Inverting Schema Mappings

Ronald Fagin
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120
fagin@almaden.ibm.com

ABSTRACT

A schema mapping is a specification that describes how data structured under one schema (the source schema) is to be transformed into data structured under a different schema (the target schema). Although the notion of an inverse of a schema mapping is important, the exact definition of an inverse mapping is somewhat elusive. This is because a schema mapping may associate many target instances with each source instance, and many source instances with each target instance. Based on the notion that the composition of a mapping and its inverse is the identity, we give a formal definition for what it means for a schema mapping $\mathcal{M}'$ to be an inverse of a schema mapping $\mathcal{M}$ for a class $\mathcal{S}$ of source instances. We call such an inverse an $\mathcal{S}$-*inverse*. A particular case of interest arises when $\mathcal{S}$ is the class of all instances, in which case an $\mathcal{S}$-inverse is a global inverse. We focus on the important and practical case of schema mappings defined by source-to-target tuple-generating dependencies, and uncover a rich theory. When $\mathcal{S}$ is defined by a set of dependencies with a finite chase, we show how to construct an $\mathcal{S}$-inverse when one exists. In particular, we show how to construct a global inverse when one exists. Given $\mathcal{M}$ and $\mathcal{M}'$, we show how to define the largest class $\mathcal{S}$ such that $\mathcal{M}'$ is an $\mathcal{S}$-inverse of $\mathcal{M}$.

Categories and Subject Descriptors: H.2.5 [Heterogeneous Databases]: Data translation; H.2.4 [Systems]: Relational databases

General Terms: Algorithms, theory

Keywords: Data exchange, chase, computational complexity, data integration, dependencies, inverse, metadata model management, schema mapping, second-order logic

1. INTRODUCTION

Data exchange is the problem of materializing an instance that adheres to a target schema, given an instance of a source schema and a schema mapping that specifies the relationship between the source and the target. This is a very old problem [13] that arises in many tasks where data must be transferred between independent applications that do not have the same data format.

Because of the extensive use of schema mappings, it has become important to develop a framework for managing schema mappings and other metadata, and operators for manipulating them. Bernstein [2] has introduced such a framework, called *model management*. Melnik et al. [12] have developed a semantics for model-management operators that allows applying the operators to executable mappings. One important schema mapping operator, at least in principle, is the inverse operator. What do we mean by an inverse of a schema mapping? This is a delicate question, since in spite of the traditional use of the name "mapping", a schema mapping is not simply a function that maps an instance of the source schema to an instance of the target schema. Instead, for each source instance, the schema mapping may associate many target instances. Furthermore, for each target instance, there may be many corresponding source instances.

As in [5, 6, 7], we study the relational case, where a schema is a sequence of distinct relational symbols. A *schema mapping* is a triple $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, where $\mathbf{S}$ (the *source schema*) and $\mathbf{T}$ (the *target schema*) are schemas with no relation symbols in common and Σ is a set of formulas of some logical formalism over $\langle \mathbf{S}, \mathbf{T} \rangle$. We say that Σ *defines* the schema $\mathcal{M}$. As in [5, 6, 7], our main focus is on the important and practical case of schema mappings where Σ is a finite set of source-to-target tuple-generating dependencies (which we shall call s-t tgds or simply tgds). These are formulas of the form $\forall \mathbf{x}(\varphi(\mathbf{x}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}))$, where $\varphi(\mathbf{x})$ is a conjunction of atomic formulas over $\mathbf{S}$, and where $\psi(\mathbf{x}, \mathbf{y})$ is a conjunction of atomic formulas over $\mathbf{T}$.[1] They have been used to formalize data exchange [5]. They have also been used in data integration scenarios under the name of GLAV (global-and-local-as-view) assertions [10].

There are other flavors of "schema mappings" that have been studied in the literature, such as view definitions, where there is a unique target instance associated with each source instance. In such cases, a schema mapping is a function in the classical sense, and so it is quite clear and unambiguous as to what an inverse mapping is. An example of such work is Hull's seminal research on information capacity of relational database schemas [9]. Although our schema mappings are not actually functions, they have the advantage

PODS'06, June 26–28, 2006, Chicago, Illinois, USA.

[1] There is also a safety condition, which says that every variable in $\mathbf{x}$ appears in φ. However, not all of the variables in $\mathbf{x}$ need to appear in ψ.

of being simpler and more flexible. In fact, LAV mappings, which have been widely used in data integration, are special cases of schema mappings defined by s-t tgds, where the left-hand side of each tgd is a single atomic formula rather than a conjunction of atomic formulas.

Let us now consider how to define the inverse in our context, where schema mappings are not actually functions. Let us associate with the schema mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ the set S_{12} of ordered pairs $\langle I, J\rangle$ such that I is a source instance, J is a target instance, and the pair $\langle I, J\rangle$ satisfy Σ_{12} (written $\langle I, J\rangle \models \Sigma_{12}$). Perhaps the most natural definition of the inverse of the schema mapping $\mathcal{M}_{12}$ would be a schema mapping $\mathcal{M}_{21}$ that is associated with the set $S_{21} = \{\langle J, I\rangle : \langle I, J\rangle \in S_{12}\}$. This reflects the standard algebraic definition of an inverse, and is the definition that Melnik [11] and Melnik et al. [12] give for the inverse. In those papers, this definition was intended for a generic model management context, where mappings can be defined in a variety of ways, including as view definitions, relational algebra expressions, etc. However, this definition does not make sense in our context. This is because S_{12}, by being associated with a schema mapping defined by s-t tgds, is automatically "closed down on the left and closed up on the right". This means that if $\langle I, J\rangle \in S_{12}$ and if $I' \subseteq I$ (that is, I' is a subinstance of I) and $J \subseteq J'$, then $\langle I', J'\rangle \in S_{12}$.[2] However, instead of being closed down on the left and closed up on the right, S_{21} is closed up on the left and closed down on the right. This is inconsistent with a schema mapping that is defined by a set of s-t tgds.

Our notion of an inverse of a schema mapping is based on another algebraic property of inverses, that the composition of a function with its inverse is the identity mapping. In our context, the identity mapping is defined by tgds that "copy" the source instance to the target instance. Our definition of inverse says that the schema mapping $\mathcal{M}_{21}$ is an inverse of the schema mapping $\mathcal{M}_{12}$ for the class $\mathcal{S}$ of source instances if the schema mapping defined by their composition is equivalent on $\mathcal{S}$ to the identity mapping. We refer then to $\mathcal{M}_{21}$ as an *$\mathcal{S}$-inverse* of $\mathcal{M}_{12}$ When $\mathcal{S}$ is the class of all source instances, then $\mathcal{M}_{21}$ is said to be a *global inverse* of $\mathcal{M}_{12}$. When $\mathcal{S}$ is a singleton set containing only the source instance I, then $\mathcal{M}_{21}$ is said to be a *local inverse*, or simply an *inverse*, of $\mathcal{M}_{12}$ for I. Note that our definition of what it means for $\mathcal{M}_{21}$ to be an inverse of $\mathcal{M}_{12}$ corresponds exactly to what we would like an inverse mapping to do in data exchange: if after applying $\mathcal{M}_{12}$, we then apply $\mathcal{M}_{21}$, the resulting effect of $\mathcal{M}_{21}$ is to "undo" the effect of $\mathcal{M}_{12}$. Fortunately, because of work by Fagin et al. [7], we now understand very well the composition of schema mappings, and so we are in a good position to study our notion of inverse. This paper is the first step in exploring the very rich theory that arises.

If $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a schema mapping, I is a source instance, and J is a target instance, then J is a *solution* for I if $\langle I, J\rangle \models \Sigma_{12}$. A simple necessary condition for $\mathcal{M}_{12}$ to have a global inverse is the *unique solutions property*, which says that no two distinct source instances have the same set of solutions. For a fixed choice of $\mathcal{M}_{12}$, let f be the set-valued function where $f(I)$ is the set of solutions for the source instance I. The unique solutions property is equivalent to the condition that f be one-to-one. The fact that this condition is necessary for there to be a global inverse is analogous to the standard algebraic condition that an invertible function be one-to-one. We show that surprisingly and pleasingly, in the important special case of LAV schema mappings, the unique solutions property is not only necessary for $\mathcal{M}_{12}$ to have a global inverse but also sufficient.

Assume that $\mathcal{M}$ is a schema mapping defined by a finite set of s-t tgds, and I is a source instance. We derive a canonical local inverse, which is a schema mapping defined by a finite set of s-t tgds that is an inverse of $\mathcal{M}$ for I if there is any such inverse. If $\mathcal{S}$ is a class of source instances defined by a set of tgds and egds that always have a finite chase, then we derive a *canonical $\mathcal{S}$-inverse*, which is a schema mapping defined by a finite set of s-t tgds that is an $\mathcal{S}$-inverse of $\mathcal{M}$ if there is any such $\mathcal{S}$-inverse. When $\mathcal{S}$ is the class of all source instances, we refer to the canonical $\mathcal{S}$-inverse as the *canonical global inverse*. On the face of it, the canonical local inverse seems to be of theoretical interest only: after all, we typically care only about an inverse that "works" for a large class, not for a single instance. However, it turns out that the canonical local inverse plays a key role in the proof of correctness of the canonical $\mathcal{S}$-inverse.

Our canonical inverses are each defined by finite sets of *full* tgds (those with no existential quantifiers). This is not an accident: we show that if $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are schema mappings that are each defined by a finite set of s-t tgds, $\mathcal{S}$ is a class of source instances, and $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$, then there is a schema mapping defined by a finite set of *full* s-t tgds and that is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$.

It is folk wisdom that an inverse can be obtained by simply "reversing the arrows" in a tgd. We show that even a weak form of this folk wisdom is false. Instead, our canonical inverses are obtained by a slightly more complicated but still very natural procedure.

Since a local inverse may be quite tailored to a particular instance, it is natural to ask whether it is possible for a schema mapping defined by a finite set of s-t tgds to have an inverse for every source instance yet not have a global inverse. We show that this can indeed happen.

Given schema mappings $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ that are each defined by a finite set of s-t tgds, an analyst might want to investigate under what conditions $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$. (We give an example later, where $\mathcal{M}_{12}$ does a projection and $\mathcal{M}_{21}$ joins the projections.) If we hold $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ fixed, then we show that the problem of deciding whether $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I is in the complexity class NP. It therefore follows from Fagin's Theorem [3] that the class $\mathcal{S}$ of source instances such that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for precisely the class $\mathcal{S}$ can be defined by a formula Γ in existential second-order logic. Remarkably, we are able to obtain such a formula Γ by a purely syntactical transformation of the formula that defines the composition of the schema mappings. Furthermore, when $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are defined by *full* s-t tgds, this formula is first-order.

Finally, we obtain other complexity results about deciding local or global invertibility.

Missing proofs are in the full version of the paper, currently on the author's website at:
`http://www.almaden.ibm.com/cs/people/fagin/inverse.pdf`

1.1 Applications of inverse mappings

There are potentially a number of applications for inverse

[2]This is why a schema mapping may associate many target instances with each source instance, and many source instances with each target instance.

mappings, especially in schema evolution. For example, assume that data has been migrated from one schema to another with a schema mapping $\mathcal{M}$. At some point, we might decide to "roll back" to the original schema, and so we might want to apply an inverse schema mapping $\mathcal{M}^{-1}$. In fact, if we think this scenario is probable, we might deliberately choose a schema mapping $\mathcal{M}$ that has an inverse $\mathcal{M}^{-1}$.

As a more intricate example, assume that there are two different schema mappings from schema $\mathbf{S_1}$: the schema mapping $\mathcal{M}_1$ from schema $\mathbf{S_1}$ to schema $\mathbf{T_1}$, and the schema mapping $\mathcal{M}'_1$ from $\mathbf{S_1}$ to $\mathbf{S'_1}$. Assume that there is also a schema mapping $\mathcal{M}_2$ from $\mathbf{T_1}$ to $\mathbf{T'_1}$. If there is an "inverse schema mapping" ${\mathcal{M}'_1}^{-1}$ of $\mathcal{M}'_1$, then these schema mappings can be composed to give a schema mapping directly from $\mathbf{S'_1}$ to $\mathbf{T'_1}$, by taking the composition of the schema mapping ${\mathcal{M}'_1}^{-1}$ (from $\mathbf{S'_1}$ to $\mathbf{S_1}$) with the schema mapping $\mathcal{M}_1$ (from $\mathbf{S_1}$ to $\mathbf{T_1}$) and composing the result with the schema mapping $\mathcal{M}_2$ (from $\mathbf{T_1}$ to $\mathbf{T'_1}$).

2. BACKGROUND

We now review basic concepts from data exchange.

A *schema* is a finite sequence $\mathbf{R} = \langle \mathtt{R}_1, \ldots, \mathtt{R}_k \rangle$ of distinct relation symbols, each of a fixed arity. An *instance* I *(over the schema* $\mathbf{R}$) is a sequence $\langle \mathtt{R}_1^I, \ldots, \mathtt{R}_k^I \rangle$ such that each $\mathtt{R}_i^I$ is a finite relation of the same arity as $\mathtt{R}_i$. We call $\mathtt{R}_i^I$ the $\mathtt{R}_i$*-relation of* I. We shall often abuse the notation and use $\mathtt{R}_i$ to denote both the relation symbol and the relation $\mathtt{R}_i^I$ that interprets it.

Let $\mathbf{S} = \langle \mathtt{S}_1, \ldots, \mathtt{S}_n \rangle$ and $\mathbf{T} = \langle \mathtt{T}_1, \ldots, \mathtt{T}_m \rangle$ be two schemas with no relation symbols in common. We write $\langle \mathbf{S}, \mathbf{T} \rangle$ to denote the schema that is the result of concatenating the members of $\mathbf{S}$ with the members of $\mathbf{T}$. If I is an instance over $\mathbf{S}$ and J is an instance over $\mathbf{T}$, then we write $\langle I, J \rangle$ for the instance K over the schema $\langle \mathbf{S}, \mathbf{T} \rangle$ such that $\mathtt{S}_i^K = \mathtt{S}_i^I$ and $\mathtt{T}_j^K = \mathtt{T}_j^J$, for $1 \leq i \leq n$ and $1 \leq j \leq m$.

If K is an instance and σ is a formula in some logical formalism, then we write $K \models \sigma$ to mean that K satisfies σ. If Σ is a set of formulas, then we write $K \models \Sigma$ to mean that $K \models \sigma$ for every formula $\sigma \in \Sigma$.

We will often drop the universal quantifiers in front of a tgd, and implicitly assume such quantification. However, we will write down all existential quantifiers.

Given a tuple $(t_1, \ldots, t_r)$ occurring in a relation $\mathtt{R}$, we denote by $\mathtt{R}(t_1, \ldots, t_r)$ the association between $(t_1, \ldots, t_r)$ and $\mathtt{R}$, and call it a *fact*. We will identify an instance with its set of facts. We call each t_i in the tuple $(t_1, \ldots, t_r)$ a *value*. We denote by $\underline{\mathsf{Const}}$ the set of all values that appear in source instances (instances of the schema $\mathbf{S}$) and we call them *constants*. In addition, we assume an infinite set $\underline{\mathsf{Var}}$ of values, which we call *nulls*, such that $\underline{\mathsf{Var}} \cap \underline{\mathsf{Const}} = \emptyset$.

If K is an instance with values in $\underline{\mathsf{Const}} \cup \underline{\mathsf{Var}}$, then $\underline{\mathsf{Var}}(K)$ denotes the set of nulls appearing in relations in K. Let K_1 and K_2 be two instances over the same schema with values in $\underline{\mathsf{Const}} \cup \underline{\mathsf{Var}}$. A *homomorphism* $h : K_1 \to K_2$ is a mapping from $\underline{\mathsf{Const}} \cup \underline{\mathsf{Var}}(K_1)$ to $\underline{\mathsf{Const}} \cup \underline{\mathsf{Var}}(K_2)$ such that: (1) $h(c) = c$, for every $c \in \underline{\mathsf{Const}}$; and (2) for every fact $\mathtt{R}(\mathbf{t})$ of K_1, we have that $\mathtt{R}(h(\mathbf{t}))$ is a fact of K_2 (where, if $\mathbf{t} = (t_1, \ldots, t_s)$, then $h(\mathbf{t}) = (h(t_1), \ldots, h(t_s))$).

Consider a schema mapping $(\mathbf{S}, \mathbf{T}, \Sigma)$, as defined in the introduction. Recall that if I is a source instance, and J is a target instance, then J is a *solution* for I if $\langle I, J \rangle \models \Sigma$. If I is a source instance, then a *universal solution for* I is a solution J for I such that for every solution J' for I, there exists a homomorphism $h : J \to J'$. When Σ is a finite set of s-t tgds, and I is a source instance, then there is always a universal solution for I [5].

Let $\mathcal{M}_{12} = (\mathbf{S_1},\ \mathbf{S_2},\ \Sigma_{12})$ and $\mathcal{M}_{23} = (\mathbf{S_2},\ \mathbf{S_3},\ \Sigma_{23})$ be two schema mappings such that the schemas $\mathbf{S_1}, \mathbf{S_2}, \mathbf{S_3}$ have no relation symbol in common pairwise. The *composition formula* [7], denoted by $\Sigma_{12} \circ \Sigma_{23}$, has the semantics that if I is an instance of $\mathbf{S_1}$ and J is an instance of $\mathbf{S_3}$, then $\langle I, J \rangle \models \Sigma_{12} \circ \Sigma_{23}$ precisely if there is an instance J' of $\mathbf{S_2}$ such that $\langle I, J' \rangle \models \Sigma_{12}$ and $\langle J', J \rangle \models \Sigma_{23}$. It is proven in [7] that when Σ_{12} and Σ_{23} are finite sets of s-t tgds, then the composition formula is given by a *second-order tgd (SO tgd)*. We give the definition of SO tgds later (Definition 10.1). We now give an example (from [7]) of an SO tgd that defines the composition formula.

EXAMPLE 2.1. Consider the following three schemas $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$. Schema $\mathbf{S}_1$ consists of a single unary relation symbol $\mathtt{Emp}$ of employees. Schema $\mathbf{S}_2$ consists of a single binary relation symbol $\mathtt{Mgr}_1$, that associates each employee with a manager. Schema $\mathbf{S}_3$ consists of a similar binary relation symbol $\mathtt{Mgr}$, that is intended to provide a copy of $\mathtt{Mgr}_1$, and an additional unary relation symbol $\mathtt{SelfMgr}$, that is intended to store employees who are their own manager. Consider now the schema mappings $\mathcal{M}_{12} = (\mathbf{S}_1,\ \mathbf{S}_2, \Sigma_{12})$ and $\mathcal{M}_{23} = (\mathbf{S}_2, \mathbf{S}_3, \Sigma_{23})$, where Σ_{12} consists of the tgd $\forall e\, (\mathtt{Emp}(e) \to \exists m \mathtt{Mgr}_1(e, m))$, and Σ_{23} consists of the two tgds $\forall e \forall m\, (\mathtt{Mgr}_1(e, m) \to \mathtt{Mgr}(e, m))$ and $\forall e (\mathtt{Mgr}_1(e, e) \to \mathtt{SelfMgr}(e))$. Then the composition formula $\Sigma_{12} \circ \Sigma_{23}$ is defined by the following second-order tgd:

$$\exists f (\forall e (\mathtt{Emp}(e) \to \mathtt{Mgr}(e, f(e))) \land \qquad (1)$$
$$\forall e (\mathtt{Emp}(e) \land (e = f(e)) \to \mathtt{SelfMgr}(e))). \quad \square$$

3. WHAT IS AN INVERSE MAPPING?

Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2},\ \Sigma_{12})$ is a schema mapping. For each relation symbol $\mathtt{R}$ of $\mathbf{S_1}$, let $\widehat{\mathtt{R}}$ be a new relation symbol (different from any relation symbol in $\mathbf{S_1}$ or $\mathbf{S_2}$) of the same arity as $\mathtt{R}$. Define $\widehat{\mathbf{S_1}}$ to be $\{\widehat{\mathtt{R}} : \mathtt{R} \in \mathbf{S_1}\}$. Thus, $\widehat{\mathbf{S_1}}$ is a schema disjoint from $\mathbf{S_1}$ and $\mathbf{S_2}$ that can be thought of as a copy of $\mathbf{S_1}$. If I is an instance of $\mathbf{S_1}$, define $\widehat{I}$ to be the corresponding instance of $\widehat{\mathbf{S_1}}$. Thus, $\widehat{\mathtt{R}}^{\widehat{I}} = \mathtt{R}^I$ for every $\mathtt{R}$ in $\mathbf{S_1}$.

Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}},\ \Sigma_{21})$ be a schema mapping, where the source schema $\mathbf{S_2}$ is the target schema of $\mathcal{M}_{12}$, where the target schema is $\widehat{\mathbf{S_1}}$, and where Σ_{21} is a finite set of s-t tgds (with source $\mathbf{S_2}$ and target $\widehat{\mathbf{S_1}}$). The issue we are concerned with is: what does it mean for $\mathcal{M}_{21}$ to be an inverse of $\mathcal{M}_{12}$, and what can we say about such inverse mappings? We are most interested in the case where Σ_{12} and Σ_{21} are finite sets of s-t tgds. We now introduce an example that we shall use as a running example to demonstrate some of the issues that arise.

EXAMPLE 3.1. Let $\mathbf{S_1}$ consist of the ternary relation symbol $\mathtt{EDL}$ ("Employee-Department-Location"). Let $\mathbf{S_2}$ consist of the binary relation symbol $\mathtt{ED}$ ("Employee-Department") and the binary relation symbol $\mathtt{DL}$ ("Department-Location"). Let Σ_{12} consist of the s-t tgd $\mathtt{EDL}(x, y, z) \to \mathtt{ED}(x, y) \land \mathtt{DL}(y, z)$, that corresponds to projecting $\mathtt{EDL}$ onto $\mathtt{ED}$ and $\mathtt{DL}$. Let Σ_{21} consist of the s-t tgd $(\mathtt{ED}(x, y) \land \mathtt{DL}(y, z)) \to$

$\widehat{\texttt{EDL}}(x, y, z)$, where the source schema is $\mathbf{S_2}$ and the target schema is $\widehat{\mathbf{S_1}}$, that corresponds to taking the join of the projections. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$.

Let Γ be the multivalued dependency[3]

$$\texttt{EDL}(x, y, z') \wedge \texttt{EDL}(x', y, z)) \rightarrow \texttt{EDL}(x, y, z). \quad (2)$$

It is known [4] that if we project the EDL relation onto ED and DL and then join the resulting projections, we obtain the original EDL relation precisely if the multivalued dependency Γ holds. We want our definition of inverse to have the property that the schema mapping $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for precisely those source instances I that satisfy Γ. □

Let us now define some preliminary notions that will allow us to define what it means for the mapping $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ to be an $\mathcal{S}$-inverse of the mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. (In Example 3.1, the class $\mathcal{S}$ would consist of those source instances that satisfy Γ.) Define Σ_{Id} (where Id stands for "identity") to consist of the tgds $\texttt{R}(x_1, \ldots, x_k) \rightarrow \widehat{\texttt{R}}(x_1, \ldots, x_k)$, where $x_1, \ldots, x_k$ are distinct variables, when R is a k-ary relation symbol of $\mathbf{S_1}$. Define the *identity mapping* to be $\mathcal{M}_{Id} = (\mathbf{S_1}, \widehat{\mathbf{S_1}}, \Sigma_{Id})$. Note that J is a solution for I under the identity mapping if and only if $\widehat{I} \subseteq J$. The reason we have $\widehat{I} \subseteq J$ rather than simply $\widehat{I} = J$ is that Σ_{Id} is a set of s-t tgds, and hence whenever J is a solution, then so is every J' with $J \subseteq J'$. Let us say that two schema mappings with the same source schema and the same target schema are *equivalent on* I if they have the same solutions for I.

We are now ready to define the notion of inverse. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be schema mappings. Let σ be the composition formula $\Sigma_{12} \circ \Sigma_{21}$ of $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$, and let $\mathcal{M}_{11} = (\mathbf{S_1}, \widehat{\mathbf{S_1}}, \sigma)$. Let I be an instance of $\mathbf{S_1}$. Let us say that $\mathcal{M}_{21}$ *is an inverse of* $\mathcal{M}_{12}$ *for* I if $\mathcal{M}_{11}$ and the identity mapping $\mathcal{M}_{Id}$ are equivalent on I. Thus, $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I precisely if for every J,

$$\langle I, J \rangle \models \sigma \text{ if and only if } \widehat{I} \subseteq J. \quad (3)$$

If $\mathcal{S}$ is a class of source instances, then we say that $\mathcal{M}_{21}$ *is an* $\mathcal{S}$*-inverse of* $\mathcal{M}_{12}$ if $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I, for each I in $\mathcal{S}$. A particularly important case arises when $\mathcal{S}$ is the class of all source instances. In that case, we say that $\mathcal{M}_{21}$ is a *global inverse* of $\mathcal{M}_{12}$.

Example 3.2. Let us return to Example 3.1. We said there that we want our definition of inverse to have the property that the schema mapping $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for precisely those source instances I that satisfy Γ. We now show that satisfying Γ is a sufficient condition for $\mathcal{M}_{21}$ to be an inverse of $\mathcal{M}_{12}$. In Example 10.7, we shall show that Γ is also a necessary condition.

If we apply the composition algorithm of [7], we find that the composition formula $\Sigma_{12} \circ \Sigma_{21}$, which we denote by σ, is

$$\texttt{EDL}(x, y, z') \wedge \texttt{EDL}(x', y, z)) \rightarrow \widehat{\texttt{EDL}}(x, y, z). \quad (4)$$

Let I be a source instance of $\mathbf{S_1}$ satisfying Γ. We must show that (3) holds. Assume first that $\langle I, J \rangle \models \sigma$; we must show that $\widehat{I} \subseteq J$. Now Σ_{Id} consists of the tgd $\texttt{EDL}(x, y, z) \rightarrow \widehat{\texttt{EDL}}(x, y, z)$. It is clear that σ logically implies Σ_{Id} (we let the roles of x' and z' be played by x and z, respectively). Therefore, since $\langle I, J \rangle \models \sigma$, it follows that $\langle I, J \rangle \models \Sigma_{Id}$. So $\widehat{I} \subseteq J$, as desired.

Assume now that $\widehat{I} \subseteq J$; we must show that $\langle I, J \rangle \models \sigma$. Thus, we must show that if $\texttt{EDL}(x, y, z')$ and $\texttt{EDL}(x', y, z)$ hold in I, then $\widehat{\texttt{EDL}}(x, y, z)$ holds in J. So assume that $\texttt{EDL}(x, y, z')$ and $\texttt{EDL}(x', y, z)$ hold in I. Since $I \models \Gamma$, it follows that $\texttt{EDL}(x, y, z)$ holds in I. Since $\widehat{I} \subseteq J$, it follows that $\widehat{\texttt{EDL}}(x, y, z)$ holds in J, as desired.

Note the unexpected similarity of the composition formula (4) and Γ (the multivalued dependency (2)). We shall explain this surprising connection between the composition formula and Γ later (in Example 10.7). □

The next example shows that there need not be a unique inverse. Therefore, we refer to "*an* inverse mapping" rather than "*the* inverse mapping".

Example 3.3. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, where $\mathbf{S_1}$ consists of the unary relation symbol R, where $\mathbf{S_2}$ consists of the binary relation symbol S, and where Σ_{12} consists of the tgd $\texttt{R}(x) \rightarrow \texttt{S}(x, x)$. Let Σ_{21} consist of the tgd $\texttt{S}(x, y) \rightarrow \widehat{\texttt{R}}(x)$, and let Σ'_{21} consist of the tgd $\texttt{S}(x, y) \rightarrow \widehat{\texttt{R}}(y)$. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, and let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. In both cases (for $\mathcal{M}_{21}$ and for $\mathcal{M}'_{21}$), the composition formula is $\texttt{R}(x) \rightarrow \widehat{\texttt{R}}(x)$, which defines the identity mapping. So both $\mathcal{M}_{21}$ and $\mathcal{M}'_{21}$ are global inverses of $\mathcal{M}_{12}$. □

4. THE UNIQUE SOLUTIONS PROPERTY

Unlike the rest of this paper, in this section we do not restrict our attention to schema mappings $(\mathbf{S}, \mathbf{T}, \Sigma)$ where Σ is a finite set of s-t tgds. Instead, we allow Σ to be an arbitrary constraint between source and target instances. Our only requirement is that the satisfaction relation between formulas and instances be preserved under isomorphism. This means that if $\langle I, J \rangle \models \Sigma$, and if $\langle I', J' \rangle$ is isomorphic to $\langle I, J \rangle$, then $\langle I', J' \rangle \models \Sigma$. This is a mild condition that is true of all standard logical formalisms, such as first-order logic, second-order logic, fixed-point logics, and infinitary logics.

Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a schema mapping, and let I be a source instance. Intuitively, as far as $\mathbf{S_2}$ is concerned, the only information about I is the set of solutions for I, that is, the set of target instances J such that $\langle I, J \rangle \models \Sigma_{12}$. Therefore, we would expect that if $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for two distinct source instances I_1 and I_2, then I_1 and I_2 would have different sets of solutions. Otherwise, intuitively, there would not be enough information to allow $\mathcal{M}_{21}$ to reconstruct I_1 after applying $\mathcal{M}_{12}$. The next theorem says that this intuition is correct.

Theorem 4.1. *Let $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ be schema mappings. Assume that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for distinct instances I_1 and I_2. Then the set of solutions for I_1 under $\mathcal{M}_{12}$ is different from the set of solutions for I_2 under $\mathcal{M}_{12}$.*

As a corollary of Theorem 4.1, we obtain a necessary condition for $\mathcal{M}_{12}$ to have an inverse for a fixed source instance. (The proof depends on our assumption of preservation under isomorphism.)

Corollary 4.2. *Let $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ be schema mappings, and let I_1 and I_2 be distinct but isomorphic source instances.*

[3]Note that Γ is not an s-t tgd, since the left-hand side and right-hand side use the same relation symbol EDL. Of course, Γ is a tgd in the classical sense of [1].

Assume that there is an inverse of $\mathcal{M}_{12}$ for I_1. Then the set of solutions for I_1 under $\mathcal{M}_{12}$ is different from the set of solutions for I_2 under $\mathcal{M}_{12}$.

The next corollary, which we shall find quite useful later, applies to schema mappings defined by tgds, and makes use of the fundamental notion of the chase. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, where Σ_{12} is a finite set of s-t tgds. Assume that I is an instance of $\mathbf{S_1}$. If the result of chasing $\langle I, \emptyset\rangle$ with Σ_{12} is $\langle I, J\rangle$, then we define $chase_{12}(I)$ to be J.[4] We may say loosely that J is the result of chasing I with Σ_{12}.

COROLLARY 4.3. *Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a schema mapping, where Σ_{12} is a finite set of s-t tgds. If $\mathcal{M}_{12}$ has an inverse for I (not necessarily defined by s-t tgds), then every value that appears in a tuple of I appears in a tuple of $chase_{12}(I)$.*

The next proposition is an interesting application of Theorem 4.1.

PROPOSITION 4.4. *There is a schema mapping defined by a finite set of full s-t tgds that has an inverse for every source instance with a schema mapping defined by a finite set of s-t tgds, but has no global inverse.*

PROOF. Let $\mathbf{S_1}$ consist of the unary relation symbols $\mathtt{P}$ and $\mathtt{Q}$, and let $\mathbf{S_2}$ consist of the binary relation symbol $\mathtt{R}$ and the unary relation symbol $\mathtt{S}$. Let $\Sigma_{12} = \{\mathtt{P}(x) \wedge \mathtt{Q}(y) \rightarrow \mathtt{R}(x, y), \mathtt{P}(x) \rightarrow \mathtt{S}(x), \mathtt{Q}(x) \rightarrow \mathtt{S}(x)\}$. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$.

We now show that for every source instance I, the schema mapping $\mathcal{M}_{12}$ has an inverse that is defined by a finite set of s-t tgds. There are three cases:

- $\mathtt{P}^I$ is empty. Then an inverse is $\mathtt{S}(x) \rightarrow \widehat{\mathtt{Q}}(x)$
- $\mathtt{Q}^I$ is empty. Then an inverse is $\mathtt{S}(x) \rightarrow \widehat{\mathtt{P}}(x)$.
- Neither $\mathtt{P}^I$ nor $\mathtt{Q}^I$ is empty. Then an inverse is $\mathtt{R}(x, y) \rightarrow \widehat{\mathtt{P}}(x) \wedge \widehat{\mathtt{Q}}(y)$.

Now we will show that $\mathcal{M}_{12}$ does not have a global inverse. Let $I_1 = \{P(0)\}$, and let $I_2 = \{Q(0)\}$. Then the set of solutions for I_1 under $\mathcal{M}_{12}$ equals the set of solutions for I_2 under $\mathcal{M}_{12}$ (both equal the set of target instances J that contain $\{S(0)\}$). It then follows from Theorem 4.1 that $\mathcal{M}_{12}$ does not have a global inverse. □

We now give a simple example of the use of Corollary 4.2.

EXAMPLE 4.5. Let $\mathbf{S_1}$ consist of the unary relation symbols $\mathtt{R}$ and $\mathtt{R}'$, let $\mathbf{S_2}$ consist of the unary relation symbol $\mathtt{S}$, and let $\Sigma_{12} = \{\mathtt{R}(x) \rightarrow \mathtt{S}(x), \mathtt{R}'(x) \rightarrow \mathtt{S}(x)\}$. Assume that the facts of I_1 are precisely $\mathtt{R}(0)$ and $\mathtt{R}'(1)$; we now show that $\mathcal{M}_{12}$ does not have an inverse for I_1. Let I_2 be the source instance whose facts are precisely $\mathtt{R}(1)$ and $\mathtt{R}'(0)$. Let J be the target instance whose facts are precisely $\mathtt{S}(0)$ and $\mathtt{S}(1)$. Then the solutions under Σ_{12} for I_1 are exactly those J' where $J \subseteq J'$. But these are also exactly the solutions for I_2. Since I_1 and I_2 are distinct isomorphic source instances with the same set of solutions, it follows from Corollary 4.2 that $\mathcal{M}_{12}$ does not have an inverse for I_1. □

Let us say that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ has the *unique solutions property* if whenever I_1 and I_2 are distinct source instances, then the set of solutions for I_1 is distinct from the set of solutions for I_2. In the case where Σ_{12} is a finite set of s-t tgds, it follows from results of [5] that I_1 and I_2 have the same set of solutions if and only if they share a universal solution. Therefore, when Σ_{12} is a finite set of tgds, the unique solutions property is equivalent to the *unique universal solutions property*, which says that whenever I_1 and I_2 are distinct source instances, then no universal solution for I_1 is a universal solution for I_2.

Theorem 4.1 implies that the unique solutions property is a necessary condition for global invertibility. Recall that a *LAV* (local-as-view) schema mapping is a schema mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ where Σ_{12} is a finite set of s-t tgds all with a singleton left-hand side. The next theorem says that for LAV schema mappings, the unique solutions property is not only necessary for global invertibility but also sufficient. This shows robustness of our notion of inverse, since (at least in the case of LAV mappings), our notion of global invertibility is equivalent to the unique solutions property, which is another natural notion.

THEOREM 4.6. *A LAV schema mapping has a global inverse if and only if it has the unique solutions property.*

The schema mapping that is a global inverse in our proof of Theorem 4.6 is rather complex (it is not defined in terms of tgds). For the rest of this paper, we shall consider only "practical" schema mappings—specifically, schema mappings $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ that are each defined by a finite set of s-t tgds.

5. CHARACTERIZING INVERTIBILITY

In this section, we give useful characterizations, in terms of the chase, of invertibility.

For the next theorem, we define $chase_{21}$ based on $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ just as we defined $chase_{12}$ based on $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$.

THEOREM 5.1. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of s-t tgds. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I if and only if $\langle I, \widehat{I}\rangle \models \Sigma_{12} \circ \Sigma_{21}$ and $\widehat{I} \subseteq chase_{21}(chase_{12}(I))$.*

As a corollary, we obtain a particularly simple characterization when Σ_{12} and Σ_{21} consist of *full* tgds.

COROLLARY 5.2. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of full s-t tgds. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I if and only if $\widehat{I} = chase_{21}(chase_{12}(I))$.*

The next result[5] gives a version of Corollary 5.2 that holds even when the tgds are not full. Two instances I_1 and I_2 are *homomorphically equivalent* if there is a homomorphism from I_1 into I_2 and a homomorphism from I_2 into I_1.

COROLLARY 5.3. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of s-t tgds. If $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I, then $\widehat{I}$ and $chase_{21}(chase_{12}(I))$ are homomorphically equivalent.*

[4]For definiteness, we use the version of the chase as defined in [7], although it does not really matter.

[5]This result is due to Lucian Popa.

The next theorem implies the falsity of the converse of Corollary 5.3, that if $\widehat{I}$ and $chase_{21}(chase_{12}(I))$ are homomorphically equivalent, then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I. In fact, the negative result we prove is even stronger: it says that even if $\widehat{I}$ and $chase_{21}(chase_{12}(I))$ are homomorphically equivalent for *every* I, there can be an I such that $\mathcal{M}_{21}$ is not inverse of $\mathcal{M}_{12}$ for I.

THEOREM 5.4. *There are schema mappings $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, where Σ_{12} and Σ_{21} are finite sets of s-t tgds, such that $\widehat{I}$ and $chase_{21}(chase_{12}(I))$ are homomorphically equivalent for every instance I of $\mathbf{S_1}$, but $\mathcal{M}_{21}$ is not an inverse of $\mathcal{M}_{12}$ for some instance I of $\mathbf{S_1}$.*

6. THE CANONICAL LOCAL INVERSE

Let $\mathcal{M}$ be a schema mapping defined by a finite set of s-t tgds, and let I be a source instance. In this section, we give a schema mapping that is guaranteed to be an inverse of $\mathcal{M}$ for I if there is any inverse at all that is defined by a finite set of s-t tgds.

We begin with a definition. Assume that I and J are instances (of different schemas) where every value that appears in a tuple of I also appears in a tuple of J. Define $\beta_{J,I}$ to be the full tgd where the left-hand side is the conjunction of the facts of J, and the right-hand side is the conjunction of the facts of I (we are treating the values in J as universally quantified variables in the tgd).

Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a schema mapping where Σ_{12} is a finite set of s-t tgds. Assume that there is a schema mapping $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ that is an inverse of $\mathcal{M}_{12}$ for I, where Σ_{21} is a finite set of s-t tgds. Let $J^* = chase_{12}(I)$. It follows from Corollary 4.3 that every value that appears in a tuple of I (and hence in a tuple of $\widehat{I}$) appears in a tuple of J^*, and so $\beta_{J^*,\widehat{I}}$ is a full tgd. Define the *canonical local inverse of* $\mathcal{M}_{12}$ *for* I to be $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \left\{\beta_{J^*,\widehat{I}}\right\})$ (we shall show that it is actually a local inverse). We call $\mathcal{M}_{21}$ the *most general s-t inverse of* $\mathcal{M}_{12}$ *for* I if Σ'_{21} logically implies Σ_{21} for every inverse $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ of $\mathcal{M}_{12}$ for I where Σ'_{21} is a finite set of s-t tgds.

THEOREM 6.1. *Let $\mathcal{M}$ be a schema mapping defined by a finite set of s-t tgds, and let I be a source instance. Assume that $\mathcal{M}$ has an inverse for I that is defined by a finite set of s-t tgds. Then the canonical local inverse of $\mathcal{M}$ for I is indeed an inverse of $\mathcal{M}$ for I, and in fact the most general s-t inverse of $\mathcal{M}$ for I.*

Of course, we are much more interested in an $\mathcal{S}$-inverse for a large class $\mathcal{S}$, rather than an inverse for a single instance I. However, the canonical local inverse is important as a tool in proving correctness of the canonical $\mathcal{S}$-inverse (including the canonical global inverse) in the next section. In fact, even the fact that the canonical local inverse is most general is needed for the proof.

7. THE CANONICAL $\mathcal{S}$-INVERSE

Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a schema mapping, where Σ_{12} is a finite set of s-t tgds. In this section, we shall consider certain classes $\mathcal{S}$ of source instances, and show how to define a *canonical $\mathcal{S}$-inverse of* $\mathcal{M}_{12}$, which is a schema mapping defined by a finite set of s-t tgds that is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$ if there is any such $\mathcal{S}$-inverse. We shall consider certain sets Γ of constraints on source instances, and let $\mathcal{S}$ be the class of source instances that satisfy Γ. When Γ is the empty set, then $\mathcal{S}$ is the class of all source instances, and so an $\mathcal{S}$-inverse is a global inverse. In this case, we shall refer to the canonical $\mathcal{S}$-inverse as the *canonical global inverse*.

Let us say that a set Γ of tgds and egds (all on the source schema) is *finitely chasable* if for every (finite) source instance I, some result of chasing I with Γ is a (finite) instance, or else some chase of I with Γ fails (by trying to equate two distinct values in I). It is not hard to see that Γ is finitely chasable if and only if for every (finite) source instance I, some result of chasing I with Γ is a (finite) instance, where we allow values in I to be equated in the chase. It follows from results in [5] that when Γ is the disjoint union of a set of egds with a weakly acyclic set of tgds (as defined in [5]), then Γ is finitely chasable. We now give a simple example where the converse fails.

EXAMPLE 7.1. Let Γ' consist of the single tgd $\mathtt{R}(x, y) \rightarrow \exists z \mathtt{R}(y, z)$. It is easy to see that Γ' is not weakly acyclic, and in fact not finitely chasable. Let Γ'' consist of the single egd $\mathtt{R}(x, y) \rightarrow (x = y)$. Now let Γ be $\Gamma' \cup \Gamma''$. Then Γ is finitely chasable (since in this case, we need only chase with Γ'' alone). However, Γ is not weakly acyclic, since Γ' is not weakly acyclic. □

Let Γ be a finitely chasable set of tgds and egds, and let $\mathcal{S}$ be the class of all source instances that satisfy Γ. Assume that $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$, and Σ_{21} is a finite set of s-t tgds. For each relational symbol $\mathtt{R}$ of $\mathbf{S_1}$, let $I_{\mathtt{R}}$ be a one-tuple instance that contains only the fact $\mathtt{R}(\mathbf{x})$, where the variables in $\mathbf{x}$ are distinct. Let $I^{\Gamma}_{\mathtt{R}}$ be a finite instance that is a result of chasing $I_{\mathtt{R}}$ with Γ, where it is all right to allow distinct variables in $\mathbf{x}$ to be equated by the chase. In our case of greatest interest, where Γ is the empty set, we have $I^{\Gamma}_{\mathtt{R}} = I_{\mathtt{R}}$. Let $J^{\Gamma}_{\mathtt{R}}$ be $chase_{12}(I^{\Gamma}_{\mathtt{R}})$, a result of chasing $I^{\Gamma}_{\mathtt{R}}$ with Σ_{12}.[6]

Since $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$, in particular $\mathcal{M}_{21}$ is a local inverse of $\mathcal{M}_{12}$ for $I^{\Gamma}_{\mathtt{R}}$ (this is because $I^{\Gamma}_{\mathtt{R}}$ is a member of $\mathcal{S}$). It follows from Corollary 4.3 that every value that appears in a tuple of $I^{\Gamma}_{\mathtt{R}}$ (and hence in a tuple of $\widehat{I^{\Gamma}_{\mathtt{R}}}$) appears in a tuple of $J^{\Gamma}_{\mathtt{R}}$. Therefore, $\beta_{J,\widehat{I}}$ is a full tgd, where I is $I^{\Gamma}_{\mathtt{R}}$, and J is $J^{\Gamma}_{\mathtt{R}}$, with $\beta_{\cdot,\cdot}$ as defined in Section 6. Let us denote this full tgd by $\delta^{\Gamma}_{\mathtt{R}}$, and let $\Sigma^{\mathcal{S}}_{12}$ consist of all of the tgds $\delta^{\Gamma}_{\mathtt{R}}$, one for every relation symbol $\mathtt{R}$ of $\mathbf{S_1}$. Define the *canonical $\mathcal{S}$-inverse of* $\mathcal{M}_{12}$ to be $\mathcal{M}^{\mathcal{S}}_{12} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^{\mathcal{S}}_{12})$ (we shall show that it is actually an $\mathcal{S}$-inverse). In the case where $\mathcal{S}$ is the class of all source instances, we may write Σ^{-1}_{12} for $\Sigma^{\mathcal{S}}_{12}$, and $\mathcal{M}^{-1}_{12}$ for $\mathcal{M}^{\mathcal{S}}_{12}$, to honor the fact that we are then dealing with a schema mapping that is a global inverse). We call $\mathcal{M}_{21}$ the *most general s-t $\mathcal{S}$-inverse of* $\mathcal{M}_{12}$ if Σ'_{21} logically implies Σ_{21} for every $\mathcal{S}$-inverse $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ of $\mathcal{M}_{12}$ where Σ'_{21} is a finite set of s-t tgds.

EXAMPLE 7.2. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. Assume that $\mathbf{S_1}$ consists of the binary relation symbol $\mathtt{R}$ and the unary relation symbol $\mathtt{S}$, and that $\mathbf{S_2}$ consists of the binary relation symbols $\mathtt{T}$ and $\mathtt{U}$. Let Σ_{12} consist of the s-t tgds $\mathtt{R}(x_1, x_2) \rightarrow \exists y(\mathtt{T}(x_1, y) \wedge \mathtt{U}(y, x_2))$, $\mathtt{R}(x, x) \rightarrow \mathtt{U}(x, x)$, and $\mathtt{S}(x) \rightarrow \exists y \mathtt{U}(x, y)$. Let Γ consist of the egd $\mathtt{R}(x_1, x_2) \rightarrow (x_1 = x_2)$.

[6]Even though $J^{\Gamma}_{\mathtt{R}}$ depends not just on $\mathtt{R}$ and Γ, but also on Σ_{12}, for simplicity we do not reflect the dependency on Σ_{12} in the notation $J^{\Gamma}_{\mathtt{R}}$.

Now $I_{\mathtt{R}}$ consists of the fact $\mathtt{R}(x_1, x_2)$, and so $I_{\mathtt{R}}^{\Gamma}$ consists of the fact $\mathtt{R}(x_1, x_1)$. Then $J_{\mathtt{R}}^{\Gamma}$ consists of the facts $\mathtt{T}(x_1, y)$, $\mathtt{U}(y, x_1)$, and $\mathtt{U}(x_1, x_1)$. So $\delta_{\mathtt{R}}^{\Gamma}$ is the tgd

$$(\mathtt{T}(x_1, y) \wedge \mathtt{U}(y, x_1) \wedge \mathtt{U}(x_1, x_1)) \rightarrow \widehat{\mathtt{R}}(x_1, x_1).$$

Also, $I_{\mathtt{S}}$ and $I_{\mathtt{S}}^{\Gamma}$ each consist of the fact $\mathtt{S}(x_1)$, and $J_{\mathtt{S}}^{\Gamma}$ consists of the fact $\mathtt{U}(x_1, y)$. So $\delta_{\mathtt{S}}^{\Gamma}$ is the tgd $\mathtt{U}(x_1, y) \rightarrow \widehat{\mathtt{S}}(x_1)$. Finally, $\mathcal{M}_{12}^{\mathcal{S}} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{12}^{\mathcal{S}})$, where $\Sigma_{12}^{\mathcal{S}}$ consists of the tgds $\delta_{\mathtt{R}}^{\Gamma}$ and $\delta_{\mathtt{S}}^{\Gamma}$. □

THEOREM 7.3. *Let $\mathcal{M}$ be a schema mapping defined by a finite set of s-t tgds. Let Γ be a finitely chasable set of tgds and egds, and let $\mathcal{S}$ be the class of source instances that satisfy Γ. Assume that $\mathcal{M}$ has an $\mathcal{S}$-inverse that is defined by a finite set of s-t tgds. Then the canonical $\mathcal{S}$-inverse of $\mathcal{M}$ is indeed an $\mathcal{S}$-inverse of $\mathcal{M}$, and in fact the most general s-t $\mathcal{S}$-inverse of $\mathcal{M}$.*

8. FULL TGDS SUFFICE FOR AN INVERSE

The canonical local inverse and the canonical $\mathcal{S}$-inverse are each defined by a finite set of *full* tgds. In this section, we show that this is no accident: if $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are schema mappings that are each defined by a finite set of s-t tgds, $\mathcal{S}$ is a class of source instances, and $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$, then there is a schema mapping $\mathcal{M}_{21}^{f}$ defined by a finite set of *full* s-t tgds and that is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$. While the canonical local inverse is tailored to a particular instance I, the mapping M_{21}^{f} is, as we shall see, constructed only from $\mathcal{M}_{21}$. From a technical point of view, this contrasts also with the canonical global inverse, which is constructed only from $\mathcal{M}_{12}$.

We begin with some definitions. Let γ be an s-t tgd. Assume that γ is $\forall\mathbf{x}(\varphi_{\mathbf{S}}(\mathbf{x}) \rightarrow \exists\mathbf{y}\psi_{\mathbf{T}}(\mathbf{x}, \mathbf{y}))$, where $\varphi_{\mathbf{S}}(\mathbf{x})$ is a conjunction of atomic formulas over $\mathbf{S}$ and $\psi_{\mathbf{T}}(\mathbf{x}, \mathbf{y})$ is a conjunction of atomic formulas over $\mathbf{T}$. Let $\psi_{\mathbf{T}}^{f}(\mathbf{x})$ be the conjunction of all atomic formulas in $\psi_{\mathbf{T}}(\mathbf{x}, \mathbf{y})$ that do not contain any variables in $\mathbf{y}$ (the f stands for "full"). Define γ^f (the *full part of* γ) to be the full tgd $\forall\mathbf{x}(\varphi_{\mathbf{S}}(\mathbf{x}) \rightarrow \psi_{\mathbf{T}}^{f}(\mathbf{x}))$. If $\psi_{\mathbf{T}}^{f}(\mathbf{x})$ is an empty conjunction, then γ^f is a *dummy tgd* where the right-hand side is "Truth" (and so the dummy tgd itself is "Truth").

Let $\psi_{\mathbf{T}}^{n}(\mathbf{x})$ be the conjunction of all atomic formulas in $\psi_{\mathbf{T}}(\mathbf{x}, \mathbf{y})$ that contain some variable in $\mathbf{y}$ (the n stands for "non-full"). Define γ^n (the *non-full part of* γ) to be the tgd $\forall\mathbf{x}(\varphi_{\mathbf{S}}(\mathbf{x}) \rightarrow \exists\mathbf{y}\psi_{\mathbf{T}}^{n}(\mathbf{x}, \mathbf{y}))$. As before, if $\psi_{\mathbf{T}}^{n}(\mathbf{x})$ is an empty conjunction, then γ^n is a *dummy tgd* where the right-hand side is "Truth"(and so the dummy tgd itself is "Truth"). If Σ is a set of tgds, let Σ^f be the set of γ^f where $\gamma \in \Sigma$ and where γ^f is not a dummy tgd. Similarly, let Σ^n be the set of γ^n where $\gamma \in \Sigma$ and where γ^n is not a dummy tgd. It is easy to see that Σ is logically equivalent to $\Sigma^f \cup \Sigma^n$. The next theorem tells us that only full tgds play a role in the inverse.

THEOREM 8.1. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of s-t tgds. Let $\mathcal{M}_{21}^{f} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21}^{f})$. If $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$, then so is $\mathcal{M}_{21}^{f}$.*

The following corollary is immediate (by letting Σ'_{21} in the corollary be Σ_{21}^{f}).

COROLLARY 8.2. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of s-t tgds. Assume that $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$. Then there is a finite set Σ'_{21} of full tgds such that $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$.*

9. REVERSING THE ARROWS (NOT!)

It is folk wisdom that simply "reversing the arrows" gives an inverse. What does this mean in our context?

Let us call a full tgd *reversible* if the same variables appear in the left-hand side as the right-hand side. If γ is a reversible tgd $\varphi \rightarrow \psi$, define *rev*(γ) to be the full tgd $\psi \rightarrow \widehat{\varphi}$, where $\widehat{\varphi}$ is the result of replacing every relational symbol $\mathtt{R}$ by $\widehat{\mathtt{R}}$. Since γ is reversible, *rev*(γ) is indeed a full tgd. We think of *rev*(γ) as the result of "reversing the arrow" of γ.

EXAMPLE 9.1. We now give a simple example that shows that $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \{rev(\gamma) : \gamma \in \Sigma_{12}\})$ is not necessarily a global inverse of $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, even when Σ_{12} consists of a finite set of reversible tgds and $\mathcal{M}_{12}$ has a global inverse that is defined by a finite set of s-t tgds. Let $\mathbf{S_1}$ consist of the unary relation symbols $\mathtt{R}_1$ and $\mathtt{R}_2$. Let $\mathbf{S_2}$ consist of the unary relation symbols $\mathtt{S}_1$, $\mathtt{S}_2$, and $\mathtt{S}_3$. Let $\Sigma_{12} = \{\mathtt{R}_1(x) \rightarrow \mathtt{S}_1(x), \mathtt{R}_2(x) \rightarrow \mathtt{S}_2(x), \mathtt{R}_1(x) \rightarrow \mathtt{S}_3(x), \mathtt{R}_2(x) \rightarrow \mathtt{S}_3(x)\}$. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. Let $\Sigma_{21} = \{\mathtt{S}_1(x) \rightarrow \widehat{\mathtt{R}_1}(x), \mathtt{S}_2(x) \rightarrow \widehat{\mathtt{R}_2}(x)\}$. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$. It is easy to see that $\mathcal{M}_{21}$ is a global inverse of $\mathcal{M}_{12}$.

Now let $\Sigma'_{21} = \{rev(\gamma) : \gamma \in \Sigma_{12}\}$. Thus $\Sigma'_{21} = \{\mathtt{S}_1(x) \rightarrow \widehat{\mathtt{R}_1}(x), \mathtt{S}_2(x) \rightarrow \widehat{\mathtt{R}_2}(x), \mathtt{S}_3(x) \rightarrow \widehat{\mathtt{R}_1}(x), \mathtt{S}_3(x) \rightarrow \widehat{\mathtt{R}_2}(x)\}$. Let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. It is easy to verify that $\mathcal{M}'_{21}$ is not a global inverse of $\mathcal{M}_{12}$. So simply "reversing the arrows" does not necessarily give a global inverse, even when there is a global inverse. □

Note that although $\{rev(\gamma) : \gamma \in \Sigma_{12}\}$ in Example 9.1 does not define a global inverse, some subset of it (namely, Σ_{21}) does. The next theorem says that there is an example where there is no subset of $\{rev(\gamma) : \gamma \in \Sigma_{12}\}$ that defines a global inverse.

THEOREM 9.2. *There is a schema mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ where each member of Σ_{12} is a reversible tgd with a singleton right-hand side, that has a global inverse defined by a finite set of s-t tgds, but where there is no subset X of Σ_{12} such that $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \{rev(\gamma) : \gamma \in X\})$ is a global inverse of $\mathcal{M}_{12}$.*

10. CHARACTERIZING THE CLASS $\mathcal{S}$

Given schema mappings $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, we would like to know the class $\mathcal{S}$ of source instances I such that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I. It is easy to see that this class is precisely the largest class $\mathcal{S}$ such that $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$. Let $\mathcal{M}_{11}$ be the schema mapping $(\mathbf{S_1}, \widehat{\mathbf{S_1}}, \sigma)$, where σ is the composition formula $\Sigma_{12} \circ \Sigma_{21}$. So the class $\mathcal{S}$ we are seeking is the class of all instances I such that $\mathcal{M}_{11}$ and the identity mapping are equivalent on I. Therefore, the class $\mathcal{S}$ is determined completely by the composition formula σ. We shall show that remarkably, there is a syntactic transformation of σ that produces a formula Γ that actually defines $\mathcal{S}$! We now begin our development.

Given a collection $\mathbf{x}$ of variables and a collection $\mathbf{f}$ of function symbols, a *term (based on* $\mathbf{x}$ *and* $\mathbf{f}$*)* is defined recursively as follows: (1) Every variable in $\mathbf{x}$ is a term, and (2) if f is a k-ary function symbol in $\mathbf{f}$ and $t_1, \ldots, t_k$ are terms, then $f(t_1, \ldots, t_k)$ is a term. We now define a second-order tgd [7].

DEFINITION 10.1. Let $\mathbf{S}$ be a source schema and $\mathbf{T}$ a target schema. A *second-order tuple-generating dependency (SO tgd)* is a formula of the form:

$$\exists \mathbf{f}((\forall \mathbf{x_1}(\varphi_1 \rightarrow \psi_1)) \wedge \cdots \wedge (\forall \mathbf{x_n}(\varphi_n \rightarrow \psi_n))),$$

where (1) each member of $\mathbf{f}$ is a function symbol; (2) each φ_i is a conjunction of (a) atomic formulas $S(y_1, \ldots, y_k)$, where S is a k-ary relation symbol of schema $\mathbf{S}$, and $y_1, \ldots, y_k$ are variables in $\mathbf{x_i}$, not necessarily distinct, and (b) equalities of the form $t = t'$ where t and t' are terms based on $\mathbf{x_i}$ and $\mathbf{f}$; (3) each ψ_i is a conjunction of atomic formulas $T(t_1, \ldots, t_l)$, where T is an l-ary relation symbol of schema $\mathbf{T}$ and $t_1, \ldots, t_l$ are terms based on $\mathbf{x_i}$ and $\mathbf{f}$; and (4) each variable in $\mathbf{x_i}$ appears in some atomic formula of φ_i.

As noted in [7], every finite set of s-t tgds is logically equivalent to an SO tgd (but not conversely).

If γ is an SO tgd, or a set of (first-order) tgds, from $\mathbf{S}$ to $\widehat{\mathbf{S}}$, define $\gamma^\sharp$ to be the source constraint that is the result of replacing each relational symbol $\widehat{\mathtt{R}}$ in γ by $\mathtt{R}$. For example, if γ is the s-t tgd (4), then $\gamma^\sharp$ is (2). The next proposition follows easily from the definitions of $\widehat{I}$ and of $\gamma^\sharp$.

PROPOSITION 10.2. *Let γ be an SO tgd, or a set of s-t tgds with source $\mathbf{S}$ and target $\widehat{\mathbf{S}}$, and let I be an instance of $\mathbf{S}$. Then $I \models \gamma^\sharp$ if and only if $\langle I, \widehat{I} \rangle \models \gamma$.*

We now need some more definitions. Let γ be an SO tgd. The *equality-free reduction*[7] γ^* of γ is obtained in multiple steps. First, we recursively replace each equality $f(t_1, \ldots, t_k) = f(t'_1, \ldots, t'_k)$ by $(t_1 = t'_1) \wedge \cdots \wedge (t_k = t'_k)$. We replace each equality $f(\mathbf{t}) = g(\mathbf{t}')$ where f and g are different function symbols by "False". Similarly, we replace each equality $f(\mathbf{t}) = x$, where x is a variable, by "False". Intuitively, only those equalities that are "forced" remain. We then "clean up" by deleting each "tgd" that appears as a conjunct of γ and that contains "False". The remaining equalities are all of the form $x = y$, where x and y are variables. Within each "tgd", we form equivalence classes of variables based on these equalities (where two variables are in the same equivalence class if they are forced to be equal by these equalities), replace each occurrence of each variable by a fixed representative of its equivalence class, and delete the equalities. The final result γ^* is an SO tgd that contains no equalities.

For example, the equality-free reduction of the SO tgd (1) is $\exists f(\forall e(\mathtt{Emp}(e) \rightarrow \mathtt{Mgr}(e, f(e)))$, the result of dropping the second clause of (1). As another example, consider the following SO tgd:

$$\exists f(\forall x \forall y(\mathtt{R}(x,y) \wedge (f(x) = f(y)) \rightarrow \mathtt{S}(x, f(x)) \wedge \mathtt{T}(x,y))) \quad (5)$$

Its quantifier-free reduction is $\exists f(\forall x(\mathtt{R}(x,x) \rightarrow \mathtt{S}(x, f(x)) \wedge \mathtt{T}(x,x)))$, which is obtained by replacing $f(x) = f(y)$ by $x = y$ and simplifying.

We now define *fulltgd*(γ), which is a set of full tgds that we associate with the SO tgd γ. To obtain *fulltgd*(γ), we first find the equality-free reduction γ^* of γ. We then rewrite γ^* so that each right-hand side is a singleton. Thus, we replace $\varphi \rightarrow (\psi_1 \wedge \cdots \wedge \psi_r)$, where $\psi_1, \ldots, \psi_r$ are atomic formulas, by $(\varphi \rightarrow \psi_1) \wedge \cdots \wedge (\varphi \rightarrow \psi_r)$. We then delete each "tgd" $\varphi \rightarrow \psi$ where the right-hand side ψ contains a function symbol. Then *fulltgd*(γ) is the set of s-t tgds that remain. These are real tgds, since there are no function symbols present. By construction, *fulltgd*(γ) is a set of full tgds with singleton right-hand sides.

As an example, when γ is the SO tgd (1), then *fulltgd*(γ) is the empty set. As another example, when γ is the SO tgd (5), then *fulltgd*(γ) contains the single tgd $\mathtt{R}(x,x) \rightarrow \mathtt{T}(x,x)$.

For each SO tgd γ where the source is $\mathbf{S}$ and the target is $\widehat{\mathbf{S}}$, we now define $\gamma^\dagger$. As we shall see, if σ is the composition formula, then $\sigma^\dagger$ plays a complementary role to $\sigma^\sharp$. For each k and each k-ary relational symbol $\mathtt{R}$ of $\mathbf{S}$, take $x_1, \ldots, x_k$ to be k distinct variables that do not appear in *fulltgd*(γ). Let $A_{\mathtt{R}}$ be the set of all tgds of *fulltgd*(γ) where the relational symbol in the right-hand side is $\widehat{\mathtt{R}}$. For each $\alpha \in A_{\mathtt{R}}$, assume that α is $\nu(\mathbf{y}) \rightarrow \widehat{\mathtt{R}}(y_1, \ldots, y_k)$, where $y_1, \ldots, y_k$ are not necessarily distinct (since α is full, every y_i appears in $\mathbf{y}$). Define μ_α to be the first-order formula $\exists \mathbf{y}(\nu(\mathbf{y}) \wedge (x_1 = y_1) \wedge \cdots \wedge (x_k = y_k))$. Define $\psi_{\mathtt{R}}$ to be $\mathtt{R}(x_1 \ldots, x_k) \rightarrow \vee \{\mu_\alpha : \alpha \in A_{\mathtt{R}}\}$. Since the empty disjunction represents "False", it follows that if $A_{\mathtt{R}} = \emptyset$, then $\psi_{\mathtt{R}}$ is equivalent to $\neg \mathtt{R}(x_1 \ldots, x_k)$. Now define $\gamma^\dagger$ to be the conjunction of the formulas $\psi_{\mathtt{R}}$ (over all relational symbols $\mathtt{R}$ of $\mathbf{S}$). Note that $\gamma^\dagger$ is a first-order formula.

EXAMPLE 10.3. Assume that there are two source relation symbols $\mathtt{R}$ and $\mathtt{S}$, and assume that *fulltgd*(γ) consists of the following tgds, which we denote by α_1, α_2:

$$(\alpha_1): \quad \mathtt{R}(y_2, y_1, y_3) \wedge \mathtt{S}(y_2, y_3, y_3) \rightarrow \widehat{\mathtt{R}}(y_1, y_1, y_2)$$
$$(\alpha_2): \quad \mathtt{S}(y_1, y_2, y_2) \rightarrow \widehat{\mathtt{R}}(y_1, y_2, y_1)$$

So $\mu_{\alpha_1}, \mu_{\alpha_2}$ are as follows:

$$(\mu_{\alpha_1}): \quad \exists y_1 \exists y_2 \exists y_3 \quad (\mathtt{R}(y_2, y_1, y_3) \wedge \mathtt{S}(y_2, y_3, y_3) \wedge (x_1 = y_1) \wedge (x_2 = y_1) \wedge (x_3 = y_2))$$
$$(\mu_{\alpha_2}): \quad \exists y_1 \exists y_2 \exists y_3 \quad (\mathtt{S}(y_1, y_2, y_2) \wedge (x_1 = y_1) \wedge (x_2 = y_2) \wedge (x_3 = y_1))$$

Then $\psi_{\mathtt{R}}$ is $\mathtt{R}(x_1, x_2, x_3) \rightarrow (\mu_{\alpha_1} \vee \mu_{\alpha_2})$. Further, $\psi_{\mathtt{S}}$ is $\neg \mathtt{S}(x_1, x_2, x_3)$, since $A_{\mathtt{S}} = \emptyset$. Finally, $\gamma^\dagger$ is $(\mathtt{R}(x_1, x_2, x_3) \rightarrow (\mu_{\alpha_1} \vee \mu_{\alpha_2})) \wedge \neg \mathtt{S}(x_1, x_2, x_3)$. (Of course, this formula is universally quantified with $\forall x_1 \forall x_2 \forall x_3$, but we suppress this as usual.) □

We now have the following proposition.

PROPOSITION 10.4. *Let γ be an SO tgd with source $\mathbf{S}$ and target $\widehat{\mathbf{S}}$, and let I be an instance of $\mathbf{S}$. Then $I \models \gamma^\dagger$ if and only if for every J such that $\langle I, J \rangle \models \gamma$, necessarily $\widehat{I} \subseteq J$.*

The next theorem gives a formula that defines the largest class $\mathcal{S}$ of source instances where $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$.

THEOREM 10.5. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings where Σ_{12} and Σ_{21} are finite sets of s-t tgds. Let σ be $\Sigma_{12} \circ \Sigma_{21}$. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I if and only if $I \models \sigma^\sharp \wedge \sigma^\dagger$.*

[7]A similar notion appears in [14] under the name "mapping reduction".

Held fixed	*Complexity*
$\mathcal{M}_{12}$ and $\mathcal{M}_{21}$	NP; may be NP-complete
Full $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$	polytime
$\mathcal{M}_{12}$	Σ_2^P; may be coNP-hard
Full $\mathcal{M}_{12}$	coNP; may be coNP-complete

Figure 1: Local invertibility: input is source instance I

Input	*Complexity*
$\mathcal{M}_{12}$ and $\mathcal{M}_{21}$	DP-hard
Full $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$	DP-complete
$\mathcal{M}_{12}$	coNP-hard
Full $\mathcal{M}_{12}$	coNP-complete

Figure 2: Global invertibility

From our earlier Theorem 5.1, we can prove that if $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are each defined by a finite set of s-t tgds and held fixed, then the problem of deciding, given I, whether $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I is in NP. (Complexity results appear in Section 11.) So by Fagin's Theorem [3], the class $\mathcal{S}$ of source instances I such that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I can be defined by a formula Γ in existential second-order logic. What is remarkable is that, as Theorem 10.5 says, there is such a formula Γ, namely $\sigma^\sharp \wedge \sigma^\dagger$, that can be obtained from the composition formula σ by a purely syntactical transformation.

The following corollary gives an important case where $\mathcal{S}$ is first-order definable.

COROLLARY 10.6. *Assume that $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are schema mappings that are each defined by a finite set of full s-t tgds. There is a first-order formula φ such that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I if and only if $I \models \varphi$.*

EXAMPLE 10.7. We continue with our running example (from Examples 3.1 and 3.2). We shall fulfill our promise to show that the schema mapping $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for precisely those source instances I that satisfy Γ. We noted that Γ (as given by (2)) looks mysteriously similar to the composition formula (as given by (4)). We shall explain this mystery.

We observed in Example 3.2 that σ logically implies Σ_{Id}. We now show that this implies that $\sigma^\dagger$ is valid. Let I be an arbitrary instance of $\mathbf{S_1}$; we must show that $I \models \sigma^\dagger$. By Proposition 10.4, we need only show that for every J such that $\langle I, J\rangle \models \sigma$, necessarily $\widehat{I} \subseteq J$. Let J be arbitrary such that $\langle I, J\rangle \models \sigma$. Since $\sigma \models \Sigma_{Id}$, it follows that $\langle I, J\rangle \models \Sigma_{Id}$. Therefore, $\widehat{I} \subseteq J$, as desired. So indeed, $\sigma^\dagger$ is valid.

Therefore, by Theorem 10.5, we know that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I if and only if $I \models \sigma^\sharp$. But $\sigma^\sharp$ is exactly Γ. This not only proves our claim that the schema mapping $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for precisely those source instances I that satisfy Γ, but also explains the mystery of the resemblance of σ and Γ. In fact, this mysterious resemblance in this example is what inspired us to search for and discover Theorem 10.5. □

11. COMPLEXITY RESULTS

We have investigated complexity issues, dealing with both local and global invertibility. In this paper, we do not consider complexity issues for $\mathcal{S}$-invertibility except when $\mathcal{S}$ is a singleton (local invertibility) and when $\mathcal{S}$ is the class of all source instances.(global invertibility). It might be interesting to consider complexity issues for other choices of $\mathcal{S}$. Our results are summarized in the tables of Figures 1 and 2. In both tables, we consider separately the cases where the tgds that define $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are full.

In the table of Figure 1, the input is a source instance I. The first line of the table says that if $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are fixed schema mappings each defined by a finite set of s-t tgds, then the problem of deciding if $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ for I is in NP, and there is a choice of $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ where the problem is NP-complete.[8] In the second line we consider the same problem as the first line, but $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are each defined by a finite set of *full* tgds. Then the complexity drops to polynomial time (in fact, by Corollary 10.6, the problem is even definable in first-order logic, which makes it logspace computable). The third line considers whether $\mathcal{M}_{12}$ has an inverse for I defined by a finite set of s-t tgds. Since we have shown how to obtain a canonical local inverse that is an inverse for I if there is any inverse for I defined by a finite set of tgds, the reader may be puzzled as to why this problem does not reduce to the problem in the first line, where $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are given. The reason is that the size of Σ_{21} that defines the canonical local inverse grows with I, unlike the situation in the first line where Σ_{21} is given and so of fixed size. There is a complexity gap in the third line, where we have an upper bound of Σ_2^P in the polynomial-time hierarchy, and a lower bound of coNP-hardness. In the fourth line we consider the same problem as the third line, but $\mathcal{M}_{12}$ is defined by a finite set of full tgds. The problem is then in coNP, and there is a choice of $\mathcal{M}_{12}$ where the problem is coNP-complete.

In the first line of the table of Figure 2, the input consists of schema mappings $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ that are each defined by a finite set of s-t tgds, and the problem is deciding whether $\mathcal{M}_{21}$ is a global inverse of $\mathcal{M}_{12}$. In the second line we consider the same problem as the first line, but $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are each defined by a finite set of full tgds, and the problem is DP-complete.[9] The third line considers whether $\mathcal{M}_{12}$ has a global inverse defined by a finite set of s-t tgds. In the fourth line we consider the same problem as the third line, but $\mathcal{M}_{12}$ is defined by a finite set of full tgds, and the problem is coNP-complete. In fact, this problem is coNP-complete even when the tgds that define $\mathcal{M}_{12}$ all have a singleton left-hand side. The first and third lines inherit their lower bounds from the full cases (the second and fourth lines, respectively).

There is a large complexity gap in the first and third lines, since it is open as to whether these problems are even decidable. When the tgds that define $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are full, the reason the problem is decidable (and in fact, DP-complete or coNP-complete) is a small model theorem that guarantees that if $\mathcal{M}_{21}$ is not a global inverse, then there is a small (polynomial-size) counterexample I. We close this section with a discussion of small model theorems, including a reason for the difficulty in proving a small model theorem when the tgds that define $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are not necessarily full.

[8]The NP-hardness result was obtained by Phokion Kolaitis.

[9]The class DP consists of all decision problems that can be written as the intersection of an NP problem and a coNP problem.

11.1 Small model theorems

We begin with a small submodel theorem for the case of schema mappings that are each defined by a finite set of *full* tgds.

THEOREM 11.1. *Let* $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ *and* $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ *be schema mappings where* Σ_{12} *and* Σ_{21} *are finite sets of full s-t tgds. Let* I *be an instance of* $\mathbf{S_1}$. *Assume that* $\mathcal{M}_{21}$ *is not an inverse of* $\mathcal{M}_{12}$ *for* I. *Then there is a subinstance* I' *of* I, *with size polynomial in the size of* Σ_{12} *and* Σ_{21}, *such that* $\mathcal{M}_{21}$ *is not an inverse of* $\mathcal{M}_{12}$ *for* I'.

As an immediate corollary of the small submodel theorem, we obtain the following small model theorem.

THEOREM 11.2. *Let* $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ *and* $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ *be schema mappings where* Σ_{12} *and* Σ_{21} *are finite sets of full s-t tgds. Assume that* $\mathcal{M}_{21}$ *is not a global inverse of* $\mathcal{M}_{12}$. *Then there is an instance* I, *with size polynomial in the size of* Σ_{12} *and* Σ_{21}, *such that* $\mathcal{M}_{21}$ *is not an inverse of* $\mathcal{M}_{12}$ *for* I.

The next theorem implies that the small submodel theorem fails dramatically when the tgds are not necessarily full.

THEOREM 11.3. *There are schema mappings* $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ *and* $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, *where* Σ_{12} *and* Σ_{21} *are finite sets of s-t tgds, and where for arbitrarily large* n, *there is an instance* I *of* $\mathbf{S_1}$ *consisting of* n *facts, such that* $\mathcal{M}_{21}$ *is not an inverse of* $\mathcal{M}_{12}$ *for* I, *but* $\mathcal{M}_{21}$ *is an inverse of* $\mathcal{M}_{12}$ *for every proper subinstance* I' *of* I.

Although the small submodel theorem fails when the tgds are not necessarily full, it is open as to whether the small model theorem holds in this case. If the small model theorem were to hold, then our problems in the first and third lines of the table of Figure 2 would be decidable (and as we can show, even in Π_2^P in the polynomial-time hierarchy).

12. SUMMARY AND OPEN PROBLEMS

We have given a formal definition for one schema mapping to be an inverse of another schema mapping for a class $\mathcal{S}$ of source instances. We have obtained a number of results about our notion of inverse, and some of these results are surprising.

There are many open problems, as we would expect from a "first step" paper like this. Section 11 gives us several open problems, including closing the complexity gaps and resolving whether the small model theorem holds when the tgds are not necessarily full. We now mention some other open problems.

- We have focused most of our attention on schema mappings defined by a finite set of s-t tgds. What about more general schema mappings? What if we allow target dependencies, such as functional dependencies?
- We have focused on right inverses, where we are given $\mathcal{M}_{12}$ and want to find a right inverse $\mathcal{M}_{21}$. It might be interesting to study the left inverse, where we are given $\mathcal{M}_{21}$ and we wish to find $\mathcal{M}_{12}$.
- Our next open problem is somewhat imprecise, but is important in practice. Assume that we are given $\mathcal{M}_{12}$. How do we find a large class $\mathcal{S}$ and a schema mapping $\mathcal{M}_{21}$ such that $\mathcal{M}_{21}$ is an $\mathcal{S}$-inverse of $\mathcal{M}_{12}$? In fact, there might be several such large classes $\mathcal{S}$ and corresponding inverse mappings. How do we find them? This problem is imprecise, because it is not clear what we mean by a "large class" $\mathcal{S}$. We should not necessarily restrict our attention to classes $\mathcal{S}$ defined by a finitely chasable set Γ of tgds and egds.
- It might be interesting to explore more fully the unique solutions property, which is an interesting notion in its own right.
- We might explore the notion of $\widehat{I}$ and $chase_{21}(chase_{12}(I))$ being homomorphically equivalent. By Theorem 5.4, this notion is strictly weaker than $\mathcal{M}_{21}$ being an inverse of $\mathcal{M}_{12}$ for I.

This paper is, we think, simply the first step in a fascinating journey!

Acknowledgments

The author is grateful to Phokion Kolaitis, Lucian Popa, and Wang-Chiew Tan for numerous invaluable discussions and suggestions.

13. REFERENCES

[1] C. Beeri and M. Y. Vardi. A Proof Procedure for Data Dependencies. *Journal of the Association for Computing Machinery (JACM)*, 31(4):718–741, 1984.

[2] P. A. Bernstein. Applying Model Management to Classical Meta-Data Problems. In *Conference on Innovative Data Systems Research (CIDR)*, pages 209–220, 2003.

[3] R. Fagin. Generalized First-Order Spectra and Polynomial-Time Recognizable Sets. In R. M. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings, Vol. 7*, pages 43–73, 1974.

[4] R. Fagin. Multivalued dependencies and a new normal form for relational databases. *ACM Transactions on Database Systems (TODS)*, 2(3):262–278, Sept. 1977.

[5] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data Exchange: Semantics and Query Answering. *Theoretical Computer Science*, pages 89–124, 2005.

[6] R. Fagin, P. G. Kolaitis, and L. Popa. Data Exchange: Getting to the Core. *ACM Transactions on Database Systems (TODS)*, 30(1):174–210, 2005.

[7] R. Fagin, P. G. Kolaitis, L. Popa, and W.-C. Tan. Composing Schema Mappings: Second-Order Dependencies to the Rescue. In *ACM Symposium on Principles of Database Systems (PODS)*, pages 83–94, 2004.

[8] P. Hell and J. Nešetřil. The Core of a Graph. *Discrete Mathematics*, 109:117–126, 1992.

[9] R. Hull. Relative Information Capacity of Simple Relational Database Schemata. *SIAM J. on Computing*, 15:856–886, 1986.

[10] M. Lenzerini. Data Integration: A Theoretical Perspective. In *ACM Symposium on Principles of Database Systems (PODS)*, pages 233–246, 2002.

[11] S. Melnik. *Generic Model Management: Concepts and Algorithms.* Springer, 2004.

[12] S. Melnik, P. Bernstein, A. Halevy, and E. Rahm. Supporting Executable Mappings in Model Management. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 167–178, 2005.

[13] N. C. Shu, B. C. Housel, R. W. Taylor, S. P. Ghosh, and V. Y. Lum. EXPRESS: A Data EXtraction, Processing, amd REStructuring System. *ACM Transactions on Database Systems (TODS)*, 2(2):134–174, 1977.

[14] C. Yu and L. Popa. Semantic Adaptation of Schema Mappings when Schemas Evolve. In *International Conference on Very Large Data Bases (VLDB)*, pages 1006–1017, 2005.