is $\mathcal{O}((1 + \min\{s, \log \bar{p}\}) \cdot n)$. □

Example 6.2 Consider a classical example [4]. Let

$$\Sigma = \{AB \twoheadrightarrow DEFG, CGJ \twoheadrightarrow ADHI\},$$

and suppose we want to compute $DepB_U(\{A, C, G, J\})$ with respect to Σ. We obtain $R_{\min} = \{A, B, C, D, E, F, G, H, I, J\}$ and

$$DepB_{R_{\min}}(\{A, C, G, J\}) = \{\{A\}, \{C\}, \{G\}, \{J\}, \{D\}, \{H, I\}, \{B, E, F\}\}$$

according to [4]. The Σ-scope of $\{A, C, G, J\}$ is

$$\{A, C, G, J\}^S = \{A, C, D, G, I, H, J\}.$$

Therefore, $\{A, C, G, J\}^S \subset R_{\min}$ and $R_{\min} - \{A, C, G, J\}^S = \{B, E, F\}$. That is,

$$DepB_U(\{A, C, G, J\}) = \{\{A\}, \{C\}, \{G\}, \{J\}, \{D\}, \{H, I\}\}$$

and the MVD $ACGJ \twoheadrightarrow BDEF$ which is $R_{\min}$-implied by Σ is not implied by Σ. □

Theorem 6.4 *Let Σ be a finite set of MVDs. The MVD $X \twoheadrightarrow Y$ is implied by Σ if and only if $Y \subseteq X^S$ and $X \twoheadrightarrow Y$ is R_{min}-implied by Σ.*

Proof If $X \twoheadrightarrow Y$ is implied by Σ, then Y is the union of some subset of $DepB_U(X) \subseteq DepB_{R_{\min}}(X)$. Consequently, $Y \subseteq X^S$ and $X \twoheadrightarrow Y$ is $R_{\min}$-implied by Σ.

If $X \twoheadrightarrow Y$ is $R_{\min}$-implied by Σ, then Y is the union of some subset of $DepB_{R_{\min}}(X)$. If $R_{\min} = X^S$, then $DepB_{R_{\min}}(X) = DepB_U(X)$ by Theorem 6.2 and $X \twoheadrightarrow Y$ is also implied by Σ according to Theorem 6.1. Otherwise $X^S \subset R_{\min}$. Since $Y \subseteq X^S$ holds as well, we have $Y \cap (R_{\min} - X^S) = \emptyset$, i.e., Y is in fact the union of some subset of

$$DepB_{R_{\min}}(X) - \{R_{\min} - X^S\} = DepB_U(X).$$

That means, $X \twoheadrightarrow Y$ is implied by Σ according to Theorem 6.1. □

Corollary 6.1 *The implication problem $\Sigma \models X \twoheadrightarrow Y$ can be decided in time $\mathcal{O}((1 + \min\{s, \log p\}) \cdot n)$ where s denotes the number of dependencies in Σ, p the number of sets in $DepB_U(X)$ that have non-empty intersection with Y and n denotes the total number of occurrences of attributes in Σ.*

Proof This result follows from the time bound in [19], Theorems 6.3 and 6.1. It can be decided in time $\mathcal{O}(n)$ whether $Y \subseteq X^S$ holds. If $Y \not\subseteq X^S$, then $X \twoheadrightarrow Y$ is not implied by Σ. Otherwise, we can decide in time $\mathcal{O}((1 + \min\{s, \log p\}) \cdot n)$ whether Y is the union of some subset of $DepB_U(X)$. Note that there is at most one more element in $DepB_{R_{\min}}(X)$ than in $DepB_U(X)$, namely $R_{\min} - X^S$. However, Y and $R_{\min} - X^S$ are disjoint since $Y \subseteq X^S$. □

If there is a linear-time algorithm for computing the dependency basis in a fixed universe, then Algorithm 1 also provides a linear-time algorithm for computing the dependency basis in undetermined universes. Vice versa, if there is a linear-time algorithm for computing the dependency basis in undetermined universes, then Theorem 6.2 shows that the dependency basis can also be computed in linear time in a fixed universe.

7 Related work and open problems

We use this final section to point the reader to earlier and related work on multivalued dependencies in relational databases, and to state some problems that warrant future research.

The original notion of MVD implication is based on a fixed underlying set of attributes, and was independently investigated by Delobel et al. [14,16,49]. The first axiomatisation of MVDs with respect to this notion of implication was presented by Beeri et al. [6]. The minimality of the inference rules were studied independently by Mendelzon and Biskup [10,34,22]. The time complexity of the implication problem for MVDs in fixed universes was studied by several authors [4,15,19,20,26,28,36,37,43].

The notion of MVD implication in undetermined universes was introduced by Biskup [11] who also proposed the first axiomatisation $\mathfrak{S}_0$ in the same paper. The results in this paper were partially presented in [31].

Lien introduced and axiomatised multivalued dependencies in the presence of null values (NMVDs) [30]. Again, the notion of NMVD implication is based on a fixed universe. Axiomatisations of NMVDs in undetermined universes can be found in [33]. Moreover, a strongly minimal axiomatisation for NMVDs in fixed universes can be found in [32].

Full hierarchical dependencies (FHDs) subsume the class of MVDs and were originally introduced by Delobel in the context of a fixed relation schema [14]. Axiomatisations for FHDs in fixed and undetermined universes can be found in [23].

Recently, MVDs have also been studied and axiomatised in the context of a fixed nested database schema that is generated from finitely many recursive applications of record and list constructor [25]. It is a challenging problem to study whether the complementation rule can be shifted towards the very end of an inference, and to find a suitable notion for MVD implication where the underlying nested database schema is left undetermined.

We will conclude the article by listing some open problems that warrant future research:

1. Are there any complete sets (not a subset of $\mathfrak{S}_U$) in which the subset rule $\mathcal{S}$ does not occur?
2. Are there any minimal sets of inference rules for MVD implication in fixed universes that are also complementary?
3. Consider MVDs in undetermined universes together with FDs. What are minimal complete sets for the implication of FDs and MVDs in undetermined universes?
4. Further investigate full hierarchical dependencies, for instance (the time complexity of) the implication problem of both FHDs and FDs, and FHDs in the presence of null values.
5. Study axiomatisability, and the time-complexity of the MVD implication problem in fixed and undetermined universes in the context of other data models, e.g.:
   - for conceptual data models such as the Entity-Relationship model [42],
   - fuzzy database models [39],
   - temporal database models [27],
   - and XML [38,45].
6. Investigate the suitability of notions for MVD implication in the context of views.
7. Find maximal sets of sound (and possibly incomplete) MVD inference rules for which the associated implication problem is still linear.
8. Characterise the expressive power for incomplete sets of MVD inference rules.

References

1. Arenas, M., Libkin, L.: A normal form for XML documents. ACM Trans. Database Syst. **29**(1), 195–232 (2004)
2. Armstrong, W.W.: Dependency structures of database relationships. Inf. Process. **74**, 580–583 (1974)
3. Armstrong, W.W., Nakamura, Y., Rudnicki, P.: Armstrong's axioms. J. Formalized Math. **14** (2002)
4. Beeri, C.: On the membership problem for functional and multivalued dependencies in relational databases. ACM Trans. Database Syst. **5**(3), 241–259 (1980)
5. Beeri, C., Bernstein, P.A.: Computational problems related to the design of normal form relational schemata. 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: Proceedings of the SIGMOD International Conference on Management of Data, pp. 47–61. ACM (1977)
7. Bernstein, P.: Synthesizing third normal form relations from functional dependencies. ACM Trans. Database Syst. **1**(4), 277–298 (1976)
8. Bernstein, P.A., Goodman, N.: What does Boyce–Codd normal form do? In: Proceedings of the 6th International Conference on Very Large Data Bases, pp. 245–259. IEEE Computer Society (1980)

9. Birkhoff, G.: Lattice Theory. American Mathematical Society (1940)
10. Biskup, J.: On the complementation rule for multivalued dependencies in database relations. Acta Inf. **10**(3), 297–305 (1978)
11. Biskup, J.: Inferences of multivalued dependencies in fixed and undetermined universes. Theor. Comput. Sci. **10**(1), 93–106 (1980)
12. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM **13**(6), 377–387 (1970)
13. Codd, E.F.: Further normalization of the database relational model. In: Courant Computer Science Symposia 6: Data Base Systems, pp. 33–64. Prentice-Hall, Englewood Cliffs (1972)
14. Delobel, C.: Normalisation and hierarchical dependencies in the relational data model. ACM Trans. Database Syst. **3**(3), 201–222 (1978)
15. Diederich, J., Milton, J.: New methods and fast algorithms for database normalization. ACM Trans. Database Syst. **13**(3), 339–365 (1988)
16. Fagin, R.: Multivalued dependencies and a new normal form for relational databases. ACM Trans. Database Syst. **2**(3), 262–278 (1977)
17. Fagin, R., Vardi, M.Y.: The theory of data dependencies: a survey. In: Mathematics of Information Processing: Proceedings of Symposia in Applied Mathematics, pp. 19–71. American Mathematical Society (1986)
18. Fischer, P.C., Saxton, L.V., Thomas, S.J., Van Gucht, D.: Interactions between dependencies and nested relational structures. J. Comput. Syst. Sci. **31**(3), 343–354 (1985)
19. Galil, Z.: An almost linear-time algorithm for computing a dependency basis in a relational database. J. ACM **29**(1), 96–102 (1982)
20. Hagihara, K., Ito, M., Taniguchi, K., Kasami, T.: Decision problems for multivalued dependencies in relational databases. SIAM J. Comput. **8**(2), 247–264 (1979)
21. Hara, C., Davidson, S.: Reasoning about nested functional dependencies. In: Proceedings of the 18th SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 91–100. ACM (1999)
22. Hartmann, S., Link, S.: On a problem of Fagin concerning multivalued dependencies in relational databases. Theor. Comput. Sci. **353**(1–3), 53–62 (2006)
23. Hartmann, S., Link, S.: Full hierarchical dependencies in fixed and undetermined universes. Ann. Math. Artif. Intell. **50**(1–2), 195–226 (2007)
24. Hartmann, S., Link, S.: Characterising nested database dependencies by fragments of propositional logic. Ann. Pure Appl. Logic **152**(1–3), 84–106 (2008)
25. Hartmann, S., Link, S., Schewe, K.-D.: Functional and multivalued dependencies in nested databases generated by record and list constructor. Ann. Math. Artif. Intell. **46**(1–2), 114–164 (2006)
26. Ito, M., Iwasaki, M., Taniguchi, K., Kasami, K.: Membership problems for data dependencies in relational expressions. Theor. Comput. Sci. **34**, 315–335 (1984)
27. Jensen, C., Snodgrass, R., Soo, M.: Extending existing dependency theory to temporal databases. IEEE Trans. Knowl. Data Eng. **8**(4), 563–582 (1996)
28. Lakshmanan, V., VeniMadhavan, C.: An algebraic theory of functional and multivalued dependencies in relational databases. Theor. Comput. Sci. **54**, 103–128 (1987)
29. Levene, M., Loizou, G.: Axiomatisation of functional dependencies in incomplete relations. Theor. Comput. Sci. **206**(1–2), 283–300 (1998)
30. Lien, Y.E.: On the equivalence of data models. J. ACM **29**(2), 333–363 (1982)
31. Link, S.: On multivalued dependencies in fixed and undetermined universes. In: FoIKS, number 3861 in Lecture Notes in Computer Science, pp. 258–277. Springer, Heidelberg (2006)
32. Link, S.: A strongly minimal axiomatisation of multivalued dependencies in incomplete database relations. In: DEXA Workshops, pp. 537–541 (2006)
33. Link, S.: On the implication of multivalued dependencies in partial database relations. Int. J. Found. Comput. Sci. **19**(3), 691–715 (2008)
34. Mendelzon, A.: On axiomatising multivalued dependencies in relational databases. J. ACM **26**(1), 37–44 (1979)
35. Paredaens, J., De Bra, P., Gyssens, M., Van Gucht, D.: The Structure of the Relational Database Model. Springer, Heidelberg (1989)
36. Parker, D., Delobel, C.: Algorithmic applications for a new result on multivalued dependencies. In: Proceedings of the 5th International Conference on Very Large Data Bases, pp. 67–74. IEEE Computer Society (1979)
37. Sagiv, Y.: An algorithm for inferring multivalued dependencies with an application to propositional logic. J. ACM **27**(2), 250–262 (1980)
38. Saxton, L., Tang, X.: Tree multivalued dependencies for XML datasets. In: Advances in Web-Age Information Management: 5th International Conference, number 3129 in Lecture Notes in Computer Science, pp. 357–367. Springer, Heidelberg (2004)

39. Sözat, M., Yazici, A.: A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations. Fuzzy Sets Syst. **117**(2), 161–181 (2001)
40. Tari, Z., Stokes, J., Spaccapietra, S.: Object normal forms and dependency constraints for object-oriented schemata. ACM Trans. Database Syst. **22**, 513–569 (1997)
41. Thalheim, B.: Dependencies in Relational Databases. Teubner-Verlag (1991)
42. Thalheim, B.: Conceptual treatment of multivalued dependencies. In: Conceptual Modeling ER 2003, 22nd International Conference on Conceptual Modeling, Chicago, IL, USA, October 13–16, 2003, Proceedings, number 2813 in Lecture Notes in Computer Science, pp. 363–375. Springer, Heidelberg (2003)
43. Vardi, M.: Inferring multivalued dependencies from functional and join dependencies. Acta Inf. **19**(2), 305–324 (1983)
44. Vardi, M.: Fundamentals of dependency theory. In: Börger, E. (ed.) Trends in Theoretical Computer Science, pp. 171–224. Computer Science Press (1987)
45. Vincent, M., Liu, J.: Multivalued dependencies in XML. In: Proceedings of the 20th British National Conference on Databases, number 2712 in Lecture Notes in Computer Science, pp. 4–18. Springer, Heidelberg (2003)
46. Vincent, M., Liu, J., Liu, C.: A redundancy free 4NF for XML. In: Proceedings of the 1st International XML Database Symposium, number 2824 in Lecture Notes in Computer Science, pp. 254–266. Springer, Heidelberg (2003)
47. Weddell, G.: Reasoning about functional dependencies generalized for semantic data models. ACM Trans. Database Syst. **17**(1), 32–64 (1992)
48. Wijsen, J.: Temporal FDs on complex objects. ACM Trans. Database Syst. **24**(1), 127–176 (1999)
49. Zaniolo, C.: Database relations with null values. J. Comput. Syst. Sci. **28**(1), 142–166 (1984)