

Limits of Schema Mappings

Phokion G. Kolaitis¹ · Reinhard Pichler² ·
Emanuel Sallinger³ · Vadim Savenkov⁴

Published online: 2 October 2017

© The Author(s) 2017. This article is an open access publication

Abstract Schema mappings have been extensively studied in the context of data exchange and data integration, where they have turned out to be the right level of abstraction for formalizing data inter-operability tasks. Up to now and for the most part, schema mappings have been studied as static objects, in the sense that each time the focus has been on a single schema mapping of interest or, in the case of composition, on a pair of schema mappings of interest. In this paper, we adopt a dynamic viewpoint and embark on a study of sequences of schema mappings and of the limiting behavior of such sequences. To this effect, we first introduce a natural notion of distance on sets of finite target instances that expresses how “close” two sets of target instances are as regards the certain answers of conjunctive queries on these sets. Using this notion of distance, we investigate pointwise limits and

This article is part of the Topical Collection on *Special Issue on Database Theory*

✉ Emanuel Sallinger
emanuel.sallinger@cs.ox.ac.uk

Phokion G. Kolaitis
kolaitis@cs.ucsc.edu

Reinhard Pichler
pichler@dbai.tuwien.ac.at

Vadim Savenkov
vadim.savenkov@wu.ac.at

¹ UC Santa Cruz and IBM Research-Almaden, Santa Cruz, CA, USA

² TU Wien, Wien, Austria

³ University of Oxford, Oxford, UK

⁴ Vienna University of Economics and Business, Wien, Austria

uniform limits of sequences of schema mappings, as well as the companion notions of pointwise Cauchy and uniformly Cauchy sequences of schema mappings. We obtain a number of results about the limits of sequences of GAV schema mappings and the limits of sequences of LAV schema mappings that reveal striking differences between these two classes of schema mappings. We also consider the completion of the metric space of sets of target instances and obtain concrete representations of limits of sequences of schema mappings in terms of generalized schema mappings, that is, schema mappings with infinite target instances as solutions to (finite) source instances.

Keywords Schema mappings · Limits · Pointwise convergence · Uniform convergence

1 Introduction

Schema mappings have been extensively studied in the context of data exchange and data integration, where they have turned out to be the right level of abstraction for formalizing data inter-operability tasks (see the surveys [11, 12] and the monograph [1]). Up to now and for the most part, schema mappings have been studied as static objects, in the sense that each time the focus has been on a single schema mapping or on a finite and, typically, small number of schema mappings. In the case of data exchange [6], a single schema mapping is used to specify the relationship between a source schema and a target schema. In the case of operators on schema mappings [3], such as the composition operator [8, 14], a fixed number of schema mappings is used as input (e.g., two schema mappings in the case of composition) and another schema mapping is returned as output. Even the case of schema-mapping evolution [9] entails a finite (but potentially large) number of schema mappings.

In this paper, we adopt a dynamic viewpoint and embark on a systematic investigation of sequences of schema mappings and of the limiting behavior of such sequences. The original motivation came from the earlier work [2, 5, 7, 10, 14] on schema-mapping optimization and the study of various notions of equivalence between schema mappings that, intuitively, stipulate that two schema mappings cannot be distinguished using conjunctive queries (CQ-equivalence) or conjunctive queries with at most n variables (CQ $_n$ -equivalence), for some fixed $n \geq 1$. In particular, in [5] and, implicitly, in [14], it was shown that, given an SO-tgd (second-order tuple-generating dependency) σ and a positive integer n , one can construct a GLAV schema mapping that is CQ $_n$ -equivalent to σ . Informally, this means that a given SO tgd can be “approximated” by GLAV schema mappings up to any fixed level of precision, even though an SO tgd is a formula of second-order logic that may not be logically equivalent to any formula of first-order logic and, in particular, to any GLAV schema mapping. A more dynamic interpretation is that, given an SO-tgd σ , one can obtain a sequence of GLAV schema mappings $(\mathcal{M}_n)_{n \geq 1}$, whose “limit” is σ .

Summary of Results Our contributions are both conceptual and technical. At the conceptual level, we develop a framework for studying sequences of schema

mappings by first introducing a natural notion of *distance* on the powerset $\mathcal{P}(\text{Inst}(\mathbf{T}))$ of the set $\text{Inst}(\mathbf{T})$ of finite instances over a schema \mathbf{T} . Intuitively, this notion of distance expresses how “close” two sets of finite \mathbf{T} -instances are as regards the certain answers of conjunctive queries on these sets. The pair $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is a pseudometric space, which means that the distance function $\text{dist}(\cdot, \cdot)$ is symmetric and obeys the triangle inequality, but different sets of finite target instances may have distance zero; however, two such sets have distance zero if and only if they are CQ-equivalent, i.e., every conjunctive query has the same certain answers on these two sets. Thus, we will also work with the metric space obtained by considering the CQ-equivalence classes of members of $\mathcal{P}(\text{Inst}(\mathbf{T}))$, and will use the same notation for it.

Sequences of functions from some set to a metric space occupy a central place in the study of metric spaces (see, e.g., [18]). In particular, there are natural notions of a *pointwise limit* and of a *uniform limit* of a sequence $(f_n)_{n \geq 1}$ of functions from some set to a metric space; moreover, there are companion notions of a *pointwise Cauchy* and of a *uniformly Cauchy* sequence of such functions. We now describe briefly how these notions can be applied to sequences of schema mappings. In its most general formulation, a schema mapping \mathcal{M} over a source schema \mathbf{S} and a target schema \mathbf{T} is a set of pairs (I, J) , where I is a finite \mathbf{S} -instance and J is a finite \mathbf{T} -instance. It follows that a schema mapping \mathcal{M} can also be viewed as a function f from the set $\text{Inst}(\mathbf{S})$ of all finite \mathbf{S} -instances to the powerset $\mathcal{P}(\text{Inst}(\mathbf{T}))$ of the set of all finite \mathbf{T} -instances, where $f(I) = \{J : (I, J) \in \mathcal{M}\}$. This way, a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings over a source schema \mathbf{S} and a target schema \mathbf{T} can be viewed as a sequence of functions from $\text{Inst}(\mathbf{S})$ to the (pseudo)metric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$.

After the conceptual framework has been laid out, we study in depth the limiting behavior of sequences of GAV mappings and the convergence of sequences of LAV mappings. We establish a number of technical results that reveal rather dramatic and perhaps unanticipated differences between GAV schema mappings and LAV schema mappings.

For sequences of GAV mappings, we point out that every uniformly Cauchy sequence of GAV mappings is eventually constant, hence it has a GAV mapping as uniform limit. We also show that every pointwise Cauchy sequence of GAV mappings has a pointwise limit, but it need not have a uniform limit; moreover, there are pointwise Cauchy sequences of GAV mappings such that no GAV mapping is their pointwise limit. This raises the question as to when a sequence of GAV mappings has a GAV mapping as a pointwise limit. We prove that a sequence of GAV mappings has a GAV mapping as a pointwise limit if and only if it has a pointwise limit that allows for CQ-rewriting.¹

For sequences of LAV mappings, we show that the notions of uniform limit and pointwise limit coincide; moreover, the same holds true for the notions of uniformly Cauchy and pointwise Cauchy sequences. However, there are uniformly Cauchy sequences of LAV mappings that have no uniform limit. We also establish that a uniformly Cauchy sequence of LAV mappings has a LAV mapping as a uniform limit if and only if it has a uniform limit that admits universal solutions. The

¹Allowing for CQ-rewriting means that the certain answers of every conjunctive query over the target schema is definable by a union of conjunctive queries over the source schema - see [19].

aforementioned results lift to sequences of *premise-bounded* sequences of GLAV mappings, i.e., sequences of GLAV mappings for which there is a $k \geq 1$ such that, for every mapping in the sequence, the left-hand side of every GLAV constraint has at most k source atoms (LAV mappings have $k = 1$).

In terms of techniques, we use systematically the structural characterizations of schema-mapping languages established in [19], thus creating a link with a different line of research.

The metric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is incomplete, i.e., there are Cauchy sequences of elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ that have no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. It is well known that every incomplete metric space (X, d) has a completion, which means that it can be embedded into a complete metric space (X^*, d^*) so that X is a dense subset of X^* . Moreover, pointwise (respectively, uniformly) Cauchy sequences of functions on X have pointwise (respectively, uniform) limits that take values in X^* . The construction of X^* from X involves equivalence classes of Cauchy sequences of elements of X , thus, in general, the members of X^* do not have a concrete representation. In the last part of the paper, we show that the members of $\mathcal{P}(\text{Inst}(\mathbf{T}))^*$ can be represented by suitably constructed infinite \mathbf{T} -instances. As a consequence of this, the pointwise (respectively, uniform) limits of Cauchy sequences of schema mappings can be represented by *generalized* schema mappings, i.e., schema mappings that allow for infinite target instances as solutions to finite source instances.

2 Preliminaries

This section contains a minimum amount of necessary background material.

Schemas, Instances, and Conjunctive Queries A *schema* \mathbf{R} is a finite sequence $\langle R_1, \dots, R_k \rangle$ of relation symbols, where each R_i has a fixed arity. An *instance* I over \mathbf{R} , or an *\mathbf{R} -instance*, is a sequence $\langle R_1^I, \dots, R_k^I \rangle$, where each R_i^I is a finite relation of the same arity as R_i . We will often use R_i to denote both the relation symbol and the relation R_i^I that interprets it. The *active domain* $\text{adom}(I)$ of an instance I is the set of all values occurring in the relations of I . A *fact* of an instance I (over \mathbf{R}) is an expression $R_i^I(t_1, \dots, t_m)$ (or simply $R_i(t_1, \dots, t_m)$), where R_i is a relation symbol of \mathbf{R} and $(t_1, \dots, t_m) \in R_i^I$.

A *conjunctive query* is a first-order formula of the form $\exists \mathbf{z} \theta(\mathbf{x}, \mathbf{z})$, where $\theta(\mathbf{x}, \mathbf{z})$ is a conjunction of atomic formulas $R_i(v_1, \dots, v_m)$ and each v_j is one of the variables in \mathbf{x} and \mathbf{z} . A *boolean conjunctive query* is a conjunctive query with no free variables. We write CQ for the class of all conjunctive queries over some schema. For every $n \geq 1$, we let CQ_n denote the class of all conjunctive queries with at most n variables. We also let CQ_0 denote the singleton consisting of a trivially true query. If I is an instance and q is a conjunctive query, then we write $q(I)$ for the result of evaluating q on I ; in particular, for boolean conjunctive queries q we have that $q(I) = \text{true}$ if and only if I satisfies q .

Schema Mappings, Universal Solutions, Certain Answers Motivated by the terminology in data exchange [6], we typically work with two schemas, a *source*

schema **S** and a target schema **T** with no relation symbols in common. We refer to **S**-instances as *source instances*, and to **T**-instances as *target instances*. We assume that the values occurring in the active domains of instances come from two fixed countably infinite disjoint sets, the set Const of all *constants* and the set Null of (*labeled*) *nulls*. We also assume that the active domains of source instances consist entirely of constants; the active domains of target instances may contain both constants and nulls.

In its most general form, a *schema mapping* \mathcal{M} between a source schema **S** and a target schema **T** is a set of pairs (I, J) , where I is source instance and J a target instance. To avoid anomalies that arise from such a relaxed notion, we will assume that a schema mapping \mathcal{M} must also possess a mild closure property, namely, that \mathcal{M} is closed under isomorphisms that rename nulls by other nulls. This is a natural “genericity” condition that is akin to the condition that database queries are closed under arbitrary isomorphisms. The precise definitions are as follows.

Definition 1 Let **S** be a source schema and **T** a target schema.

- An *isomorphism that renames nulls* between two target instances J and J' is a one-to-one and onto function $h : \text{adom}(J) \rightarrow \text{adom}(J')$ such that:
 - (i) If c is a constant in $\text{adom}(J)$, then $h(c) = c$.
 - (ii) If w is a null in $\text{adom}(J)$, then $h(w)$ is also a null.
 - (iii) For every relation symbol R of **T** of arity m and for every tuple (a_1, \dots, a_m) of constants and nulls, we have that $R^J(a_1, \dots, a_m)$ is a fact of J if and only if $R^{J'}(h(a_1), \dots, h(a_m))$ is a fact of J' .
 In this case, we write $J' = h(J)$ and say that J' is an *isomorphic copy of J via an isomorphism that renames nulls*.
- A *schema mapping* \mathcal{M} between **S** and **T** is a set of pairs (I, J) , where I is source instance and J a target instance, such that the following holds: if a pair (I, J) is in \mathcal{M} and if $J' = h(J)$ is an isomorphic copy of J via an isomorphism h that renames nulls, then also (I, J') is in \mathcal{M} .

A schema mapping is often (but not always) given as a triple $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, where Σ is a set of formulas in some logical formalism such that $(I, J) \in \mathcal{M}$ if and only if $I \cup J \models \Sigma$. Clearly, if Σ is a set of first-order formulas or a set of second-order formulas, then \mathcal{M} is indeed closed under isomorphisms that rename nulls holds.

Let \mathcal{M} be a fixed schema mapping. In data exchange, the main problem is, given a source instance I , to find a *solution* for I w.r.t. \mathcal{M} , that is, a target instance J such that $(I, J) \in \mathcal{M}$ (or determine that no solution exists). We use the notation $\text{Sol}(I, \mathcal{M}) = \{J \mid (I, J) \in \mathcal{M}\}$ to denote the set of all solutions for I w.r.t. \mathcal{M} . In data integration, the main problem is to compute the *certain answers* of queries [12]. Specifically, given a query q over the target schema and a source instance I , the *certain answers of q on I w.r.t. \mathcal{M}* is the set

$$\text{cert}(q, I, \mathcal{M}) = \bigcap \{q(J) \mid J \in \text{Sol}(I, \mathcal{M})\}.$$

If q is a boolean conjunctive query, then $\text{cert}(q, I, \mathcal{M}) = \text{true}$, if $q(J) = \text{true}$, for every solution J for I w.r.t. \mathcal{M} ; otherwise, $\text{cert}(q, I, \mathcal{M}) = \text{false}$. Note also if q is a non-boolean conjunctive query, then either $\text{cert}(q, I, \mathcal{M}) = \emptyset$ or every tuple $\mathbf{t} \in \text{cert}(q, I, \mathcal{M})$ is null-free, that is, it consists entirely of constants. This is a consequence of the closure of \mathcal{M} under isomorphisms that rename nulls. Indeed, assume that $\mathbf{t} \in \text{cert}(q, I, \mathcal{M})$. Let J be a solution for I w.r.t. \mathcal{M} and let $J' = h(J)$ be a target instance that is an isomorphic copy of J via an isomorphism h that renames nulls from the active domain of J to nulls outside the active domain of J (such a target instance J' and such an isomorphism h exist because J is a finite set of facts, hence its active domain is a finite set). By the closure property of \mathcal{M} , the target instance J' is also a solution for I w.r.t. \mathcal{M} , hence $\mathbf{t} \in q(J')$. It follows that \mathbf{t} must consist of values in the intersection $\text{adom}(J) \cap \text{adom}(J')$ of the active domains of J and J' , hence \mathbf{t} must consist entirely of constants. Note that the only property of conjunctive queries used in this argument is that they are *safe*, that is, they return tuples from the active domain of the instance on which they are evaluated.

On the face of it, the definition of certain answers may entail computing an intersection of infinitely many sets. One of the main findings in [6] is that there is a notion of a “good” solution in data exchange, called *universal solution*, that can also be used to compute the certain answers of conjunctive queries in a much more direct way.

Let J_1 and J_2 be two target instances. A function h is a *homomorphism* from J_1 to J_2 if the following hold: (i) for every constant c , we have that $h(c) = c$; and (ii) for every relation symbol R in \mathbf{R} and every tuple $(a_1, \dots, a_n) \in R^{J_1}$, we have that $(h(a_1), \dots, h(a_n)) \in R^{J_2}$. We write $J_1 \rightarrow J_2$ to denote that there is a homomorphism from J_1 to J_2 . We say that J_1 is *homomorphically equivalent* to J_2 , written $J_1 \leftrightarrow J_2$, if $J_1 \rightarrow J_2$ and $J_2 \rightarrow J_1$.

Let I be a source instance. A *universal solution* for I w.r.t. \mathcal{M} is a solution J such that for every solution $J' \in \text{Sol}(I, \mathcal{M})$, we have that $J \rightarrow J'$. Intuitively, a universal solution for I is a “most general” solution for I . We write $\text{UnivSol}(I, \mathcal{M})$ to denote the set of all universal solutions for I w.r.t. \mathcal{M} (note that universal solutions need not always exist, so it is possible that $\text{UnivSol}(I, \mathcal{M}) = \emptyset$). The following useful property of universal solutions was first identified in [6].

Proposition 1 *Assume that \mathcal{M} is a schema mapping, I is a source instance, and J is a universal solution for I w.r.t. \mathcal{M} . If q is a conjunctive query, then $\text{cert}(q, I, \mathcal{M}) = q(J)_\downarrow$, where $q(J)_\downarrow$ is the set of all null-free tuples in $q(J)$.*

Proof First, assume that $\mathbf{t} \in \text{cert}(q, I, \mathcal{M})$. Then, as discussed earlier, \mathbf{t} must be a null-free tuple. Since J is a solution for I w.r.t. \mathcal{M} , we have that $\mathbf{t} \in q(J)$, hence we have that $\mathbf{t} \in q(J)_\downarrow$. Next, assume that \mathbf{t} is a null-free tuple in $q(J)$. If J' is an arbitrary solution for I w.r.t. \mathcal{M} , then, since J is a universal solution for I w.r.t. \mathcal{M} , there is a homomorphism h from J to J' . Since conjunctive queries are preserved under homomorphisms, it follows that $h(\mathbf{t}) = \mathbf{t} \in q(J')$. Thus, $\mathbf{t} \in \text{cert}(q, I, \mathcal{M})$. \square

Structural Properties of Schema Mappings We now present a number of structural properties that a schema mapping may or may not possess. These properties

were investigated in their own right in [19], where they were used to obtain characterizations of schema-mapping languages that will be of great interest to us in this paper.

Let \mathcal{M} be a schema mapping.

- \mathcal{M} allows for CQ-rewriting if for every target conjunctive query q , there exists a union q' of source conjunctive queries such that $\text{cert}(I, \mathcal{M}, q) = q'(I)$, for every source instance I .
- \mathcal{M} admits universal solutions if for every source instance I , there is a universal solution for I w.r.t. \mathcal{M} .
- \mathcal{M} is closed under target homomorphisms if $(I, J) \in \mathcal{M}$ and $J \rightarrow J'$ imply that $(I, J') \in \mathcal{M}$.
- \mathcal{M} is closed under unions if $(I_1, J_1) \in \mathcal{M}$ and $(I_2, J_2) \in \mathcal{M}$ imply that $(I_1 \cup I_2, J_1 \cup J_2) \in \mathcal{M}$.
- \mathcal{M} is closed under target intersections if $J_1 \in \text{Sol}(I, \mathcal{M})$ and $J_2 \in \text{Sol}(I, \mathcal{M})$ imply that $(J_1 \cap J_2) \in \text{Sol}(I, \mathcal{M})$.
- \mathcal{M} is n -modular if whenever $(I, J) \notin \mathcal{M}$, there is a subinstance $I' \subseteq I$ with at most n elements in its active domain such that $(I', J) \notin \mathcal{M}$ (“small counterexample”).

Schema Mapping Languages A GLAV (*global-and-local-as-view*) constraint is a first-order formula 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 atoms over the source schema \mathbf{S} , each variable in \mathbf{x} occurs in at least one atom in $\varphi(\mathbf{x})$, and $\psi(\mathbf{x}, \mathbf{y})$ is a conjunction of atoms over the target schema \mathbf{T} with variables in \mathbf{x} and \mathbf{y} . We refer to $\varphi(\mathbf{x})$ as the *left-hand side*, or *premise*, and $\exists \mathbf{y}\psi(\mathbf{x}, \mathbf{y})$ as the *right-hand side*, or *conclusion* of the constraint. Another name for GLAV constraints is *source-to-target tuple-generating dependencies* or, in short, *s-t tgds*.

A LAV (*local-as-view*) constraint is a GLAV constraint whose left-hand side is a single atom over the source, while a GAV (*global-as-view*) constraint is a GLAV constraint whose right-hand side contains no existential quantifiers and consists of a single atom over the target. For example, $\forall x, y(E(x, y) \rightarrow \exists z(F(x, z) \wedge F(z, y)))$ is a LAV constraint, and $\forall x, y, z(E(x, z) \wedge E(z, y) \rightarrow F(x, y))$ is a GAV constraint.

A GLAV (*global-and-local-as-view*) mapping is a schema mapping $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ such that Σ is a finite set of GLAV constraints. The notions of a LAV mapping and of a GAV mapping are defined analogously.

Every GLAV mapping \mathcal{M} admits universal solutions [6]; furthermore, given a source instance I , a *canonical universal solution chase* $\text{chase}(I, \mathcal{M})$ can be produced via the *oblivious chase procedure* as follows: whenever the antecedent of an s-t tgd in \mathcal{M} becomes true, fresh null values are introduced and facts involving these nulls are added to $\text{chase}(I, \mathcal{M})$, so that the conclusion of the s-t tgd becomes true. Every GLAV mapping is also known to allow for CQ-rewriting and to be n -modular, for some $n \geq 1$. Moreover, every LAV mapping is closed under unions, while every GAV mapping is closed under target intersections.

Second-Order tgds, or *SO tgds*, were introduced in [8] and were shown to be exactly the constraints needed to express the composition of a finite number of GLAV

mappings. Instead of giving the precise definition of an SO tgd, we illustrate this notion with an example from [8]. The formula

$$\exists f(\forall e(\text{Emp}(e) \rightarrow \text{Mgr}(e, f(e))) \wedge \forall e(\text{Emp}(e) \wedge (e = f(e)) \rightarrow \text{SelfMgr}(e)))$$

expresses the property that every employee has a manager, and if an employee is the manager of himself/herself, then this employee is a self-manager. Clearly, SO tgds are existential second-order formulas with existentially quantified function symbols, which can be thought of as acting like Skolem functions. The use of these function symbols, however, is limited by the syntax of SO tgds: they can only appear in equations between terms in the antecedent of an implication or as arguments of atoms in the conclusion of an implication. As regards expressive power, SO tgds are, in general, strictly more expressive than GLAV constraints, but less expressive than arbitrary existential second-order formulas. In particular, the above formula is an SO tgd that is not logically equivalent to any (finite or infinite) set of GLAV constraints [8].

Every SO tgd allows for CQ-rewriting and admits universal solutions; however, an SO tgd may not be closed under target homomorphisms and there may not exist any $n \geq 1$ such that the SO tgd is n -modular (see [8, 19]).

Pseudometric Spaces and Metric Spaces A *pseudometric space* is a pair (X, d) , where X is a set and d is a function from $X \times X$ to the set R^+ of non-negative real numbers with the following properties: (i) $d(x, x) = 0$, for every x in X ; (ii) $d(x, y) = d(y, x)$, for every x and y in X ; (iii) $d(x, y) \leq d(x, z) + d(y, z)$, for every x, y, z in X (triangle inequality). A *metric space* is a pseudometric space (X, d) such that if $d(x, y) = 0$, then $x = y$. It is easy to see that if (X, d) is a pseudometric space, then the relation $R_d = \{(x, y) \in X \times X \mid d(x, y) = 0\}$ is an equivalence relation on X . From this, it follows that every pseudometric space (X, d) gives rise to a metric space $(\widehat{X}, \widehat{d})$, where \widehat{X} is the set of equivalence classes of elements of X modulo the equivalence relation R_d and $\widehat{d}([x], [y]) = d(x, y)$.

Let (X, d) be a pseudometric space. A sequence of elements x_1, x_2, \dots of X *converges* to an element x of X , denoted by $\lim_{n \rightarrow \infty} x_n = x$, if for every $\epsilon > 0$, there is an integer n_0 such that $d(x_n, x) < \epsilon$, for every $n \geq n_0$. We say that x is a *limit* of this sequence. The limit is unique if (X, d) is a metric space. A sequence x_1, x_2, \dots of elements of X is *Cauchy* if for every $\epsilon > 0$, there is an integer n_0 such that $d(x_n, x_{n'}) < \epsilon$, for every $n, n' \geq n_0$.

Using the triangle inequality, it is easy to see that if a sequence of elements in a (pseudo)metric space has a limit, then the sequence is Cauchy. The converse, however, does not hold for arbitrary (pseudo)metric spaces. A (pseudo)metric space (X, d) is *complete* if every Cauchy sequence of elements of X has a limit in X ; otherwise, it is *incomplete*.

It is well known that every incomplete (pseudo)metric space (X, d) can be embedded into a complete (pseudo)metric space (X^*, d^*) , called the *completion* of (X, d) , in such a way that X is a *dense* subset of X^* , i.e., every member of X^* is the limit of a sequence of members of X . The members of X^* are equivalence classes of Cauchy sequences of X , where two Cauchy sequences x_1, x_2, \dots and y_1, y_2, \dots of elements

of X are *equivalent* if $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$, while the distance function d^* is defined as $d^*([x_1, x_2, \dots], [y_1, y_2, \dots]) = \lim_{n \rightarrow \infty} d(x_n, y_n)$. The proof of correctness of this construction can be found in [18] or any other book on metric spaces.

As a concrete example, the metric space of the real numbers is the completion of the metric space of the rational numbers (both with the standard distance).

3 Metric Space of Target Instances

To study the limits of sequences of schema mappings, we first introduce a pseudometric space of sets of target instances. By considering schema mappings as functions that map each source instance to the set of its solutions, we can view sequences of schema mappings as *sequences of functions*. The (pointwise or uniform) limit of a sequence of schema mappings is then simply defined in the standard way as the limit of a sequence of functions taking values in a pseudometric space. Moreover, by passing to the associated metric space of equivalence classes of sets of target instances, we ensure the uniqueness of the limit. If \mathbf{T} is a schema, we write $\text{Inst}(\mathbf{T})$ for the set of all finite instances of \mathbf{T} . We also write $\mathcal{P}(\text{Inst}(\mathbf{T}))$ for the power set of $\text{Inst}(\mathbf{T})$. The notion of distance on $\mathcal{P}(\text{Inst}(\mathbf{T}))$ that we are about to introduce is heavily based on the notion of the certain answers to conjunctive queries and on the idea that two members \mathcal{J} and \mathcal{J}' of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ are “close” to each other if only “big” conjunctive queries can yield different certain answers on \mathcal{J} and \mathcal{J}' .

Definition 2 Let \mathbf{T} be a schema.

- Let q be a query over \mathbf{T} and let \mathcal{J} be a member of $\mathcal{P}(\text{Inst}(\mathbf{T}))$. The *certain answers* of q over \mathcal{J} are defined as

$$\text{cert}(q, \mathcal{J}) = \bigcap \{q(J) \mid J \in \mathcal{J}\}.$$

- We say that two sets of instances \mathcal{J} and \mathcal{J}' in $\mathcal{P}(\text{Inst}(\mathbf{T}))$ are *CQ-equivalent*, denoted $\mathcal{J} \equiv_{\text{CQ}} \mathcal{J}'$, if $\text{cert}(q, \mathcal{J}) = \text{cert}(q, \mathcal{J}')$ holds for all conjunctive queries q .
- We say that \mathcal{J} and \mathcal{J}' are *CQ_n-equivalent*, denoted $\mathcal{J} \equiv_{\text{CQ}_n} \mathcal{J}'$, if it holds that $\text{cert}(q, \mathcal{J}) = \text{cert}(q, \mathcal{J}')$ for all conjunctive queries q with at most n variables (i.e., for all q in CQ_n .)

Definition 3 Let \mathcal{J} and \mathcal{J}' be two sets of instances in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. The *similarity* $\text{sim}(\mathcal{J}, \mathcal{J}')$ and the *distance* $\text{dist}(\mathcal{J}, \mathcal{J}')$ between \mathcal{J} and \mathcal{J}' are defined as follows:

- $\text{sim}(\mathcal{J}, \mathcal{J}') = \max\{k \mid \mathcal{J} \equiv_{\text{CQ}_k} \mathcal{J}'\}$;
- $\text{dist}(\mathcal{J}, \mathcal{J}') = 2^{-\text{sim}(\mathcal{J}, \mathcal{J}')}.$

It is easy to verify that the pair $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is a pseudometric space; in fact, dist is an *ultrametric* distance function, that is,

$$\text{dist}(\mathcal{J}, \mathcal{J}') \leq \max\{\text{dist}(\mathcal{J}, \mathcal{J}''), \text{dist}(\mathcal{J}'', \mathcal{J}')\}$$

holds for all $\mathcal{J}, \mathcal{J}', \mathcal{J}''$ in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Moreover, $\text{dist}(\mathcal{J}, \mathcal{J}') = 0$ if and only if \mathcal{J} and \mathcal{J}' are CQ-equivalent.

Definition 4 Let \mathbf{T} be a schema. If J is a \mathbf{T} -instance, then we write $v(J)$ to denote the member of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ consisting of all isomorphic copies of J via isomorphisms that rename nulls. In other words, $v(J)$ consists of all \mathbf{T} -instances J' such that J' is isomorphic to J via an isomorphism h that maps each constant to itself and maps each null to a (possibly different) null.

The next lemma will be used repeatedly in the sequel.

Lemma 1 *Let \mathbf{T} be a schema.*

1. *If J is a \mathbf{T} -instance whose active domain consists entirely of nulls and q is a non-boolean conjunctive query, then $\text{cert}(q, v(J)) = \emptyset$.*
2. *If J is a \mathbf{T} -instance whose active domain consists entirely of nulls and q is a boolean conjunctive query, then $\text{cert}(q, v(J)) = q(J)$.*
3. *If J and J' are \mathbf{T} -instances whose active domains consist entirely of nulls, then, for every $k \geq 1$, the following statements are equivalent:*
 - (a) $v(J) \equiv_{\text{CQ}_k} v(J')$.
 - (b) J and J' satisfy the same boolean conjunctive queries in CQ_k .

Proof For the first two parts of the lemma, let J be a \mathbf{T} -instance whose active domain consists entirely of nulls. For every non-boolean query q in CQ_k , we have that $\text{cert}(q, v(J)) = \emptyset$, because $v(J)$ contains instances with disjoint active domains. For every boolean query q , we have $\text{cert}(q, v(J)) = q(J)$ for the following reason: first, J is a member of $v(J)$, so if $\text{cert}(q, v(J)) = \text{true}$, then $q(J) = \text{true}$ as well; second, since every member of $v(J)$ is isomorphic to J and since boolean conjunctive queries are preserved under isomorphisms, we have that if $q(J) = \text{true}$, then $\text{cert}(q, v(J)) = \text{true}$.

For the third part of the lemma, let J and J' be \mathbf{T} -instances whose active domains consist entirely of nulls and let k be a positive integer. If $v(J) \equiv_{\text{CQ}_k} v(J')$, then J and J' must satisfy the same boolean conjunctive queries in CQ_k because $J \in v(J)$ and $J' \in v(J')$. For the converse, assume that J and J' satisfy the same boolean conjunctive queries in CQ_k . We have to show that $\text{cert}(q, v(J)) = \text{cert}(q, v(J'))$, for every conjunctive query q in CQ_k . If q is a non-boolean conjunctive query in CQ_k , then, by the first part of the lemma, we have that $\text{cert}(q, v(J)) = \emptyset = \text{cert}(q, v(J'))$. If q is a boolean query in CQ_k , then, by the second part of the lemma and the hypothesis about J and J' , we have that $\text{cert}(q, v(J)) = q(J) = q(J') = \text{cert}(q, v(J'))$. □

The preceding lemma will be used in the next example, which presents a sequence from $\mathcal{P}(\text{Inst}(\mathbf{T}))$ that has a limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$.

Example 1 Let \mathbf{T} be a schema consisting of a single binary relation E and let C_m be the undirected cycle of length m , $m \geq 1$, where the vertices of the cycle are pairwise

distinct labeled nulls. Consider the sequence $(v(C_{2n+1}))_{n \geq 1}$ arising from the cycles of odd size. Then, for every $m \geq 1$, we have that $\lim_{n \rightarrow \infty} v(C_{2n+1}) = v(C_{2m})$. In particular, $\lim_{n \rightarrow \infty} v(C_{2n+1}) = v(C_2)$.

We first show that $v(C_{2m}) \equiv_{\text{CQ}} v(C_2)$, for every $m \geq 1$. By Lemma 1, it suffices to show that C_{2m} and C_2 satisfy the same boolean conjunctive queries. This is true because C_{2m} and C_2 are homomorphically equivalent (and boolean conjunctive queries are preserved under homomorphisms). Indeed, there is a homomorphism from C_2 to C_{2m} because C_2 is a subgraph of C_{2m} , and there is a homomorphism from C_{2m} to C_2 because C_{2m} is 2-colorable.

We will show that $\lim_{n \rightarrow \infty} v(C_{2n+1}) = v(C_2)$ by showing that for every k , there exists n_0 such that for all $n \geq n_0$, we have that $v(C_{2n+1}) \equiv_{\text{CQ}_k} v(C_2)$. For this, we take $n_0 = k$ and show that if $n \geq k$, then $v(C_{2n+1}) \equiv_{\text{CQ}_k} v(C_2)$. By the third part of Lemma 1, it suffices to show if q is a boolean conjunctive query in CQ_k , then $q(C_{2n+1}) = q(C_2)$. Since C_2 is a subgraph of C_{2n+1} , we have that if $q(C_2) = \text{true}$, then also $q(C_{2n+1}) = \text{true}$. Assume that $q(C_{2n+1}) = \text{true}$. Since $q \in \text{CQ}_k$, there is a subgraph H of C_{2n+1} with at most k distinct nodes such that $q(H) = \text{true}$. Since $2n + 1 > n \geq k$, we have that H is a proper subgraph of C_{2n+1} . Consequently, H is 2-colorable and so there is a homomorphism from H to C_2 , which, in turn, implies that $q(C_2) = \text{true}$.

In contrast to what we have just seen, there are Cauchy sequences of elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ that have no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Thus, the pseudo-metric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is incomplete.

Proposition 2 *Let \mathbf{T} be a schema consisting of a single binary relation E and let K_n be the clique of size n , for $n \geq 1$, where the vertices are pairwise distinct labeled nulls. The sequence $(v(K_n))_{n \geq 1}$ is Cauchy, but has no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$.*

Proof The sequence $(v(K_n))_{n \geq 1}$ is Cauchy, because if $m \geq n$, then $v(K_m)$ and $v(K_n)$ satisfy the same conjunctive queries in CQ_n . To show this, by the third part of Lemma 1, it suffices to show that if $m \geq n$, then K_m and K_n satisfy the same boolean conjunctive queries in CQ_n . Let q be a boolean conjunctive query in CQ_n . Since K_n is a subgraph of K_m , if $q(K_n) = \text{true}$, then $q(K_m) = \text{true}$. Conversely, if $q(K_m) = \text{true}$, then there is a subgraph H of K_m with at most n distinct nodes such that $q(H) = \text{true}$. But then H is also a subgraph of K_n , hence $q(K_n) = \text{true}$.

It remains to show that the sequence $(v(K_n))_{n \geq 1}$ has no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Assume to the contrary that there does exist a set \mathcal{J} of finite instances over \mathbf{T} such that $\lim_{n \rightarrow \infty} v(K_n) = \mathcal{J}$. We distinguish three cases.

First, if $\mathcal{J} = \emptyset$, then $\text{cert}(q, \mathcal{J}) = \text{true}$, for every conjunctive query q . In particular, this holds for the query $q = \exists x E(x, x)$, which asserts the existence of a self-loop. In contrast, for this conjunctive query, we have that $\text{cert}(q, v(K_n)) = \text{false}$, for every $n \geq 1$, since $K_n \in v(K_n)$ and none of the graphs $K_n, n \geq 1$ contains a self-loop.

Second, if $\mathcal{J} \neq \emptyset$ and if every member J of \mathcal{J} contains a self-loop, then we again consider the query $q = \exists x E(x, x)$. We thus have $\text{cert}(q, \mathcal{J}) = \text{true}$, whereas $\text{cert}(q, v(K_n)) = \text{false}$, for every $n \geq 1$.

It remains to consider the case that $\mathcal{J} \neq \emptyset$ and at least one member $J \in \mathcal{J}$ does not contain a self-loop. Let m be the biggest integer such that J contains a clique of size m . We define the query q as

$$q = \exists x_1, \dots, x_{m+1} \bigwedge_{i \neq j} (E(x_i, x_j) \wedge E(x_j, x_i)).$$

For graphs without self-loops, q asserts the existence of a clique of size $m + 1$. We now have that q evaluates to false overy J . Hence, $\text{cert}(q, \mathcal{J}) = \text{false}$ holds, while $\text{cert}(q, v(K_n)) = \text{true}$, for every $n \geq m + 1$. □

Since $(v(K_n)_{n \geq 1})$ is a Cauchy sequence, it has a limit in the completion of $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$. As we will see in Section 6, a concrete representation of this limit is the set consisting of all disjoint unions of cliques of all finite sizes in which every node is a null.

The following definitions are perfectly meaningful for every pseudometric space (X, d) and for every sequence of functions taking values in X . For concreteness, we give the definitions for sequences of functions taking values in $\mathcal{P}(\text{Inst}(\mathbf{T}))$.

Definition 5 Let A be a set, let $(f_n)_{n \geq 1}$ be a sequence of functions from A to $\mathcal{P}(\text{Inst}(\mathbf{T}))$, and let f be a function from A to $\mathcal{P}(\text{Inst}(\mathbf{T}))$.

- We say that $(f_n)_{n \geq 1}$ *converges pointwise* to f , denoted as $\lim_{n \rightarrow \infty}^p f_n = f$, if for every element $x \in A$, we have that $\lim_{n \rightarrow \infty} f_n(x) = f(x)$.
- We say that $(f_n)_{n \geq 1}$ *converges uniformly* to f , denoted as $\lim_{n \rightarrow \infty}^u f_n = f$, if for every $\epsilon > 0$, there exists an integer $n_0 \geq 1$ such that for every integer $n \geq n_0$ and for every element $x \in A$, we have $\text{dist}(f_n(x), f(x)) < \epsilon$.
- We say that $(f_n)_{n \geq 1}$ is *pointwise Cauchy*, if for every element $x \in A$, the sequence $(f_n(x))_{n \geq 1}$ is Cauchy.
- We say that $(f_n)_{n \geq 1}$ is *uniformly Cauchy*, if for every $\epsilon > 0$, there exists an integer $n_0 \geq 1$ such that for all integers $n, n' \geq n_0$ and for every element $x \in A$, we have $\text{dist}(f_n(x), f_{n'}(x)) < \epsilon$.

Clearly, if $(f_n)_{n \geq 1}$ converges pointwise (resp., uniformly), then $(f_n)_{n \geq 1}$ is pointwise (resp., uniformly) Cauchy. The converse is not in general true for arbitrary (pseudo)metric spaces; in particular, it is not true for the pseudometric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$, as we shall see later on.

We now bring schema mappings into the picture. Every schema mapping \mathcal{M} over a source schema \mathbf{S} and a target schema \mathbf{T} can be identified with a function $f: \text{Inst}(\mathbf{S}) \rightarrow \mathcal{P}(\text{Inst}(\mathbf{T}))$, where $f(I) = \text{Sol}(I, \mathcal{M})$ (recall that $\text{Sol}(I, \mathcal{M})$ is the set of all solutions of I w.r.t. \mathcal{M} , i.e., the set of all finite \mathbf{T} instances J such that $(I, J) \in \mathcal{M}$). Thus, a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings over a source schema \mathbf{S} and target schema \mathbf{T} can be viewed as a sequence of functions from $\text{Inst}(\mathbf{S})$ to $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Therefore, we can talk about a sequence of schema mappings being pointwise Cauchy and uniformly Cauchy if the sequence of the associated functions

has these properties. Similarly, we say that a sequence of schema mappings has a pointwise limit (resp., a uniform limit) if the sequence of the associated functions converges pointwise (resp., converges uniformly) to a schema mapping.

The preceding notion of convergence of a sequence of schema mappings allows us to draw a connection to earlier work on schema mapping optimization [5, 7]. Here, we are considering CQ-equivalence and CQ_n-equivalence of sets of instances. In previous works, these notions of equivalence have been mainly applied to schema mappings (see, e.g., [5, 7, 14]). Specifically, two schema mappings $\mathcal{M}, \mathcal{M}'$ are CQ-equivalent (resp., CQ_n-equivalent) if for every target conjunctive query q (resp., every target conjunctive query q in CQ_n) and every source instance I , we have that $cert(q, I, \mathcal{M}) = cert(q, I, \mathcal{M}')$. In this case, we write $\mathcal{M} \equiv_{CQ} \mathcal{M}'$ (resp., $\mathcal{M} \equiv_{CQ_n} \mathcal{M}'$). The notion of CQ_n-equivalence has been studied in the context of schema mapping optimization [5, 7]. Below we discuss its relationship to the convergence of schema mappings.

Proposition 3 Consider a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings and a schema mapping \mathcal{M} . Then $\lim_{n \rightarrow \infty} \mathcal{M}_n = \mathcal{M}$ if and only if for every integer $k \geq 1$, there is an integer $n_0 \geq 1$ such that for all integers $n \geq n_0$, we have that $\mathcal{M}_n \equiv_{CQ_k} \mathcal{M}$.

Proof The result follows by unfolding and comparing the definitions. Specifically, $\lim_{n \rightarrow \infty} \mathcal{M}_n = \mathcal{M}$ means that for every $\epsilon > 0$, there is an integer $n_0 \geq 1$ such that for every integer $n \geq n_0$ and for every source instance I we have that $dist(Sol(I, \mathcal{M}_n), Sol(I, \mathcal{M})) < \epsilon$. In turn, this means that for every integer $k \geq 1$, there is an integer $n_0 \geq 1$ such that for every integer $n \geq n_0$ and for every source instance I we have that $Sol(I, \mathcal{M}_n) \equiv_{CQ_k} Sol(I, \mathcal{M})$. Thus, for every integer $k \geq 1$, there is an integer $n_0 \geq 1$ such that for every integer $n \geq n_0$, we have that $\mathcal{M}_n \equiv_{CQ_k} \mathcal{M}$. □

Intuitively, the preceding proposition states that it takes bigger and bigger conjunctive queries to distinguish the members of a sequence $(\mathcal{M}_n)_{n \geq 1}$ from its uniform limit.

Although never explicitly introduced, the notion of uniform convergence was implicit in [5], where it was shown that for every SO tgd σ and for every $n \geq 1$, there is a GLAV mapping \mathcal{M}_n such that $\sigma \equiv_{CQ_n} \mathcal{M}_n$. From this, it is easy to see that $\lim_{n \rightarrow \infty} \mathcal{M}_n = \sigma$. Thus, we have the following result.

Theorem 1 (implicit in [5]) Every SO tgd is a uniform limit of a sequence of GLAV mappings.

There are SO tgds that are not CQ-equivalent to any GLAV mapping. Indeed, from Example 4.6 and Theorem 4.10 in [7], it follows that the SO-tgd

$$\exists f \forall x \forall y (S(x, y) \rightarrow T(x, f(y)))$$

is not CQ-equivalent to any GLAV mapping. Thus, the point of Theorem 1 is that SO tgds can be “approximated” up to any level of CQ_k-equivalence by GLAV mappings, which are both syntactically simpler and generally more well-behaved.

As stated earlier, $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is a pseudometric space since it cannot distinguish CQ-equivalent sets of instances. Consequently, the limit of a sequence of sets of instances and the (uniform or pointwise) limit of a sequence of mappings need not be unique. However, the limit is unique up to CQ-equivalence and, as described in Section 2, there is an associated metric space $(\widehat{\mathcal{P}(\text{Inst}(\mathbf{T}))}, \widehat{\text{dist}})$ obtained by considering the equivalence classes of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ modulo the equivalence relation R_{dist} , where $(\mathcal{J}, \mathcal{J}') \in R_{\text{dist}}$ if and only if $\text{dist}(\mathcal{J}, \mathcal{J}') = 0$ (i.e., if and only if $\mathcal{J} \equiv_{\text{CQ}} \mathcal{J}'$).

In subsequent sections, we will work with the metric space $(\widehat{\mathcal{P}(\text{Inst}(\mathbf{T}))}, \widehat{\text{dist}})$. Moreover, we will be interested in schema mappings modulo CQ-equivalence, which means that from now on we will view schema mappings as functions from source instances to equivalence classes of sets of target instances modulo CQ-equivalence. However, for notational simplicity, we will work each time with representatives of the equivalence classes. By a slight abuse of notation, we will write $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$, instead of $(\widehat{\mathcal{P}(\text{Inst}(\mathbf{T}))}, \widehat{\text{dist}})$. Likewise, we will not explicitly distinguish between a schema mapping \mathcal{M} and the equivalence class of the schema mappings that are CQ-equivalent to \mathcal{M} .

4 Limits of Sequences of GAV Mappings

Our goal in this section is to analyze sequences of GAV mappings. To this effect, we first investigate the existence of limits of such sequences and then examine the definability of limits. As discussed in Section 3, if a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings has a pointwise (resp., uniform) limit, then the sequence is pointwise (resp., uniformly) Cauchy. The next result asserts that the converse holds for sequences of GAV mappings.

Theorem 2 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of GAV mappings.*

- *If $(\mathcal{M}_n)_{n \geq 1}$ is pointwise Cauchy, then it has a pointwise limit.*
- *If $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy, then it is eventually constant and thus has a GAV schema mapping as a uniform limit.*

Proof We consider GAV mappings over a source schema \mathbf{S} and a target schema \mathbf{T} . Let r denote the maximum arity of the relation symbols in \mathbf{T} . For showing the first claim, assume that $(\mathcal{M}_n)_{n \geq 1}$ is a pointwise Cauchy sequence of schema mappings and let I be a source instance. For each $n \geq 1$, consider the universal solution $\text{chase}(I, \mathcal{M}_n)$ for I w.r.t. \mathcal{M}_n obtained by using the oblivious chase procedure. Since each \mathcal{M}_n is a GAV schema mapping, we have that $\text{chase}(I, \mathcal{M}_n)$ contains constants from the active domain of I and no nulls. We claim that there exists some n_0 such that for all $n \geq n_0$, we have that $\text{chase}(I, \mathcal{M}_n) = \text{chase}(I, \mathcal{M}_{n_0})$. In other words, we

claim that the sequence $(chase(I, \mathcal{M}_n))_{n \geq 1}$ is eventually constant (does not oscillate). Since every instance in the sequence $(chase(I, \mathcal{M}_n))_{n \geq 1}$ has no nulls, it can be identified by evaluating on that instance the atomic queries $R(x_1, \dots, x_k)$, where R ranges over the relation symbols of \mathbf{T} and k (with $k \leq r$) denotes the arity of R . The assumption that the sequence $(chase(I, \mathcal{M}_n))_{n \geq 1}$ is pointwise Cauchy implies that there exists a positive integer n_0 (that depends on I and r) such that for every integer $n \geq n_0$ and every conjunctive query $q \in \mathbf{CQ}_r$, we have that $cert(q, I, \mathcal{M}_n) = cert(q, I, \mathcal{M})$. This implies that $q(chase(I, \mathcal{M}_n)) = q(chase(I, \mathcal{M}_{n_0}))$ and, consequently, for every $n \geq n_0$, we have that $chase(I, \mathcal{M}_n) = chase(I, \mathcal{M}_{n_0})$.

We have just shown that if $(\mathcal{M}_n)_{n \geq 1}$ is a pointwise Cauchy sequence of GAV mappings, then for every I , there exists a positive integer m_I such that $chase(I, \mathcal{M}_{m_I}) = chase(I, \mathcal{M}_n)$, for all $n \geq m_I$. It follows that the schema mapping $\mathcal{M} = \{(I, chase(I, \mathcal{M}_{m_I})) \mid I \text{ is a source instance}\}$ is a pointwise limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$. Note that \mathcal{M} is indeed a schema mapping because $chase(I, \mathcal{M}_{m_I})$ contains no nulls.

For showing the second claim, assume that $(\mathcal{M}_n)_{n \geq 1}$ is a uniformly Cauchy sequence of GAV mappings. We claim that $(\mathcal{M}_n)_{n \geq 1}$ is eventually constant, i.e., there is some n_0 such that for all $n \geq n_0$, $\mathcal{M}_n \equiv_{\mathbf{CQ}} \mathcal{M}_{n_0}$ holds. For this, we repeat the previous argument, but also note that, since the sequence $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy, there exists a positive integer n_0 that depends only on r such that for every source instance I , for every integer $n \geq n_0$ and every conjunctive query $q \in \mathbf{CQ}_r$, we have that $cert(q, I, \mathcal{M}_n) = cert(q, I, \mathcal{M})$. This implies that for every source instance I and every $n \geq n_0$, we have that $q(chase(I, \mathcal{M}_n)) = q(chase(I, \mathcal{M}_{n_0}))$; consequently, for every source instance I and every $n \geq n_0$, we have that $chase(I, \mathcal{M}_n) = chase(I, \mathcal{M}_{n_0})$. □

Next, we point out that, for sequences of GAV mappings, the notions of pointwise convergence and uniform convergence are genuinely different.

Proposition 4 *There exists a sequence of GAV mappings that has a GAV mapping as a pointwise limit, but has no uniform limit.*

Proof For every $n \geq 2$, let $q_n = \bigwedge_{1 \leq i < j \leq n} (E(x_i, x_j) \wedge E(x_j, x_i))$. Intuitively, if E is interpreted as edge relation, then q_n yields a non-empty answer over any graph that contains a self-loop or a clique of size n . Let \mathbf{S} be a source schema consisting of a binary relation symbol E and a unary relation symbol P , let \mathbf{T} be a target schema consisting of a unary relation symbol P' . Let $(\mathcal{M}_n)_{n \geq 1}$ be the sequence of GAV mappings, where \mathcal{M}_n is specified by the constraint $\forall x \forall x_1, \dots, x_{n+1} (P(x) \wedge q_{n+1} \rightarrow P'(x))$. Intuitively, \mathcal{M}_n is a “copy” schema mapping, but the copying action is triggered only if the source instance contains a self-loop or a clique of size $n + 1$. We will show that the GAV schema mapping $\mathcal{M} = \{\forall x \forall y (P(x) \wedge E(y, y) \rightarrow P'(x))\}$ is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$, but that this pointwise limit is not a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$ and thus no uniform limit of $(\mathcal{M}_n)_{n \geq 1}$ exists.

We first show that the GAV mapping \mathcal{M} is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. Given a source instance I , we consider two cases.

- If I contains a self-loop, then $J = \{P'(x) \mid P(x) \in I\}$ is a universal solution for I w.r.t. \mathcal{M} and w.r.t. \mathcal{M}_n , for all n . Thus, $\text{cert}(q, I, \mathcal{M}) = \text{cert}(q, I, \mathcal{M}_n)$, for all n .
- If I is self-loop free, let n_0 be such that no clique larger than n_0 exists in I . Then, $J = \emptyset$ is a universal solution for I w.r.t. \mathcal{M} and w.r.t. \mathcal{M}_n , for all $n \geq n_0$. Thus, $\text{cert}(q, I, \mathcal{M}) = \text{cert}(q, I, \mathcal{M}_n)$, for all $n \geq n_0$.

Next, we show that $(\mathcal{M}_n)_{n \geq 1}$ has no uniform limit. Towards a contradiction, suppose that such a uniform limit exists. Every uniform limit is also a pointwise limit; moreover, pointwise and uniform limits are unique up to CQ-equivalence. Hence, since the schema mapping \mathcal{M} defined above is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$, it follows that \mathcal{M} is also a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$. Let $m = 1$. Then there exists an n_0 such that for all $n \geq n_0$ we have that $\mathcal{M}_n \equiv_{\text{CQ}_1} \mathcal{M}$. Take $n = n_0$. Let I be the source instance $K_{n+1} \cup \{P(c)\}$ and let q be the target conjunctive query $\exists x P'(x)$. We now claim that $\text{cert}(q, I, \mathcal{M}_n) \neq \text{cert}(q, I, \mathcal{M})$, which contradicts the previously derived fact that $\mathcal{M}_n \equiv_{\text{CQ}_1} \mathcal{M}$. Indeed, since I contains a clique of size $n + 1$, we have $P(c)$ is a universal solution for I w.r.t. \mathcal{M}_n , hence $\text{cert}(q, I, \mathcal{M}_n) = \text{true}$. However, since I contains no self-loop, we have that \emptyset is a universal solution for I w.r.t. \mathcal{M} , hence $\text{cert}(q, I, \mathcal{M}) = \text{false}$. □

Proposition 4 and Theorem 2 imply that the sequence of GAV mappings in the proof of Proposition 4 is an example of a pointwise Cauchy sequence that is not uniformly Cauchy. Theorem 2 also implies that if a sequence of GAV mappings has a uniform limit, then it must have a GAV mapping as such a limit. In turn, this gives rise to the following natural question concerning the definability of pointwise limits: if a sequence of GAV mappings has a pointwise limit, does it have a GAV mapping as such a limit? We answer this question in the negative by showing that even the much richer language of SO tgds cannot express pointwise limits of sequences of GAV mappings.

Proposition 5 *There is a pointwise Cauchy sequence of GAV schema mappings such that no SO tgd is a pointwise limit of that sequence.*

Proof Consider a source schema \mathbf{S} consisting of a binary relation symbol E , and a target schema \mathbf{T} consisting of a binary relation F . For every $n \geq 1$, let $P_n(x, y)$ be the conjunctive query expressing the property “there is an E -path of length n from x to y ”, and let \mathcal{M}_n be the GAV mapping specified by the set $\{\forall x, y (P_i(x, y) \rightarrow F(x, y)) \mid 1 \leq i \leq n\}$. Consider the schema mapping

$$\mathcal{M}^* = \{(I, J) \mid F^J \text{ contains the transitive closure } TC(I) \text{ of } E^I\}.$$

It is easy to see that \mathcal{M}^* is a pointwise limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$; the reason for this is that, for every source instance I and for every $n \geq |\text{adom}(I)|^2$, we have

that $chase(I, \mathcal{M}_n) = TC(I)$. However, \mathcal{M}^* is not CQ-equivalent to any schema mapping \mathcal{M}' that allows for CQ-rewriting: if it were, then there would exist a union q of conjunctive queries over the source such that, for every source instance I ,

$$cert(F(x, y), I, \mathcal{M}^*) = TC(I) = cert(F(x, y), I, \mathcal{M}') = q(I).$$

Consequently, the transitive closure of I would be first-order definable over the source, which is not the case. Since every SO tgd allows for CQ-rewriting, no SO tgd is a pointwise limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$. □

We have just seen that there are sequences of GAV mappings that have a pointwise limit, but no such limit is definable by a GAV mapping. This raises the question of finding necessary and sufficient conditions guaranteeing that a sequence of GAV mappings has a GAV mapping as a pointwise limit. The next result provides an answer to this question.

Theorem 3 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of GAV mappings. The following statements are equivalent:*

1. $(\mathcal{M}_n)_{n \geq 1}$ has a GAV mapping as a pointwise limit.
2. $(\mathcal{M}_n)_{n \geq 1}$ has a pointwise limit that allows for CQ-rewriting.

Proof Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of schema mappings. As shown in the proof of Theorem 2, for every source instance I , there is a positive integer m_I , such that for all $n \geq m_I$ the equality $chase(I, \mathcal{M}_{m_I}) = chase(I, \mathcal{M}_n)$ holds for the respective elements \mathcal{M}_{m_I} and \mathcal{M}_n of $(\mathcal{M}_n)_{n \geq 1}$. Moreover, the schema mapping

$$\mathcal{M} = \{(I, chase(I, \mathcal{M}_{m_I})) \mid I \text{ is a source instance}\}$$

is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. Consider the following schema mapping \mathcal{M}^* :

$$\mathcal{M}^* = \{(I, J) \mid chase(I, \mathcal{M}_{m_I}) \subseteq J\}$$

It is clear that \mathcal{M}^* is also a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. The result we seek is an immediate consequence of the fact that the following four statements are equivalent:

- (a) $(\mathcal{M}_n)_{n \geq 1}$ has a GAV mapping as a pointwise limit.
- (b) $(\mathcal{M}_n)_{n \geq 1}$ has a pointwise limit that allows for CQ-rewriting.
- (c) \mathcal{M}^* allows for CQ-rewriting.
- (d) \mathcal{M}^* is logically equivalent to a GAV mapping.

We now show that these four conditions are equivalent.

- (a) \Rightarrow (b) This is true because every GAV mapping allows for CQ-rewriting.
- (b) \Rightarrow (c) This is true because if \mathcal{M}' is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$ that allows for CQ-rewriting, then so does \mathcal{M}^* since $\mathcal{M}' \equiv_{\text{CQ}} \mathcal{M}^*$.
- (c) \Rightarrow (d) This is the most involved part of the proof. Let us examine the structural properties that the schema mapping \mathcal{M}^* possesses. By hypothesis, \mathcal{M}^* allows for CQ-rewriting. By construction, \mathcal{M}^* admits universal solutions, since

$chase(I, \mathcal{M}_{m_I})$ is a universal solution for I w.r.t. \mathcal{M}^* , for every source instance I . Moreover, it is clear from its definition that \mathcal{M}^* is closed under target homomorphisms. Finally, we claim that \mathcal{M}^* is closed under target intersections. Indeed, assume that both (I, J_1) and (I, J_2) are in \mathcal{M}^* . Then $chase(I, \mathcal{M}_{m_I})$ is contained in both J_1 and J_2 , hence $chase(I, \mathcal{M}_{m_I})$ is contained in $J_1 \cap J_2$, hence $J_1 \cap J_2$ is a solution for I w.r.t. \mathcal{M}^* .

Thus, \mathcal{M}^* allows for CQ-rewriting, admits universal solutions, and is closed under both target homomorphisms and target intersections. Theorem 3.2 in [19] asserts that a schema mapping is logically equivalent to a GAV schema mapping if and only if it allows for CQ-rewriting, admits universal solutions, and is closed under both target homomorphisms and target intersections. It follows that \mathcal{M}^* is logically equivalent to a GAV mapping.

(d) \Rightarrow (a) This is obvious since \mathcal{M}^* is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. □

Observe that Theorem 3 (and its proof) provide necessary and sufficient conditions for a pointwise Cauchy sequence of GAV mappings to have a GAV mapping as a pointwise limit, but these conditions are on the pointwise limit and not on the sequence itself. By analyzing the proof of Theorem 3, however, it is possible to extract a necessary and sufficient condition on the sequence itself. For this, we need to introduce the following concept.

Definition 6 Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of schema mappings. We say that $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting if for every target conjunctive query q , there is a union q' of source conjunctive queries having the following property: for every source instance I , there is a positive integer n_I such that $cert(q, I, \mathcal{M}_n) = q'(I)$, for every $n \geq n_I$.

Let \mathcal{M} be a pointwise limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings. It is easy to show that \mathcal{M} allows for CQ-rewriting if and only if $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting. Indeed, assume first that \mathcal{M} allows for CQ-rewriting. To show that $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting, let q be a conjunctive query and let q' be a union of conjunctive queries such that $cert(q, I, \mathcal{M}) = q'(I)$, for every source instance I . Since \mathcal{M} is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$, for every instance I , there is a positive integer n'_I such that $cert(q, I, \mathcal{M}) = cert(q, I, \mathcal{M}_n)$, for every $n \geq n'_I$. It follows that $cert(q, I, \mathcal{M}_n) = q'(I)$, for every $n \geq n'_I$, which shows that $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting. In the other direction, assume that $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting. To show that \mathcal{M} allows for CQ-rewriting, let q be a conjunctive query and let q' be a union of conjunctive queries such that for every source instance I , there is a positive integer n_I such that $cert(q, I, \mathcal{M}_n) = q'(I)$, for every $n \geq n_I$. By the pointwise convergence of $(\mathcal{M}_n)_{n \geq 1}$ to \mathcal{M} , for every source instance I , there is a positive integer n'_I such that $cert(q, I, \mathcal{M}) = cert(q, I, \mathcal{M}_n)$, for every $n \geq n'_I$. Let I be a source instance. By taking any $n \geq \max\{n_I, n'_I\}$, we have that $cert(q, I, \mathcal{M}) = cert(q, I, \mathcal{M}_n)$ and $cert(q, I, \mathcal{M}_n) = q'(I)$, hence $cert(q, I, \mathcal{M}) = q'(I)$, which shows that \mathcal{M} allows for CQ-rewriting.

By combining the preceding remarks with Theorems 2 and 3, we obtain the following result.

Corollary 1 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of GAV mappings. The following statements are equivalent:*

1. $(\mathcal{M}_n)_{n \geq 1}$ has a GAV mapping as a pointwise limit.
2. $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting.

Since every schema mapping specified by an SO tgd allows for CQ-rewriting, Theorem 3 also implies the following result.

Corollary 2 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of GAV mappings. The following statements are equivalent:*

1. $(\mathcal{M}_n)_{n \geq 1}$ has a GAV mapping as a pointwise limit.
2. $(\mathcal{M}_n)_{n \geq 1}$ has an SO tgd as a pointwise limit.

Finally, we note that Proposition and Theorem 3 yield a fairly complete picture of the definability of pointwise limits of GAV mappings. Specifically, there are two mutually exclusive possibilities:

- (1) No pointwise limit allows for CQ-rewriting and no GAV mapping is a pointwise limit.
- (2) Every pointwise limit admits CQ-rewriting and there is a GAV mapping that is a pointwise limit. Moreover, this happens precisely when the schema mapping \mathcal{M}^* in the proof of Theorem 3 allows for CQ-rewriting or, equivalently, when \mathcal{M}^* is logically equivalent to a GAV mapping.

5 Limits of Sequences of LAV Mappings

In this section, we investigate the existence and definability of limits of sequences of LAV mappings. In fact, we will consider a much broader class of GLAV mappings, namely *k-premise-bounded* GLAV mappings for arbitrary $k \geq 1$. LAV mappings correspond to the special case of $k = 1$.

Definition 7 Let \mathcal{M} be a GLAV mapping and k a positive integer. We call \mathcal{M} a *k-premise-bounded* GLAV mapping if the premise of every constraint in \mathcal{M} has at most k atoms.

Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of GLAV mappings. We say that $(\mathcal{M}_n)_{n \geq 1}$ is *premise-bounded* if there exists an integer k such that every element \mathcal{M}_n of $(\mathcal{M}_n)_{n \geq 1}$ is k -premise bounded.

Unlike the case of GAV mappings, the notions of pointwise Cauchy and uniformly Cauchy sequences of premise-bounded GLAV mappings coincide. Moreover, the same holds true for the notions of pointwise limit and uniform limit of sequences of such schema mappings.

Theorem 4 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of premise-bounded GLAV mappings.*

- (1) The sequence $(\mathcal{M}_n)_{n \geq 1}$ is pointwise Cauchy if and only if it is uniformly Cauchy.
- (2) The sequence $(\mathcal{M}_n)_{n \geq 1}$ has a pointwise limit if and only if it has a uniform limit.

Proof We prove the first part and then use it to prove the second part.

Part 1. It is obvious that every uniformly Cauchy sequence of mappings is also pointwise Cauchy. We focus on the reverse direction. Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of premise bounded GLAV mappings. We have to show that for every m , there is an N_0 such that for all $n, n' \geq N_0$, we have that $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}_{n'}$.

Fix an integer m . Since $(\mathcal{M}_n)_{n \geq 1}$ is pointwise Cauchy, for every source instance I , there is an integer $n_0(I)$ such that for all $n, n' \geq n_0(I)$ and for every conjunctive query q in CQ_m , we have that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M}_{n'})$. Let p be the number of relation symbols in the target schema, let r be their maximum arity, and let k be the bound on the number of atoms in the premises of the members of the sequence $(\mathcal{M}_n)_{n \geq 1}$. We write \mathcal{I} to denote the class of all source instances with at most $k \cdot p \cdot m^r$ atoms. Clearly, up to isomorphism, there are only finitely many instances $I \in \mathcal{I}$. Moreover, if $I' \cong I''$, then $n_0(I') = n_0(I'')$. Consequently, the quantity $N_0 = \max\{n_0(I) \mid I \in \mathcal{I}\}$ is a positive integer. We claim that for all $n, n' \geq N_0$, we have that $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}_{n'}$.

Let I be an arbitrary source instance and let q be an arbitrary conjunctive query in CQ_m . We have to show that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M}_{n'})$, for all $n, n' \geq N_0$. Let \mathbf{a} be a tuple of constants such that $\mathbf{a} \in \text{cert}(q, I, \mathcal{M}_n)$, hence $\mathbf{a} \in q(\text{chase}(I, \mathcal{M}_n))_{\downarrow}$. Since the query q has at most m variables, it must consist of at most $p \cdot m^r$ atoms. Let $h : \text{atoms}(q) \rightarrow \text{chase}(I, \mathcal{M}_n)$ be a homomorphism establishing that $\mathbf{a} \in q(\text{chase}(I, \mathcal{M}_n))$. It follows that there are at most $p \cdot m^r$ facts in $\text{chase}(I, \mathcal{M}_n)$ witnessing that $\mathbf{a} \in q(\text{chase}(I, \mathcal{M}_n))_{\downarrow}$. Each of these facts must be produced in a single step while chasing the source instance I with \mathcal{M}_n , which implies that each of these facts is produced using at most k facts from I . Let I^* be the subinstance of I consisting of all the aforementioned facts of I used to produce the facts in $\text{chase}(I, \mathcal{M}_n)$ witnessing that $\mathbf{a} \in q(\text{chase}(I, \mathcal{M}_n))_{\downarrow}$. We then have that $|I^*| \leq k \cdot p \cdot m^r$ and $\mathbf{a} \in q(\text{chase}(I^*, \mathcal{M}_n))_{\downarrow}$. Since $n, n' \geq N_0$, we have that $q(\text{chase}(I^*, \mathcal{M}_n))_{\downarrow} = q(\text{chase}(I^*, \mathcal{M}'_n))_{\downarrow}$, hence $\mathbf{a} \in q(\text{chase}(I^*, \mathcal{M}'_n))_{\downarrow}$. By the monotonicity of the chase procedure, we have that $\mathbf{a} \in q(\text{chase}(I, \mathcal{M}'_n))_{\downarrow}$. It follows that $q(\text{chase}(I, \mathcal{M}_n))_{\downarrow} \subseteq q(\text{chase}(I, \mathcal{M}'_n))_{\downarrow}$. A symmetric argument establishes the containment $q(\text{chase}(I, \mathcal{M}'_n))_{\downarrow} \subseteq q(\text{chase}(I, \mathcal{M}_n))_{\downarrow}$, hence $q(\text{chase}(I, \mathcal{M}_n))_{\downarrow} = q(\text{chase}(I, \mathcal{M}'_n))_{\downarrow}$, which, in turn, implies that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M}_{n'})$.

Part 2. It is obvious that if a sequence of schema mappings has a uniform limit, then it has a pointwise limit. We focus on the reverse direction. Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of premise bounded GLAV mappings that has a pointwise limit \mathcal{M} . We claim that \mathcal{M} is also a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$.

Since $(\mathcal{M}_n)_{n \geq 1}$ has a pointwise limit, we have that $(\mathcal{M}_n)_{n \geq 1}$ is pointwise Cauchy. The previous part implies that $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy as well. Fix an integer m . Since $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy, there exists an n_0 such that for all $n, n' \geq n_0$, we have that $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}_{n'}$. We claim that also $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}$ holds, for every

$n \geq n_0$. To show this, fix some $n \geq n_0$ and let I be a source instance and q a conjunctive query in \mathbf{CQ}_m . We have to show that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. Since \mathcal{M} is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$, there is an $n'_0(I)$ such that for all $n' \geq n'_0(I)$, we have that $\text{cert}(q, I, \mathcal{M}_{n'}) = \text{cert}(q, I, \mathcal{M})$. Take an integer n' such that $n' \geq \max\{n_0, n'_0(I)\}$. Since $n' \geq n_0$, we have that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M}_{n'})$. Since $n' \geq n'_0(I)$, we have that $\text{cert}(q, I, \mathcal{M}_{n'}) = \text{cert}(q, I, \mathcal{M})$. Thus, $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. \square

Note that the preceding proof of Part 2 used only the hypothesis that the sequence $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy and the fact that the sequence $(\mathcal{M}_n)_{n \geq 1}$ has a pointwise limit, as we have proved in Part 1. As a matter of fact, this is an instance of a general result about pseudometric spaces, namely, that if a uniformly Cauchy sequence of functions converges pointwise, then it also converges uniformly.

The following two propositions further demarcate the differences between GAV and premise-bounded GLAV mappings. In fact, these differences are already witnessed by sequences of LAV mappings. The first difference concerns the existence of limits of uniformly Cauchy sequences. In contrast to the GAV case, uniformly Cauchy sequences of LAV mappings may have no uniform limit; in fact, they may not even have a pointwise limit.

Proposition 6 *There exists a uniformly Cauchy sequence of LAV mappings that has no pointwise limit; in particular, it has no uniform limit either.*

Proof Let \mathbf{S} be a source schema consisting of a binary relation symbol E and let \mathbf{T} be a target schema consisting of a binary relation F . For every $n \geq 1$, let \mathcal{M}_n be the LAV mapping specified by the constraint

$$\forall x, y (E(x, y) \rightarrow q_{n+1})$$

where $q_n = \exists z_1, \dots, z_n \bigwedge_{1 \leq i < j \leq n} (F(z_i, z_j) \wedge F(z_j, z_i))$ is the boolean conjunctive query which is satisfied by the graphs containing a self-loop or a clique of size n (now considering F as the edge relation).

We first show that the sequence $(\mathcal{M}_n)_{n \geq 1}$ is uniformly Cauchy. Let $k \geq 1$. We claim that if we take $n_0 = k$, then for every source instance I , for every $n, m \geq n_0$, and every $q \in \mathbf{CQ}_k$, we have that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M}_m)$. To see this, note that for every source instance I and for every $t \geq 1$, the universal solutions of I w.r.t. \mathcal{M}_t have active domains consisting entirely of labeled nulls. Hence, only boolean queries may return a non-empty result. Moreover, observe that these universal solutions have no self-loops, i.e., they contain no atoms of the form $F(v, v)$ for some labeled null v .

We now distinguish two cases: First, suppose that $q \in \mathbf{CQ}_k$ is a boolean conjunctive query which contains a “self-loop”, i.e., an atom of the form $F(z, z)$ for some variable z . Then we clearly have $\text{cert}(q, I, \mathcal{M}_n) = \text{false} = \text{cert}(q, I, \mathcal{M}_m)$. It remains to consider the case that $q \in \mathbf{CQ}_k$ is a boolean CQ containing no self-loop. Then we clearly have $\text{cert}(q, I, \mathcal{M}_n) = \text{true} = \text{cert}(q, I, \mathcal{M}_m)$, since we are assuming that $m, n \geq k$ holds.

Using an argument similar to the one in the proof of Proposition 2, we now show that the sequence $(\mathcal{M}_n)_{n \geq 1}$ has no pointwise limit. Towards a contradiction, assume that $(\mathcal{M}_n)_{n \geq 1}$ does have a pointwise limit \mathcal{M} . Let I be a non-empty source instance. We consider three cases.

First, assume that $\text{Sol}(I, \mathcal{M})$ is empty. Then, for every boolean conjunctive query q , it holds trivially that $\text{cert}(q, I, \mathcal{M}) = \text{true}$. This is, in particular, the case for the query $q = \exists z F(z, z)$, which asks for the existence of a self-loop. However, for this query q , we have that $\text{cert}(q, I, \mathcal{M}_n) = \text{false}$ for every $n \geq 1$.

Second, assume that $\text{Sol}(I, \mathcal{M})$ is non-empty and that all solutions $J \in \text{Sol}(I, \mathcal{M})$ contain a self-loop. For the query $q = \exists z F(z, z)$ as above, we again have $\text{cert}(q, I, \mathcal{M}) = \text{true}$, whereas $\text{cert}(q, I, \mathcal{M}_n) = \text{false}$, for every $n \geq 1$.

Finally, assume that $\text{Sol}(I, \mathcal{M})$ is non-empty and that at least one solution $J \in \text{Sol}(I, \mathcal{M})$ does not contain a self-loop. Let m be the biggest integer such that J contains a clique of size m . Consider the conjunctive query

$$q = \exists z_1, \dots, z_{m+1} \bigwedge_{1 \leq i < j \leq m+1} (F(z_i, z_j) \wedge F(z_j, z_i)).$$

Then q evaluates to false over J and we have $\text{cert}(q, I, \mathcal{M}) = \text{false}$. On the other hand, for all $n \geq m + 1$ we have $\text{cert}(q, I, \mathcal{M}_n) = \text{true}$. Again, this contradicts our assumption that \mathcal{M} is the pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. □

The next difference is the definability of uniform limits. In Section 4, we saw that if a sequence of GAV mappings has a uniform limit, then it is eventually constant, hence it has a GAV mapping as a uniform limit. This property need not hold for sequences of LAV mappings (hence, it need not hold for sequences of premise-bounded schema mappings).

Proposition 7 *There exists a sequence $(\mathcal{M}_n)_{n \geq 1}$ of LAV mappings that has a uniform limit, but no uniform limit of $(\mathcal{M}_n)_{n \geq 1}$ admits universal solutions. In particular, no SO tgds is a uniform limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$.*

Proof For every $n \geq 1$, let \mathcal{M}_n be the LAV mapping specified by the constraint

$$\forall x (V(x) \rightarrow \exists P_n)$$

where $\exists P_n$ is the conjunctive query $\exists z_1 \dots \exists z_n (F(z_1, z_2) \wedge \dots \wedge F(z_{n-1}, z_n))$ asserting that there is a “path” (possibly with repeated vertices) of length n in the target instance. We now show that the sequence $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit, but no uniform limit of this sequence admits universal solutions.

Part 1. For the first part of the claim, consider the schema mapping

$$\mathcal{M} = \{(I, J) \mid I \neq \emptyset \text{ and } J \in v(C_k), k > 1\},$$

where C_k is a target instance consisting of a simple cycle of nulls of size k and $v(C_k)$ is the set of all isomorphic copies of C_k via isomorphisms that rename nulls. We will show that \mathcal{M} is a uniform limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$. Specifically, we will show for every m , there exists n_0 such that for all $n \geq n_0$, we have that $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}$.

Let $n_0 = m$. Since each \mathcal{M}_n has solutions consisting entirely of nulls, it suffices to consider boolean CQs only. Let q be a boolean CQ with m variables and assume that $\text{cert}(q, I, \mathcal{M}_n) = \text{true}$, where $n \geq m$. This implies that there is a homomorphism from the body of q into P_n , where P_n is the simple path with n nodes. In turn, this implies that $C_k \models q$, for every k . Thus, $\text{cert}(q, I, \mathcal{M}) = \text{true}$ as well. In the other direction, assume that $\text{cert}(q, I, \mathcal{M}) = \text{true}$. Note that q cannot contain a directed cycle, since no directed cycle can be mapped homomorphically in every cycle of length greater than one. Let h be a homomorphism from the body of q into C_{m+1} . Since $q \in \text{CQ}_m$, the variables of q have at most m distinct images among the nodes of C_{m+1} . This means that $\tilde{C}_{m+1} \models q$, where \tilde{C}_{m+1} is obtained from C_{m+1} by removing the facts that contain at least one element that is not the image of one of the variables of q under h . Note that \tilde{C}_{m+1} has at least one fact less than C_{m+1} , and so it is a collection of simple paths of length at most m ; therefore, there is a homomorphism from \tilde{C}_{m+1} to P_n , hence $P_n \models q$.

Part 2. For the second part of the claim and towards a contradiction, assume that \mathcal{M}' is a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$ such that there exists a non-empty source instance I and a finite universal solution J for I w.r.t. \mathcal{M}' . Note that for every i , we have that $\text{cert}(\exists P_i, I, \mathcal{M}') = \text{true}$, because \mathcal{M}' is a (uniform and, hence also pointwise) limit of the sequence $(\mathcal{M}_n)_{n \geq 1}$. Then we also have that $J \models \exists P_i$, since J is universal. Since J is finite, this is possible only if J contains a directed cycle.

We can now derive a contradiction as follows. For each positive integer l , let $\exists C_l$ be the boolean conjunctive query asserting the existence of a cycle of length l . Then there is no n such that $\text{cert}(\exists C_l, I, \mathcal{M}_n) = \text{true}$. Thus, $\text{cert}(\exists C_l, I, \mathcal{M}') = \text{false}$ must hold for every l , since \mathcal{M}' is a limit of $(\mathcal{M}_n)_{n \geq 1}$. Hence, J cannot contain cycles.

Since every SO tgds admits universal solutions, it follows that no SO tgds is a (uniform or pointwise) limit of $(\mathcal{M}_n)_{n \geq 1}$. □

By Theorem 1, every SO tgds is the uniform limit of a sequence of GLAV mappings. Proposition 7 implies that the converse is false, even for sequences of LAV mappings.

In the previous section, we showed that a sequence of GAV mappings has a GAV mapping as a pointwise limit if and only if it has a pointwise limit that allows for CQ-rewriting. Is there some structural property that characterizes when a sequence of premise-bounded GLAV mappings has a GLAV mapping as a pointwise limit (which, for premise-bounded mappings, is the same as a uniform limit)? We will show that the property of admitting universal solutions is the key to this question. Specifically, we have the following result.

Theorem 5 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a premise-bounded sequence of GLAV mappings. The following statements are equivalent.*

- (1) $(\mathcal{M}_n)_{n \geq 1}$ has a GLAV mapping \mathcal{M} as a uniform limit.
- (2) $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit that admits universal solutions.

Moreover, if $(\mathcal{M}_n)_{n \geq 1}$ is a sequence of LAV mappings, then $(\mathcal{M}_n)_{n \geq 1}$ has a LAV mapping as a uniform limit if and only if $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit that admits universal solutions.

We now give two lemmas which will be used in the proof of Theorem 5, but are also of interest in their own right.

Lemma 2 *If \mathcal{M} is the uniform limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings each of which allows for CQ-rewriting, then also \mathcal{M} allows for CQ-rewriting.*

Proof Let q be a target conjunctive query with m variables. Since \mathcal{M} is a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$, there exists an integer n_0 such that for every $n \geq n_0$ and every source instance I , we have that $\text{cert}(q, I, \mathcal{M}) = \text{cert}(q, I, \mathcal{M}_n)$. In particular, $\text{cert}(q, I, \mathcal{M}) = \text{cert}(q, I, \mathcal{M}_{n_0})$. Since \mathcal{M}_{n_0} allows for CQ-rewriting, there is a source conjunctive query q' such that $\text{cert}(q, I, \mathcal{M}_{n_0}) = q'(I)$, for every source instance I . Hence, $\text{cert}(q, I, \mathcal{M}) = q'(I)$ holds, for every source instance I . \square

It should be noted that the conclusion of Lemma 2 does not hold, in general, if \mathcal{M} is a pointwise limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings each of which allows for CQ-rewriting. Indeed, if $(\mathcal{M}_n)_{n \geq 1}$ is the sequence of GAV mappings in the proof of Proposition 5, then Theorem 3 and Proposition 5 imply that no pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$ allows for CQ-rewriting.

Lemma 3 *Let \mathcal{M} be a uniform limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$ of LAV mappings. If \mathcal{M} admits universal solutions, then it is closed under unions.*

Proof The proof proceeds through several stages and involves four claims, each of which builds on preceding ones. We first state the claims without proof and then use the last claim to show the desired conclusion. After this, we complete the proof of the lemma by proving each claim.

We first modify the notion of CQ-equivalence by limiting the number of atoms of CQs, rather than the number of variables. This yields an equivalent notion of uniform limit.

- For $\ell \geq 1$, we define $\text{CQ}'_\ell = \{q \in \text{CQ} \mid \text{length}(q) \leq \ell\}$, where $\text{length}(q)$ denotes the number of atoms in q .
- We say that two schema mappings \mathcal{M}_1 and \mathcal{M}_2 are CQ'_ℓ -equivalent, denoted by $\mathcal{M}_1 \equiv_{\text{CQ}'_\ell} \mathcal{M}_2$, if for every source instance I and for every $q \in \text{CQ}'_\ell$, we have that $\text{cert}(q, I, \mathcal{M}_1) = \text{cert}(q, I, \mathcal{M}_2)$.
- We say that \mathcal{M} is the u' -limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$, denoted by $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$, if for every ℓ , there exists n_0 such that for all $n \geq n_0$, it holds that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$.

Claim A. The notions of u' -limit and uniform limit coincide. Formally, for every sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings and every schema mapping \mathcal{M} , we have that $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$ if and only if $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$.

Next, we use the given sequence $(\mathcal{M}_n)_{n \geq 1}$ to construct another sequence $(\mathcal{M}'_n)_{n \geq 1}$ of LAV mappings that possesses some desirable properties. To define the sequence $(\mathcal{M}'_n)_{n \geq 1}$, we need another claim.

Claim B. Assume that $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$. Then, there exists a strictly increasing sequence $(n_i)_{i \geq 1}$ of positive integers, such that for every $\ell \geq 1$ and for every $n \geq n_\ell$, we have that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$.

Let $(n_i)_{i \geq 1}$ be the strictly increasing sequence of positive integers according to Claim B. We define the sequence $(\mathcal{M}'_n)_{n \geq 1}$ of LAV mappings as follows:

$$\mathcal{M}'_n = \begin{cases} \mathcal{M}_n & \text{if } n < n_1 \\ \bigcup_{\tau \in \mathcal{M}_n} T(\tau, \ell) & \text{otherwise, if } n_\ell \leq n < n_{\ell+1} \end{cases}$$

Here, $T(\tau, \ell)$ contains all LAV constraints obtained from τ by restricting the conclusion to at most ℓ atoms. Formally, let $\tau = A(\mathbf{x}) \rightarrow \exists \mathbf{y} A_1(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge A_r(\mathbf{x}, \mathbf{y})$ and let $\{j_1, \dots, j_p\} \subseteq \{1, \dots, r\}$ for $p \geq 1$. Define $\tau[j_1, \dots, j_p] := A(\mathbf{x}) \rightarrow \exists \mathbf{y} A_{j_1}(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge A_{j_p}(\mathbf{x}, \mathbf{y})$. We define

$$T(\tau, \ell) := \{\tau[j_1, \dots, j_p] \mid \{j_1, \dots, j_p\} \subseteq \{1, \dots, r\} \text{ and } p \leq \ell\}.$$

Claim C. Let $(n_i)_{i \geq 1}$ be the strictly increasing sequence of positive integers according to Claim B and let $(\mathcal{M}'_n)_{n \geq 1}$ be the sequence of LAV mappings constructed above. Then, for every $\ell \geq 1$, the following properties hold: (i) for every $n \geq n_\ell$, we have that $\mathcal{M}'_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$; (ii) the conclusion of every LAV constraint in \mathcal{M}'_{n_ℓ} is of length at most ℓ .

We now make the following claim about the sequence $(\mathcal{M}'_n)_{n \geq 1}$.

Claim D. For every source instance I , there exists an integer $n_0 \geq 1$ such that for every $I' \subseteq I$, we have that $\text{Sol}(I', \mathcal{M}'_{n_0}) = \text{Sol}(I', \mathcal{M})$.

Next, we use Claim D to show that \mathcal{M} is closed under unions, i.e., given $(I_1, J_1) \in \mathcal{M}$ and $(I_2, J_2) \in \mathcal{M}$, we must show that $(I, J) \in \mathcal{M}$ with $I = I_1 \cup I_2$ and $J = J_1 \cup J_2$. From Claim D, we know that there exists n_0 such that $\text{Sol}(I', \mathcal{M}'_{n_0}) = \text{Sol}(I', \mathcal{M})$, for every $I' \subseteq I$. In particular, $I_1, I_2 \subseteq I$. Hence, for each $i \in \{1, 2\}$, we have $J_i \in \text{Sol}(I_i, \mathcal{M}'_{n_0})$, that is, $(I_1, J_1) \in \mathcal{M}'_{n_0}$ and $(I_2, J_2) \in \mathcal{M}'_{n_0}$. Since \mathcal{M}'_{n_0} is a LAV mapping, it is closed under unions. Hence, $(I, J) \in \mathcal{M}'_{n_0}$, and, since $\text{Sol}(I, \mathcal{M}'_{n_0}) = \text{Sol}(I, \mathcal{M})$, we conclude that $J \in \text{Sol}(I, \mathcal{M})$, i.e., $(I, J) \in \mathcal{M}$.

To complete the proof of the lemma, it remains to prove Claims A-D.

Claim A. The notions of u' -limit and uniform limit coincide. Formally, for every sequence $(\mathcal{M}_n)_{n \geq 1}$ of schema mappings and every schema mapping \mathcal{M} , we have

$$\text{that } \mathcal{M}_n \xrightarrow{u} \mathcal{M} \text{ if and only if } \mathcal{M}_n \xrightarrow{u'} \mathcal{M}.$$

(\Rightarrow) Assume $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$. We have to show that also $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$ holds. Consider an arbitrary $\ell \geq 1$ and let r be the maximal arity of the target schema of \mathcal{M} . Any conjunctive query with at most ℓ atoms can have at most $m = \ell \cdot r$ variables. Hence, the inclusion $\text{CQ}'_\ell \subseteq \text{CQ}_m$ holds.

We are assuming $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$. Hence, there exists $n_0(m)$ such that for all $n \geq n_0(m)$, we have that $\mathcal{M}_n \equiv_{\text{CQ}_m} \mathcal{M}$. That is, for all $q \in \text{CQ}_m$ and for all I , it holds

that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. Since $\text{CQ}'_\ell \subseteq \text{CQ}_m$, we may conclude that for all $q \in \text{CQ}'_\ell$ and for all I , it holds that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. Hence, $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$ indeed holds.

(\Leftarrow) Assume $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$. We have to show that also $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$ holds. Consider an arbitrary $m \geq 1$. As above, let r be the maximal arity of the target schema of \mathcal{M} . Moreover, let p be the number of target relation symbols. Any conjunctive query with at most m variables can have at most $\ell = p \cdot m^r$ atoms. Hence, the inclusion $\text{CQ}_m \subseteq \text{CQ}'_\ell$ holds.

We are assuming $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$. Hence, there exists $n_0(\ell)$ such that for all $n \geq n_0(\ell)$, we have that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$. That is, for all $q \in \text{CQ}'_\ell$ and for all I , it holds that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. Since $\text{CQ}_m \subseteq \text{CQ}'_\ell$, we may conclude that for all $q \in \text{CQ}_m$ and for all I , it holds that $\text{cert}(q, I, \mathcal{M}_n) = \text{cert}(q, I, \mathcal{M})$. Hence, $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$ indeed holds.

Claim B. Assume that $\mathcal{M}_n \xrightarrow{u} \mathcal{M}$. Then, there exists a strictly increasing sequence $(n_i)_{i \geq 1}$ of positive integers, such that for every $\ell \geq 1$ and for every $n \geq n_\ell$, we have that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$.

Since $\mathcal{M}_n \xrightarrow{u'} \mathcal{M}$, for each $\ell \geq 1$ there exists an integer n'_ℓ such that for all $n \geq n'_\ell$, we have that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$. We may choose n_ℓ as follows to ensure strict monotonicity:

$$\begin{aligned} n_1 &:= n'_1 \\ &\dots \\ n_\ell &:= \max(n_{\ell-1} + 1, n'_\ell) \end{aligned}$$

Then the sequence $(n_i)_{i \geq 1}$ is strictly increasing and for all $\ell \geq 1$ and for all $n \geq n_\ell$, we have that $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$.

Claim C. Let $(n_i)_{i \geq 1}$ be the strictly increasing sequence of positive integers according to Claim B and let $(\mathcal{M}'_n)_{n \geq 1}$ be the sequence of LAV mappings constructed above. Then, for every $\ell \geq 1$, the following properties hold: (i) for every $n \geq n_\ell$, we have that $\mathcal{M}'_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$; (ii) the conclusion of every LAV constraint in \mathcal{M}'_{n_ℓ} is of length at most ℓ .

Consider an arbitrary $\ell \geq 1$. By the construction of the sequence $(\mathcal{M}'_n)_{n \geq 1}$, every LAV constraint in \mathcal{M}'_n has a conclusion of length at most ℓ . Hence, property (ii) clearly holds.

To prove property (i), consider an arbitrary $n \geq n_\ell$. We have to show that $\mathcal{M}'_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$, i.e., for arbitrary source instance I and arbitrary conjunctive query $q \in \text{CQ}'_\ell$, we have to show that $\text{cert}(q, I, \mathcal{M}'_n) = \text{cert}(q, I, \mathcal{M})$. By Claim B, we have $\mathcal{M}_n \equiv_{\text{CQ}'_\ell} \mathcal{M}$. Hence, it suffices to show that $\text{cert}(q, I, \mathcal{M}'_n) = \text{cert}(q, I, \mathcal{M}_n)$ holds. We prove the two inclusions separately.

By the construction of \mathcal{M}'_n , we clearly have $\text{chase}(I, \mathcal{M}'_n) \rightarrow \text{chase}(I, \mathcal{M}_n)$. From this, it follows immediately that $\text{cert}(q, I, \mathcal{M}'_n) \subseteq \text{cert}(q, I, \mathcal{M}_n)$.

For the reverse inclusion, consider an arbitrary tuple $\mathbf{a} \in \text{cert}(q, I, \mathcal{M}_n)$. Then, there exists a homomorphism $h_n: q \rightarrow \text{chase}(I, \mathcal{M}_n)$ with $h(\mathbf{z}) = \mathbf{a}$, where \mathbf{z}

denotes the free variables of q . Let $h_n(q) = \{A_1, \dots, A_k\} \subseteq \text{chase}(I, \mathcal{M}_n)$ with $k \leq \ell$. By construction, \mathcal{M}'_n is obtained by restricting the conclusions of the LAV constraints $\tau \in \mathcal{M}_n$ in all possible ways to at most ℓ atoms. Hence, since $k \leq \ell$, we have that also $\text{chase}(I, \mathcal{M}'_n)$ contains the set $\{A_1, \dots, A_k\}$ of atoms (up to renaming of labeled nulls). Thus, there exists a homomorphism $h: \{A_1, \dots, A_k\} \rightarrow \text{chase}(I, \mathcal{M}'_n)$ and $h(h_n(\cdot))$ is a homomorphism $q \rightarrow \text{chase}(I, \mathcal{M}'_n)$ with $h(\mathbf{z}) = \mathbf{a}$. Therefore, $\mathbf{a} \in \text{cert}(q, I, \mathcal{M}'_n)$ holds.

Before presenting the proof of Claim D, we need to bring the notion of *fact block size* into the picture; this notion was introduced in [7].

Fact Blocks. Let J be an instance. The *Gaifman graph of facts* G_J of J is the graph whose nodes are the facts of J and there is an edge between two facts if they have a null in common. The fact blocks (or f-blocks) of J are the sets of nodes of the connected components of G_J . The *block size* of an undirected graph G is the size of the maximal connected component of G , where the size of a component is given as the number of nodes. The *fact block size* (f-block size) of an instance J is the block size of the Gaifman graph of facts of J .

Claim D. For every source instance I , there exists an integer $n_0 \geq 1$ such that for every $I' \subseteq I$, we have that $\text{Sol}(I', \mathcal{M}'_{n_0}) = \text{Sol}(I', \mathcal{M})$.

Consider an arbitrary $I' \subseteq I$. Let J denote a universal solution for I' w.r.t. \mathcal{M} and let $J' = \text{chase}(I', \mathcal{M}'_{n_0})$. We set $\ell = \text{size}(J)$, where $\text{size}(J)$ denotes the number of atoms in J . Moreover, we set $n_0 = n_\ell$ from the construction of $(\mathcal{M}'_n)_{n \geq 1}$. We claim that n_0 has the desired property. The proof proceeds in three steps, namely, we will show (i) $J \rightarrow J'$, (ii) $J' \rightarrow J$, and, finally, (iii) $\text{Sol}(I', \mathcal{M}'_{n_0}) = \text{Sol}(I', \mathcal{M})$.

- (i) Let $\mathbf{u} = (u_1, \dots, u_i)$ in J denote the labeled nulls in J and let $\mathbf{y} = (y_1, \dots, y_i)$ denote a vector of pairwise distinct variables. Consider the boolean conjunctive query $\exists \mathbf{y} q_J$ whose atoms are the atoms in J where we instantiate the labeled nulls $\mathbf{u} = (u_1, \dots, u_i)$ with $\mathbf{y} = (y_1, \dots, y_i)$. Clearly $q_J \rightarrow J$ holds and, therefore, also $\text{cert}(\exists \mathbf{y} q_J, I', \mathcal{M}) = \text{true}$.

Since $\mathcal{M}'_{n_0} \equiv_{\text{CQ}'_\ell} \mathcal{M}$ and $\exists \mathbf{y} q_J \in \text{CQ}'_\ell$, also $\text{cert}(\exists \mathbf{y} q_J, I', \mathcal{M}'_{n_0}) = \text{true}$ holds. Hence, there exists a homomorphism $h': q_J \rightarrow J'$, which can be easily transformed into a homomorphism $h: J \rightarrow J'$ by setting $h(u_\alpha) = h'(y_\alpha)$ for every $\alpha \in \{1, \dots, i\}$.

- (ii) For every f-block F' of J' , we consider the boolean conjunctive query $\exists \mathbf{z} q_{F'}$ whose atoms are the atoms in F' and $\mathbf{z} = (z_1, \dots, z_i)$ instantiates the labeled nulls $\mathbf{v} = (v_1, \dots, v_i)$ in F' with pairwise distinct variables. Clearly, for every F' , we have $q_{F'} \rightarrow J'$ and, therefore, also $\text{cert}(\exists \mathbf{z} q_{F'}, I', \mathcal{M}) = \text{true}$.

Since all LAV-constraints in \mathcal{M}'_{n_0} have conclusion size bounded by ℓ , the number of atoms in any f-block of J' is bounded by ℓ . Hence, for every F' , the corresponding conjunctive query $q_{F'}$ is in CQ'_ℓ . Since $\mathcal{M}'_{n_0} \equiv_{\text{CQ}'_\ell} \mathcal{M}$, we have that $\text{cert}(\exists \mathbf{z} q_{F'}, I', \mathcal{M}) = \text{true}$. Hence, for every f-block F' of J' , there exists a homomorphism $h_{F'}: q_{F'} \rightarrow J$, which can easily be transformed into a homomorphism

$g_{F'}: F' \rightarrow J'$ by setting $g_{F'}(v_\alpha) = h_{F'}(z_\alpha)$ for every $\alpha \in \{1, \dots, i\}$. These homomorphisms from the f-blocks of J' to J can be combined to the desired homomorphism $h' = \bigcup g_{F'}$ with $h': J' \rightarrow J$.

(iii) Finally, we show that $\text{Sol}(I', \mathcal{M}'_{n_0}) = \text{Sol}(I', \mathcal{M})$ holds.

“ \subseteq ”: Let $K \in \text{Sol}(I', \mathcal{M}'_{n_0})$. Since J' is a universal solution for I' w.r.t. \mathcal{M}'_{n_0} , there exists a homomorphism $g' : J' \rightarrow K$. By composing g' with the homomorphism $h : J \rightarrow J'$, we obtain a homomorphism from J to K . By the closure under target homomorphisms, we conclude that $K \in \text{Sol}(I', \mathcal{M})$

“ \supseteq ”: Now let $K \in \text{Sol}(I', \mathcal{M})$. Since J is a universal solution for I' w.r.t. \mathcal{M} , there exists a homomorphism $g : J \rightarrow K$. By composing g with the homomorphism $h' : J' \rightarrow J$, we obtain a homomorphism from J' to K . Since LAV mapping \mathcal{M}'_{n_0} is closed under target homomorphisms, we conclude that $K \in \text{Sol}(I', \mathcal{M}'_{n_0})$.

The proof of Lemma 3 is now complete. □

We now have all the tools needed to present the proof of Theorem 5. Before doing so and for the sake of readability, we reproduce its statement.

Let $(\mathcal{M}_n)_{n \geq 1}$ be a premise-bounded sequence of GLAV mappings. The following statements are equivalent.

- (1) $(\mathcal{M}_n)_{n \geq 1}$ has a GLAV mapping \mathcal{M} as a uniform limit.
- (2) $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit that admits universal solutions.

Moreover, if $(\mathcal{M}_n)_{n \geq 1}$ is a sequence of LAV mappings, then $(\mathcal{M}_n)_{n \geq 1}$ has a LAV mapping as a uniform limit if and only if $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit that admits universal solutions.

Proof of Theorem 5 The direction (1) \Rightarrow (2) is obvious. For the direction (2) \Rightarrow (1), we start with the case when $(\mathcal{M}_n)_{n \geq 1}$ is a sequence of LAV mappings.

Assume that \mathcal{M} is a uniform limit of a sequence $(\mathcal{M}_n)_{n \geq 1}$ of LAV mappings and that \mathcal{M} admits universal solutions. Without loss of generality, we may also assume that \mathcal{M} is closed under target homomorphism. Indeed, if we let \mathcal{M}' be the schema mapping obtained by closing \mathcal{M} under target homomorphisms, then \mathcal{M}' is also a uniform limit of $(\mathcal{M}_n)_{n \geq 1}$ and it admits universal solutions; this is so because the notion of uniform limit is based on CQ-equivalence and also conjunctive queries are preserved under homomorphisms. Then the schema mapping \mathcal{M} has the following properties:

1. \mathcal{M} allows for CQ-rewriting (by Lemma 2);
2. \mathcal{M} admits universal solutions (by hypothesis);
3. \mathcal{M} is closed under target homomorphisms (by hypothesis);
4. \mathcal{M} is closed under unions (by Lemma 3).

Theorem 3.1 in [19] asserts that if a schema mapping admits universal solutions, allows for query rewriting, and is closed under both target homomorphisms and unions, then it is logically equivalent to a LAV mapping. Consequently, we have that \mathcal{M} is logically equivalent to a LAV mapping.

For the case when $(\mathcal{M}_n)_{n \geq 1}$ is a sequence of premise-bounded GLAV mappings (but not necessarily LAV mappings), we apply yet another structural characterization of GLAV mappings from [19], namely, Theorem 3.9, which asserts that if a schema mapping allows for CQ-rewriting, admits universal solutions, is closed under target homomorphisms, and is n -modular, for some fixed n , then it is logically equivalent to a GLAV mapping.

Let k be the constant bounding the length of premises in $(\mathcal{M}_n)_{n \geq 1}$. We proceed exactly as in the proof of Lemma 3 and construct a sequence $(\mathcal{M}'_n)_{n \geq 1}$, in which the premises of tgds are the same as in tgds in $(\mathcal{M}_n)_{n \geq 1}$, hence each tgd in $(\mathcal{M}'_n)_{n \geq 1}$ has at most k atoms in its premise. We proceed exactly as in the proof of Lemma 3 to establish the following analog of Claim D.

Claim D (in the proof of Lemma 3) For every source instance I , there exists an integer $n_0 \geq 1$ such that for every $I' \subseteq I$, we have that $Sol(I', \mathcal{M}'_{n_0}) = Sol(I', \mathcal{M})$.

Now, since each tgd in every element of $(\mathcal{M}'_n)_{n \geq 1}$ has at most k atoms in its premise, it follows that there is a positive integer N_k so that each mapping \mathcal{M}'_n in $(\mathcal{M}'_n)_{n \geq 1}$ is N_k -modular. It is easy to see that $N_k \leq k \cdot r$ holds where r is the maximum relation arity in the source schema.

We now prove that \mathcal{M} is N_k -modular. Assume that J is not a solution for I w.r.t. to \mathcal{M} . Take an integer n_0 as in Claim D and consider \mathcal{M}'_{n_0} . It follows that J is not a solution for I w.r.t. \mathcal{M}'_{n_0} . Since \mathcal{M}'_{n_0} is N_k -modular, there is a subinstance I' of I such that J is not a solution for I' w.r.t. \mathcal{M}'_{n_0} and $|dom(I')| \leq N_k$. Again by Claim D, we have that J is not a solution for I' w.r.t. \mathcal{M} , hence \mathcal{M} is N_k -modular.

Thus, \mathcal{M} has the following properties: it admits CQ-rewriting (since it is the uniform limit of GLAV mappings that admit CQ-rewriting), it admits universal solutions, is closed under target homomorphisms (if it is not, we take its closure before we begin the construction), and, as just shown, it is N_k -modular. Consequently, by Theorem 3.9 in [19], we have that \mathcal{M} is logically equivalent to a GLAV schema mapping, which completes the proof. \square

We conclude this section with a conjecture concerning uniform limits of arbitrary sequences of GLAV mappings.

Conjecture 1 *The following statements are equivalent for a sequence $(\mathcal{M}_n)_{n \geq 1}$ of GLAV mappings.*

1. $(\mathcal{M}_n)_{n \geq 1}$ has an SO tgd as a uniform limit.
2. $(\mathcal{M}_n)_{n \geq 1}$ has a uniform limit that admits universal solutions.

It is not hard to show that the preceding conjecture is implied by a conjecture in [2] to the effect that the language of *plain* SO-tgds² can be characterized by the following three properties: allowing for CQ-rewriting, admitting universal solutions, and closure under target homomorphisms.

²A *plain* SO tgd is an SO tgd that contains no nested terms and no equalities. Every SO tgd is known to be CQ-equivalent to a plain one [2].

6 Metric Space Completion and Generalized Schema Mappings

Let \mathbf{T} be a schema containing a binary relation symbol. By Proposition 2, the metric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$ is not complete, i.e., there are Cauchy sequences of elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ that have no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Let $(\mathcal{P}(\text{Inst}(\mathbf{T}))^*, \text{dist}^*)$ be the completion of $(\mathcal{P}(\text{Inst}(\mathbf{T})), \text{dist})$. As described in Section 2, the elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))^*$ are the equivalence classes of Cauchy sequences of elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))$, where two Cauchy sequences $\mathcal{I}_1, \mathcal{I}_2, \dots$ and $\mathcal{J}_1, \mathcal{J}_2, \dots$ are equivalent if $\lim_{n \rightarrow \infty} \text{dist}(\mathcal{I}_n, \mathcal{J}_n) = 0$. Clearly, this is a rather abstract description of $\mathcal{P}(\text{Inst}(\mathbf{T}))^*$. In this section we show that, in many cases, the elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))^*$ can be represented by suitably constructed infinite \mathbf{T} -instances. In turn, this result and basic results about complete metric spaces imply that the (pointwise or uniform) limits of a Cauchy sequence of schema mappings can be represented by a *generalized* schema mapping, that is, a schema mapping in which infinite solutions are allowed. We also establish a tight connection between these results and the representation of structural limits in the monograph by Nešetřil and Ossona de Mendez [15].

6.1 Representing Limits of Cauchy Sequences in the Metric Completion

Let \mathbf{T} be a schema. Recall that, by definition, a \mathbf{T} -instance is a finite set of facts. In what follows, we will also consider infinite \mathbf{T} -instances, where, by definition, an *infinite* \mathbf{T} -instance is an infinite set I of facts $R_i(t_1, \dots, t_m)$. The term \mathbf{T} -instance will continue to denote a finite \mathbf{T} -instance, but, at times and for emphasis or disambiguation, we will also use the term finite \mathbf{T} -instance, especially in contexts in which infinite \mathbf{T} -instances are also considered. According to Definitions 2 and 3, the notion of the distance between two sets of finite instances has been defined using the notion of CQ_n -equivalence, where two sets \mathcal{J} and \mathcal{J}' of finite \mathbf{T} -instances are CQ_n -equivalent, denoted $\mathcal{J} \equiv_{\text{CQ}_n} \mathcal{J}'$, if it holds that $\text{cert}(q, \mathcal{J}) = \text{cert}(q, \mathcal{J}')$, for all $q \in \text{CQ}_n$. The notion of CQ_n -equivalence naturally extends to arbitrary (i.e., finite or infinite) \mathbf{T} -instances. Hence, also the notions of similarity and distance, both of which were defined via CQ_n -equivalence, immediately carry over to sets of arbitrary \mathbf{T} -instances. Furthermore, the set of sets of arbitrary \mathbf{T} -instances forms a pseudometric space, in which we can speak about Cauchy sequences and limits.

Definition 8 Let \mathbf{T} be a schema.

- Let \mathcal{X} and \mathcal{Y} be two sets of finite \mathbf{T} -instances. We say that \mathcal{Y} is an *isomorphic copy of \mathcal{X} with nulls named apart* if
 1. For every member J of \mathcal{X} , there is a member J' of \mathcal{Y} that is an isomorphic copy of J via an isomorphism that renames nulls.
 2. Every member J' of \mathcal{Y} is an isomorphic copy of some member J of \mathcal{X} via an isomorphism that renames nulls.
 3. No two members of \mathcal{Y} have nulls in common.
- If \mathcal{Y} is a set of finite \mathbf{T} instances, then $\bigcup \mathcal{Y}$ denotes the union of all members of \mathcal{Y} (where each member of \mathcal{Y} is viewed as a set of facts).

- If \mathcal{X} is a set of finite **T**-instances, then $\bigoplus \mathcal{X}$ denotes the set consisting of the unions of isomorphic copies of \mathcal{X} with nulls named apart, i.e.,

$$\bigoplus \mathcal{X} = \left\{ \bigcup \mathcal{Y} \mid \mathcal{Y} \text{ is an isomorphic copy of } \mathcal{X} \text{ with nulls named apart} \right\}.$$

Several remarks are in order now.

- Let \mathcal{Y} be a set of finite **T**-instances. Clearly, if \mathcal{Y} is finite, then $\bigcup \mathcal{Y}$ is a finite **T**-instance, while if \mathcal{Y} is infinite, then $\bigcup \mathcal{Y}$ is an infinite **T**-instance. Note also that if \mathcal{X} is a set of finite **T**-instances such that at least one instance in \mathcal{X} contains nulls, then $\bigoplus \mathcal{X}$ is infinite (even if \mathcal{X} is a finite set).
- According to Definition 4, if J is a **T**-instance whose active domain contains nulls only, then $v(J)$ is the set of all **T**-instances that are isomorphic copies of J via an isomorphism that renames nulls. This notation makes sense also for infinite **T**-instances J whose active domains contain nulls only. With this in mind, observe that if \mathcal{X} is a finite set of **T**-instances and if \mathcal{Y} is an isomorphic copy of \mathcal{X} with nulls named apart, then

$$\bigoplus \mathcal{X} = v\left(\bigcup \mathcal{Y}\right).$$

- As a concrete example, if $\mathcal{K} = \{K_n \mid n \geq 1\}$, where K_n is a clique of size n in which every node is a null, then the members of $\bigoplus \mathcal{K}$ are precisely the disjoint unions of cliques of all finite sizes in which every node is a null.

Definition 9 Let q be a conjunctive query over the schema **T** with k free variables, $k \geq 0$, and let \mathbf{a} be a k -tuple of constants (if $k = 0$, then $\mathbf{a} = ()$, i.e., \mathbf{a} is the empty tuple).

We write $q(\mathbf{a})$ to denote the **T**-instance J obtained from q and \mathbf{a} by (i) substituting the free variables of q by the respective elements of \mathbf{a} ; (ii) replacing the existential variables of q by fresh distinct labeled nulls; and (iii) treating the resulting body atoms of q as facts of the **T**-instance J .

Note that if q is a boolean query (in which case $\mathbf{a} = ()$), then $q(())$ is the *canonical database* of q , i.e., the **T**-instance whose active domain is the set of variables of q viewed as distinct nulls and whose facts are the atoms of q . Conversely, every **T**-instance J whose active domain consists entirely of nulls is the canonical database of a boolean conjunctive query.

Before stating the main result of this section, we need to introduce one more concept. Let \mathcal{J} be a set of finite or infinite **T**-instances. We say that \mathcal{J} is *closed under isomorphisms that rename nulls* if for every (finite or infinite) **T**-instance J in \mathcal{J} and for every (finite or infinite) **T**-instance J' that is an isomorphic copy of J via an isomorphism that renames nulls, we have that J' is also in \mathcal{J} . Note that if \mathcal{X} is a set of finite **T**-instances, then $\bigoplus \mathcal{X}$ is closed under isomorphisms that rename nulls. Moreover, if \mathcal{M} is a schema mapping between **S** and **T**, then, for every source instance I , the set $\text{Sol}(I, \mathcal{M})$ of the solutions of I w.r.t. \mathcal{M} is closed under isomorphisms that rename nulls (see Definition 1).

Theorem 6 *Let $(\mathcal{J}_n)_{n \geq 1}$ be a Cauchy sequence of elements of $\mathcal{P}(\text{Inst}(\mathbf{T}))$ such that each \mathcal{J}_n is closed under isomorphisms that rename nulls. Then the limit of the sequence $(\mathcal{J}_n)_{n \geq 1}$ is the set $\bigoplus \mathcal{J}^*$, where*

$$\mathcal{J}^* = \{q(\mathbf{a}) \mid q \in \text{CQ} \text{ and there is an integer } p \text{ such that } \mathbf{a} \in \text{cert}(q, \mathcal{J}_i), \text{ for every } i \geq p\}.$$

Proof We have to show that, for every $m \geq 1$, there is some n_0 such that for every $q \in \text{CQ}_m$ and every $n \geq n_0$, we have that $\text{cert}(q, \mathcal{J}_n) = \text{cert}(q, \bigoplus \mathcal{J}^*)$. This will be done in two steps, as follows.

Step 1: We will show that, for every $m \geq 1$, there is some n_1 such that $\text{cert}(q, \mathcal{J}_n) \subseteq \text{cert}(q, \bigoplus \mathcal{J}^*)$, for every $q \in \text{CQ}_m$ and every $n \geq n_1$.

Step 2: We will show that, for every $m \geq 1$, there is some n_2 such that $\text{cert}(q, \bigoplus \mathcal{J}^*) \subseteq \text{cert}(q, \mathcal{J}_n)$, for every $q \in \text{CQ}_m$ and every $n \geq n_2$.

Then, given $m \geq 1$, we can take $n_0 = \max\{n_1, n_2\}$.

We start by pointing out that for every $n \geq 1$ and every $q \in \text{CQ}$, the certain answers $\text{cert}(q, \mathcal{J}_n)$ consist entirely of null-free tuples. This follows from the assumption that \mathcal{J}_n is closed under isomorphisms that rename nulls (the proof is essentially the same as the proof of Proposition 1 in Section 2). Moreover, for every $q \in \text{CQ}$, the certain answers $\text{cert}(q, \bigoplus \mathcal{J}^*)$ also consist entirely of null-free tuples. This is so because $\bigoplus \mathcal{J}^*$ contains isomorphic copies of \mathcal{J}^* having no nulls in common (e.g., if v_1, \dots, v_n, \dots is a list of all nulls, then $\bigoplus \mathcal{J}^*$ contains an isomorphic copy of \mathcal{J}^* in which all nulls have even index and an isomorphic copy of \mathcal{J}^* in which all nulls have odd index). Thus, we only need to focus on tuples of constants as possible certain answers.

To prove Step 1, since the sequence $(\mathcal{J}_n)_{n \geq 1}$ is Cauchy, for every $m \geq 1$, there is some n_1 such that if $s \geq n_1$ and $t \geq n_1$, then $\mathcal{J}_s \equiv_{\text{CQ}_m} \mathcal{J}_t$. We now claim that $\text{cert}(q, \mathcal{J}_n) \subseteq \text{cert}(q, \bigoplus \mathcal{J}^*)$, for every $q \in \text{CQ}_m$ and every $n \geq n_1$. Indeed, assume that $q \in \text{CQ}_m$ and let \mathbf{a} be a (possibly empty) tuple of constants in $\text{cert}(q, \mathcal{J}_n)$, where $n \geq n_1$. It follows that $\mathbf{a} \in \text{cert}(q, \mathcal{J}_j)$, for every $j \geq n_1$, hence the finite \mathbf{T} -instance $q(\mathbf{a})$ is in the set \mathcal{J}^* . Consequently, $\mathbf{a} \in q(\bigcup \mathcal{V})$, for every isomorphic copy \mathcal{V} of \mathcal{J}^* with nulls named apart, which implies that $\mathbf{a} \in \text{cert}(q, \bigoplus \mathcal{J}^*)$.

To prove Step 2, we will first show that the set D of constants occurring in \mathcal{J}^* is finite (note that D is also the set of constants occurring in $\bigoplus \mathcal{J}^*$). As a stepping stone, we will show the finiteness of a set D' that is defined next.

A *single-atom* conjunctive query is a query of the form $\exists \mathbf{y}R(\mathbf{x}, \mathbf{y})$, where R is a relation symbol in the schema \mathbf{T} . Let D' be the set of all constants b for which there is a single-atom query q and an index p , such that b occurs in $\text{cert}(q, \mathcal{J}_i)$, for all $i \geq p$. We claim that the set D' is finite. To see this, observe first that every single-atom query has at most r variables, where r is the maximum arity of the relation symbols in \mathbf{T} . Since the sequence $(\mathcal{J}_n)_{n \geq 1}$ is Cauchy, there exists an integer p_r such that $\mathcal{J}_i \equiv_{\text{CQ}_r} \mathcal{J}_{p_r}$, for all $i \geq p_r$. This implies that the certain answers to single-atom conjunctive queries become fixed in $(\mathcal{J}_n)_{n \geq 1}$ starting from the index p_r , which depends only on

the schema **T**. By definition, the certain answers hold in every instance in \mathcal{J}_{p_r} . Since \mathcal{J}_{p_r} consists entirely of finite instances, the set D' must be finite as well.

To complete the proof of the finiteness of D , we will show that $D \subseteq D'$. Let \mathbf{a} be a tuple of constants for which there is a conjunctive query q and an index p , such that $\mathbf{a} \in cert(q, \mathcal{J}_i)$, for all $i \geq p$. Let s be the number of atoms of q and consider the single-atom queries $q'_1(\mathbf{y}_1), \dots, q'_s(\mathbf{y}_s)$ that cover q in the following sense: for every j with $1 \leq j \leq s$, the atom of q'_j is the j -th atom of q , and \mathbf{y}_j contains exactly the free variables of q that occur in this atom. Let \mathbf{a}_j be the tuple of elements from \mathbf{a} assigned to the variables \mathbf{y}_j . Clearly, every element of \mathbf{a} is an element of some \mathbf{a}_j , $1 \leq j \leq s$. Observe that $\mathbf{a}_j \in cert(q, \mathcal{J}_i)$ implies that $\mathbf{a}_j \in cert(q'_j, \mathcal{J}_i)$, hence we have that $\mathbf{a}_j \in cert(q'_j, \mathcal{J}_i)$, for every $i \geq p$. Thus, each element of \mathbf{a}_j , $1 \leq j \leq s$, is an element of D' . This shows that $D \subseteq D'$ holds, hence D is a finite set.

We now return to the proof of Step 2. We will show that for every $m \geq 1$, there is some n_2 such that $cert(q, \bigoplus \mathcal{J}^*) \subseteq cert(q, \mathcal{J}_n)$, for every $q \in \mathbf{CQ}_m$ and every $n \geq n_2$. Assume that $q \in \mathbf{CQ}_m$ and let \mathbf{a} be a tuple of constants such that $\mathbf{a} \in cert(q, \bigoplus \mathcal{J}^*)$. Then, for every instance $J \in \bigoplus \mathcal{J}^*$, we have that $\mathbf{a} \in q(J)$, hence there is a homomorphism h from the variables of q to the active domain of J such that the tuple of the free variables of q is mapped to \mathbf{a} and the atoms of q are mapped to facts of J . Let s be the number of atoms of q and let f_1, \dots, f_s be the facts of J that are the images of the atoms of q under the homomorphism h . Up to renaming nulls, each fact f_j is a fact of some finite **T**-instance of the form $q_j(\mathbf{b}_j)$, where q_j is a conjunctive query and \mathbf{b}_j is a tuple of constants such that $\mathbf{b}_j \in cert(q_j, \mathcal{J}_i)$, for all sufficiently large i . Let $n_{q(\mathbf{a})}$ be an index such that for every $i \geq n_{q(\mathbf{a})}$, we have that $\mathbf{b}_j \in cert(q_j, \mathcal{J}_i)$ holds, for $1 \leq j \leq s$. Furthermore, let n_2 be the maximum such index $n_{q(\mathbf{a})}$, for all q in \mathbf{CQ}_m and for all tuples \mathbf{a} in D . Such an index exists (i.e., it is a finite number) because both the set \mathbf{CQ}_m and the set of tuples of elements D of length at most m is finite.

Observe that n_2 has been chosen so that for every tuple \mathbf{a} and for every $q \in \mathbf{CQ}_m$ with a homomorphism h mapping $q(\mathbf{a})$ to some instance in $\bigoplus \mathcal{J}^*$ (and thus to every instance in $\bigoplus \mathcal{J}^*$, by renaming the nulls in the co-domain of h accordingly), every fact f_j in $h(q(\mathbf{a}))$ can be mapped further to every instance \mathcal{J}_n , $n \geq n_2$, via a homomorphism h_i defined on the entire f -block of f_j . (Recall that, by the definition of $\bigoplus \mathcal{J}^*$, each fact f_j instantiates an atom of some conjunctive query q_j whose certain answers persist in the sequence $(\mathcal{J}_n)_{n \geq 1}$; the bodies of these queries are mapped into instances of $\bigoplus \mathcal{J}^*$ after renaming apart the nulls in them, thus ensuring that no two distinct queries end up in the same f -block of an instance of $\bigoplus \mathcal{J}^*$).

The union of two homomorphisms h_1, h_2 defined on two distinct f -blocks B_1, B_2 is unambiguously defined, and it is a homomorphism on the instance $B_1 \cup B_2$, since homomorphisms are the identity on constants and f -blocks do not share nulls. Thus, for an instance $J \in \bigoplus \mathcal{J}^*$ and for the image $\{f_1, \dots, f_s\}$ of $q(\mathbf{a})$ under some homomorphism h , we also have a homomorphism from $q(\mathbf{a})$ to \mathcal{J}_n , $n \geq n_2$, obtained by composing h with a union $h_1 \cup \dots \cup h_s$ of homomorphisms from the f -blocks of the atoms f_1, \dots, f_s to \mathcal{J}_n . It follows that $\mathbf{a} \in cert(q, \mathcal{J}_n)$, for every $n \geq n_2$. This establishes the inclusion $cert(q, \mathcal{J}^*) \subseteq cert(q, \mathcal{J}_n)$, for $n \geq n_2$, and completes the proof of the theorem. □

Recall the sequence $(v(K_n))_{n \geq 1}$ in Proposition 2, where K_n is the clique of size n whose vertices are pairwise distinct labeled nulls. By Proposition 2, this sequence is Cauchy, but has no limit in $\mathcal{P}(\text{Inst}(\mathbf{T}))$. Theorem 6 tells us how to find the limit in the complete metric space via the conjunctive queries with non-empty certain answers over all but finitely many members of the sequence. Since the instances K_n , $n \geq 1$, have active domains consisting entirely of nulls, Lemma 1 tells us that we only need to consider boolean conjunctive queries and, moreover, it suffices to evaluate them on each K_n . These queries can only use the edge relation E , thus they can be considered as graphs - with the variables representing the vertices. If a query contains a self-loop (i.e., an atom of the form $E(z, z)$ for some variable z), then the query evaluates to false over every K_n . On the other hand, if a query contains no self-loop, then it evaluates to true over all but finitely many instances K_n . Indeed, let q be a conjunctive query without self-loop and suppose that q contains m variables. It is easy to verify that q evaluates to true over all instances K_n with $n \geq m$. Hence, by Theorem 6, the limit of $(v(K_n))_{n \geq 1}$ is $\bigoplus \mathcal{G}$, where \mathcal{G} is a set of graphs with the following properties: (i) every member of \mathcal{G} is a graph with no self-loops and with labelled nulls as vertices; (ii) every graph with no self-loops is isomorphic to a graph in \mathcal{G} . Clearly, $\bigoplus \mathcal{K}$ is also the limit of $(v(K_n))_{n \geq 1}$, where \mathcal{K} is a set of graphs with the following properties: (i) every member of \mathcal{K} is a clique with labelled nulls as vertices; (ii) every clique is isomorphic to a graph in \mathcal{K} . Thus, the limit of $(v(K_n))_{n \geq 1}$ is the set consisting of all disjoint unions of cliques of all finite sizes in which every node is a null. At any rate, it is clear that infinite instances have to be used to represent the limit of $(v(K_n))_{n \geq 1}$.

Next, we extend our results about limits of Cauchy *sequences of instances* to limits of Cauchy *sequences of mappings*. To this end, we first recall two basic results about complete metric spaces.

Proposition 8 *Let (Y, d) be a complete metric space and let $(f_n)_{n \geq 1}$ be a sequence of functions from a set X to Y .*

- *If $(f_n)_{n \geq 1}$ is a pointwise Cauchy sequence, then $(f_n)_{n \geq 1}$ has a pointwise limit $f : X \rightarrow Y$, where $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, for every $x \in X$.*
- *If $(f_n)_{n \geq 1}$ is a uniformly Cauchy sequence, then $(f_n)_{n \geq 1}$ has a uniform limit. Moreover, the pointwise limit $f : X \rightarrow Y$ of $(f_n)_{n \geq 1}$ is also the uniform limit of $(f_n)_{n \geq 1}$.*

The proof of the first part of Proposition 8 is immediate from the definitions; the proof of the second part can be found in any standard book on metric spaces (see, e.g., Proposition 3.6.6 in [18]). In fact, the argument is essentially the same as the one given in the proof of Part 2 of Theorem 4. Note that the second part of Proposition 8 is known as the *Cauchy criterion*.

We are now ready to obtain concrete representations of the (pointwise or uniform) limits of Cauchy sequences of schema mappings.

Definition 10 *Let \mathbf{S}, \mathbf{T} be two schemas. A generalized schema mapping is a set \mathcal{M} of pairs (I, J) such that I is a finite \mathbf{S} -instance, J is a finite or infinite \mathbf{T} -instance,*

and \mathcal{M} has the following closure property: if $(I, J) \in \mathcal{M}$ and if J' is an isomorphic copy of J via an isomorphism that renames nulls, then $(I, J') \in \mathcal{M}$.

Corollary 3 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a sequence of schema mappings. Consider the generalized schema mapping*

$$\mathcal{M} = \left\{ (I, J) \mid J \in \bigoplus \{q(\mathbf{a}) \mid q \in \mathbf{CQ} \wedge \exists p \forall i \geq p (\mathbf{a} \in \text{cert}(q, I, \mathcal{M}_i))\} \right\}.$$

- *If $(\mathcal{M}_n)_{n \geq 1}$ is a pointwise Cauchy sequence, then the schema mapping \mathcal{M} is the pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$.*
- *If $(\mathcal{M}_n)_{n \geq 1}$ is a uniformly Cauchy sequence, then the schema mapping \mathcal{M} is the uniform limit of $(\mathcal{M}_n)_{n \geq 1}$.*

Proof The first part follows from Theorem 6 and the definitions. The second part follows from the first part and Proposition 8. □

Finally, we consider (pointwise or uniformly) Cauchy sequences of schema mappings admitting universal solutions and obtain a different representation of their limits.

Corollary 4 *Let $(\mathcal{M}_n)_{n \geq 1}$ be a pointwise Cauchy sequence of schema mappings over a source schema \mathbf{S} and a target schema \mathbf{T} , each admitting universal solutions.*

1. *For every $I \in \text{Inst}(\mathbf{S})$, the sequence $(\text{UnivSol}(I, \mathcal{M}_n))_{n \geq 1}$ is Cauchy, and hence it has a limit $\lim_{n \rightarrow \infty} (\text{UnivSol}(I, \mathcal{M}_n))$ in the complete metric space $(\mathcal{P}(\text{Inst}(\mathbf{T}))^*, \text{dist}^*)$.*
2. *The generalized schema mapping*

$$\mathcal{M}^* = \{(I, J) \mid I \in \text{Inst}(\mathbf{S}), J \in \lim_{n \rightarrow \infty} (\text{UnivSol}(I, \mathcal{M}_n))\}$$

is a pointwise limit of $(\mathcal{M}_n)_{n \geq 1}$. Moreover, if $(\mathcal{M}_n)_{n \geq 1}$ is a uniformly Cauchy sequence, then \mathcal{M}^ is its uniform limit.*

6.2 Connections with Representations of Structural Limits

In their recent monograph [15], Nešetřil and Ossona de Mendez considered a notion of distance between instances, as well as sequences of instances and limits of such sequences. In what follows, we describe the main differences between their setting and ours.

- The first main difference is that they did not distinguish two classes of domain elements (namely, constants and nulls), as we did here. As a result, in the definition of homomorphism in [15], no special treatment of constants is needed, while, in our setting, constants must always be mapped to themselves. Their notion of homomorphism coincides with ours on instances whose active domains consist of labeled nulls only. Note that this is exactly the scenario we had in Example 1 and Proposition 2, which are both inspired by results in [15].

- The second main difference is that the notion of distance in [15] is between a pair of two instances, while our notion of distance is between a pair of two *sets* of instances. This, of course, raises the question of how the two notions compare if, in our setting, both sets are singletons. We will address this question soon.
- The third main difference is that, when cast in terms of the certain answers of conjunctive queries, the notion of distance in [15] involves boolean conjunctive queries only, while ours involves all conjunctive queries (boolean and non-boolean ones).

In what follows, we recall the definition of the similarity measure and the metric from [15] and briefly sketch the approach that Nešetřil and Ossona de Mendez took in representing limits of Cauchy sequences of instances via infinite instances.

Let \mathbf{T} be a schema and let J and J' be two \mathbf{T} -instances. By a slight abuse of notation, we write $J \rightarrow J'$ to denote the existence of a homomorphism from J to J' in the sense of Nešetřil and Ossona de Mendez (i.e., not distinguishing two types of domain elements). As mentioned before, if the active domains of J and J' contain nulls only, then this notion of homomorphism coincides with the one considered in the context of schema mappings and data exchange (which is the one we used here).

Definition 11 [Left distance in [15]] Let \mathbf{T} be a schema and let J, J' be two \mathbf{T} -instances.

- The *similarity* $sim_h(J, J')$ between J and J' is the size of the active domain of a smallest instance B such that one of the following two conditions holds: (a) $B \rightarrow J$ and $B \not\rightarrow J'$; (b) $B \not\rightarrow J$ and $B \rightarrow J'$. If no such finite instance B exists, we let $sim_h(J, J') = \infty$.
- The *distance* $dist_h(J, J')$ between J and J' is the quantity $dist_h(J, J') = 2^{-sim_h(J, J')}$.

Nešetřil and Ossona de Mendez call this distance the “left distance”, because it is defined in terms of homomorphisms *from* other structures. This is to distinguish the notion from the “right distance” which is defined in terms of homomorphisms *to* other structures. For our purposes here, only the left distance is relevant. Because of the basic connection between homomorphisms and boolean conjunctive queries, it is easy to see that if J and J' are \mathbf{T} -instances, then the following statements are equivalent.

- $sim_h(J, J') = m$.
- m is the largest number such that J and J' satisfy the same boolean conjunctive queries with at most $m - 1$ variables.

How do the notions of sim_h of similarity and $dist_h$ of distance compare with our notions sim of similarity and $dist$ of distance? Clearly, this comparison is meaningful only when, in our setting, we consider singletons of instances and, moreover, the active domains of these instances contain nulls only. Recall that, according to the notation introduced in Definition 4, if J is a \mathbf{T} -instance whose active domain contains nulls only, then $v(J)$ is the set of all \mathbf{T} -instances that are isomorphic copies of J via

an isomorphism that renames nulls. The next observation is a direct consequence of Definitions 3 and 11, Lemma 1, and the preceding remarks.

Proposition 9 *Let \mathbf{T} be a schema and let J and J' be two \mathbf{T} -instances whose active domains contain nulls only. Then the following statements are true.*

- $sim(v(J), v(J')) = sim_h(J, J') - 1.$
- $dist(v(J), v(J')) = 2 \cdot dist_h(I, I').$

In what follows, we will write $\text{NInst}(\mathbf{T})$ to denote the set of all \mathbf{T} -instances whose active domain consists entirely of nulls. The pair $(\text{NInst}(\mathbf{T}), dist_h)$ is a pseudometric space, so a metric space can be obtained from it by passing to the equivalence classes $[J]$ of target instances J , where $[J]$ consists of all target instances that are homomorphically equivalent to J . As we did for the distance $dist$ and the pseudometric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), dist)$, we will identify each equivalence class with one of its members.

Cauchy sequences and limits arising from $dist_h$ are called *left Cauchy sequences* and *left limits* in [15]. Proposition 9 implies that if $(J_n)_{n \geq 1}$ is a sequence of elements of $\text{NInst}(\mathbf{T})$, then $(J_n)_{n \geq 1}$ is Cauchy with respect to the distance $dist_h$ if and only if the sequence $(v(J_n))_{n \geq 1}$ is Cauchy with respect to the distance $dist$. If $(J_n)_{n \geq 1}$ is a sequence of elements of $\text{NInst}(\mathbf{T})$, then we will write $\lim_{n \rightarrow \infty}^h J_n$ for the limit of the sequence $(J_n)_{n \geq 1}$ in the metric completion $(\text{NInst}(\mathbf{T})^*, dist_h^*)$ of the space $(\text{NInst}(\mathbf{T}), dist_h)$. Nešetřil and Ossona de Mendez obtained representations of the left limits of Cauchy sequences of instances by an approach that is based on the *homomorphism preorder* on instances and on *ideals* of partial orders.

The existence of homomorphisms between structures gives rise to the preorder \leq_h , where $L \leq_h J$ if $L \rightarrow J$. By passing to the equivalence classes $[J]$ of instances J in $\text{NInst}(\mathbf{T})$ modulo homomorphic equivalence, the preorder \leq_h becomes a partial order (also denoted by \leq_h), where $[L] \leq_h [J]$ means that there is a homomorphism from some member of $[L]$ to some member of $[J]$; this is the same as asserting that, for every pair (L', J') with $L' \in [L]$ and $J' \in [J]$, there is a homomorphism from L' to J' . As before, we will not distinguish between equivalence classes and their members. The partial order \leq_h extends to a partial order \leq_h^* on the metric completion $(\text{NInst}(\mathbf{T})^*, dist_h^*)$ of $(\text{NInst}(\mathbf{T}), dist_h)$ in the following way.

- If $(J_n)_{n \geq 1}$ and $(L_n)_{n \geq 1}$ are two Cauchy sequences from $\text{NInst}(\mathbf{T})$, then $\lim_{n \rightarrow \infty}^h L_n \leq_h^* \lim_{n \rightarrow \infty}^h J_n$ if for every m , there is a positive integer p such that for every $i \geq p$, we have that $\min\{|B| : B \rightarrow L_i \text{ and } B \not\rightarrow J_i\} \geq m$.
- As a special case, it is easy to see that if L is an element of $\text{NInst}(\mathbf{T})$ and $(J_n)_{n \geq 1}$ is a Cauchy sequence from $\text{NInst}(\mathbf{T})$, then $L \leq_h^* \lim_{n \rightarrow \infty}^h J_n$ holds if and only if there is a positive integer p such that for every $i \geq p$, we have that $L \rightarrow J_i$ (this is the special case of $\lim_{n \rightarrow \infty}^h L_n \leq_h^* \lim_{n \rightarrow \infty}^h J_n$ in which $L_n = L$, for all n).

Let (X, \leq) be a (finite or infinite) partially ordered set.

- A *downset* is a subset F of X with the property that for all $x \in F$ and $y \leq x$, also $y \in F$ holds.
- An *ideal* is a downset F with the additional property that for all x and y in F , there exists z in F such that both $x \leq z$ and $y \leq z$ hold.

In [15], it is shown that there is a correspondence between left limits of Cauchy sequences from $\text{NInst}(\mathbf{T})$ and ideals in the partial order $(\text{NInst}(\mathbf{T}), \leq_h)$. Before presenting this correspondence, we need to introduce a piece of notation.

If \mathcal{X} is a set of \mathbf{T} -instances, then the *disjoint union* $\biguplus \mathcal{X}$ is the set $\bigcup \mathcal{Y}$, where \mathcal{Y} is an isomorphic copy of \mathcal{X} with nulls named apart. In other words, $\biguplus \mathcal{X}$ is the union of copies of all elements of \mathcal{X} (one copy of each element of \mathcal{X}) so that no two members in the union have nulls in common. Clearly, $\biguplus \mathcal{X}$ is unique up to isomorphisms that rename nulls.

Let $(J_n)_{n \geq 1}$ be a Cauchy sequence from $\text{NInst}(\mathbf{T})$ and let $\lim_{n \rightarrow \infty}^h J_n$ be the left limit of $(J_n)_{n \geq 1}$ in the metric completion $(\text{NInst}(\mathbf{T})^*, \text{dist}_h^*)$ of the space $(\text{NInst}(\mathbf{T}), \text{dist}_h)$. Consider the set

$$\{L \in \text{NInst}(\mathbf{T}) \mid L \leq_h^* \lim_{n \rightarrow \infty}^h J_n\}.$$

It is easy to see that this set is an ideal of $(\text{NInst}(\mathbf{T}), \leq_h)$. Indeed, it is a downset because homomorphisms compose. Moreover, if L_1 and L_2 are in $\{L \in \text{NInst}(\mathbf{T}) \mid L \leq_h^* \lim_{n \rightarrow \infty}^h J_n\}$, then so is the disjoint union $\biguplus\{L_1, L_2\}$ of L_1 and L_2 ; moreover, $L_i \leq_h \biguplus\{L_1, L_2\}$, for $i = 1, 2$. The following is a consequence of Lemma 9.6 and Corollary 9.3 in [15].

Proposition 10 ([15]) *The following statements are true for the complete metric space $(\text{NInst}(\mathbf{T})^*, \text{dist}_h^*)$ and the partial order $(\text{NInst}(\mathbf{T})^*, \leq_h^*)$.*

- *There is a bijection \mathcal{F} between $\text{NInst}(\mathbf{T})^*$ and the set of ideals of $(\text{NInst}(\mathbf{T}), \leq_h)$ given by $\mathcal{F}(\lim_{n \rightarrow \infty}^h J_n) = \{L \in \text{NInst}(\mathbf{T}) \mid L \leq_h^* \lim_{n \rightarrow \infty}^h J_n\}$.*
- *If $\lim_{n \rightarrow \infty}^h J_n$ is the left limit of a Cauchy sequence $(J_n)_{n \geq 1}$ from $\text{NInst}(\mathbf{T})$, then $\lim_{n \rightarrow \infty}^h J_n$ can be represented as the disjoint union of the associated ideal $\mathcal{F}(\lim_{n \rightarrow \infty}^h J_n)$, namely,*

$$\lim_{n \rightarrow \infty}^h J_n = \biguplus\{L \in \text{NInst}(\mathbf{T}) \mid L \leq_h^* \lim_{n \rightarrow \infty}^h J_n\}.$$

We now have all the conceptual and technical apparatus needed to establish a tight connection between the representations of limits given in Theorem 6 and the representation of limits given in Proposition 10.

Let $\{J_n\}_{n \geq 1}$ be a Cauchy sequence (w.r.t. the distance function dist_h) such that each J_n is a member of $\text{NInst}(\mathbf{T})$, i.e., each J_n is a \mathbf{T} -instance whose active domain

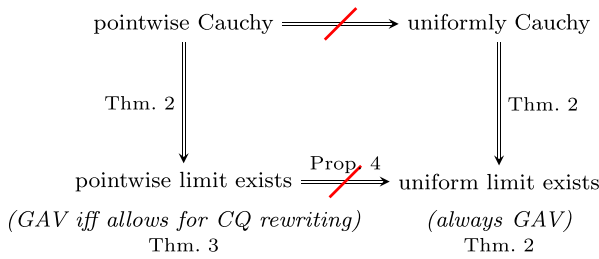


Fig. 1 Overall picture for GAV schema mappings

consists entirely of nulls. Let $\lim_{n \rightarrow \infty}^h J_n$ be its left-limit in the metric completion of $(NInst(\mathbf{T}), dist_h)$. As discussed earlier, the sequence $\{v(J_n)\}_{n \geq 1}$ is Cauchy (w.r.t. the distance function $dist$), so it has a limit $\lim_{n \rightarrow \infty} v(J_n)$ in the metric completion of $(P(Inst(\mathbf{T})), dist)$. The following proposition establishes the close relationship between these two limits.

Proposition 11 *Let $\{J_n\}_{n \geq 1}$ be a Cauchy sequence (w.r.t. the distance function $dist_h$) such that each J_n is a member of $NInst(\mathbf{T})$. Then*

$$\lim_{n \rightarrow \infty} v(J_n) = v(\lim_{n \rightarrow \infty}^h J_n).$$

Proof Theorem 6 tells us that

$$\lim_{n \rightarrow \infty} v(J_n) = \bigoplus \{q(\mathbf{a}) \mid q \in \mathbf{CQ} \wedge \exists p \forall i \geq p (\mathbf{a} \in cert(q, \{v(J_i)\}))\}.$$

Since the active domains of the elements of $v(J_n)$ consist entirely of nulls and are pairwise disjoint, we have that only boolean conjunctive queries q contribute to this expression. Moreover, by Lemma 1, the condition $\mathbf{a} \in cert(q, \{v(J_i)\})$ means that $cert(q, \{J_i\}) = true$ or, equivalently, that $J_i \models q$. As mentioned earlier, every boolean conjunctive query q can be identified with its canonical database D_q . Moreover, $J_i \models q$ if and only if $D_q \rightarrow J_i$. Thus, the preceding equation becomes

$$\lim_{n \rightarrow \infty} v(J_n) = \bigoplus \{L \in NInst(\mathbf{T}) \mid \exists p \forall i \geq p (L \rightarrow J_i)\}.$$

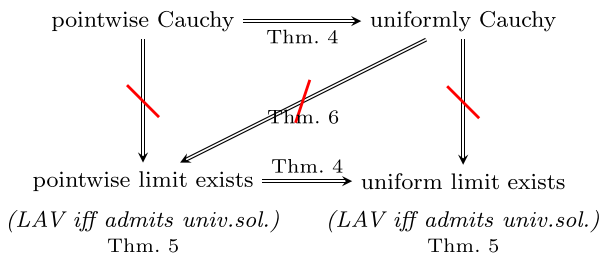


Fig. 2 Overall picture for LAV schema mappings

As explained earlier, the condition $\exists p \forall i \geq p (L \rightarrow J_i)$ is equivalent to the condition $L \leq_h \lim_{n \rightarrow \infty} J_n$, hence the preceding equation becomes

$$\lim_{n \rightarrow \infty} v(J_n) = \bigoplus \{L \in \text{NInst}(\mathbf{T}) \mid L \leq_h \lim_{n \rightarrow \infty} J_n\}.$$

This last equation and the second remark after Definition 8 imply that indeed

$$\lim_{n \rightarrow \infty} v(J_n) = v(\lim_{n \rightarrow \infty} J_n).$$

□

7 Concluding Remarks

In this paper, we have embarked on a systematic study of the limiting behavior of sequences of schema mappings using concepts and tools from metric spaces. For the important special cases of GAV and LAV mappings, our main results are summarized in Figs. 1 and 2.

In words, we have shown that, for GAV mappings, a pointwise Cauchy sequence need not be uniformly Cauchy; moreover, the existence of a pointwise limit does not imply the existence of a uniform limit. This cannot happen for LAV mappings. On the other side, a uniformly Cauchy sequence of LAV mappings need not even have a pointwise limit, which cannot happen for GAV mappings. We have also shown that structural properties of schema mappings can be used to characterize when the limit of a pointwise Cauchy sequence of GAV (or of LAV) mappings is equivalent to a GAV (or to a LAV) mapping. Finally, we have shown that infinite target instances and generalized mappings (i.e., schema mappings where target instances may be infinite) can be used to represent limits of Cauchy sequences of sets of target instances and limits of Cauchy sequences of arbitrary schema mappings.

We believe that the work reported here has laid the foundation for several interesting lines of subsequent investigations. We have seen that our results about sequences of LAV mappings extend in a natural way to sequences of premise-bounded GLAV mappings; an analogous extension of our results about sequences of GAV mappings to sequences of *conclusion-bounded* GLAV mappings is left for future work. We have also seen that there are sequences of LAV mappings for which no SO tgd is a uniform limit. Are there structural properties that characterize when a sequence of GLAV mappings has an SO tgd as a pointwise limit? In this vein, we have offered Conjecture 1. A related interesting open problem is whether schema mappings with target constraints are powerful enough to express pointwise limits or uniform limits of sequences of arbitrary GLAV schema mappings. We have some preliminary evidence that this is plausible, but much more work remains to be done.

We believe that the work reported in this paper provides a new perspective on the study of schema mappings by examining them from a dynamic viewpoint. As stated earlier, our original motivation came from schema-mapping optimization and, in particular, from the idea that “complex” schema mappings can be “approximated”

by “simpler” ones. It remains to be seen whether the work reported here will lead to applications to schema-mapping optimization. We believe, however, that the study of the limiting behavior of schema mappings via metric spaces is interesting in its own right.

We also note there are several areas in theoretical computer science where the study of limiting behavior of objects has produced results that were significant in their own right and also had fruitful consequences. For example, starting with the work of Fagin [4], there has been an extensive investigation of the asymptotic probabilities of logical properties and of 0-1 laws for various logics of interest in computer science. More recently, there has been a study of *profinite words*, which has found applications to automata theory and to the satisfiability problem for variants of monadic second-order logic (see, e.g., [17, 20]). Note that the profinite words form the completion of a metric space on words in which the distance is based on the size of the largest deterministic finite automaton needed to separate two words. Finally, the connection between graph limits in the monograph [15] by Nešetřil and Ossona de Mendez and the completion of the metric space $(\mathcal{P}(\text{Inst}(\mathbf{T})), d)$, which was mentioned in the previous section, may merit further exploration. It should also be pointed out that, motivated by the study of large-scale networks, there has been an extensive body of work on a notion of graph limits arising from converging sequences of *homomorphism densities*; a detailed account of this work is given in the monograph [13] by Lovász. In addition, Nešetřil and Ossona de Mendez [16] developed a general framework for limits of graphs and relational structures; in that framework, different fragments of first-order logic are used to define different notions of limits arising from converging sequences of the frequencies that first-order formulas in the fragment at hand are satisfied by an assignment (homomorphism densities correspond to the fragment consisting of all quantifier-free conjunctive queries). Homomorphisms, metric completions, and representations of limits of finite structures play a central role in [13, 16].

Acknowledgements The research of Reinhard Pichler, Emanuel Sallinger, and Vadim Savenkov was supported by the Austrian Science Fund, projects (FWF):P25207-N23 and (FWF):Y698, and the Vienna Science and Technology Fund, project ICT12-015. The research of Phokion Kolaitis on this paper was partially supported by NSF Grant IIS-1217869. The full version was completed while Kolaitis was visiting the Simons Institute for the Theory of Computing during the fall of 2016. The research of Emanuel Sallinger was supported by the EPSRC programme grant EP/M025268/1.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Arenas, M., Barceló, P., Libkin, L., Murlak, F.: Foundations of data exchange. Cambridge University Press, Cambridge (2014)
2. Arenas, M., Pérez, J., Reutter, J., Riveros, C.: The language of plain SO-tgds: composition, inversion and structural properties. *J. Comput. Syst. Sci.* **79**(6), 763–784 (2013)

3. Bernstein, P.A.: Applying model management to classical meta data problems. In: CIDR (2003)
4. Fagin, R.: Probabilities on finite models. *J. Symb. Log.* **41**(1), 50–58 (1976)
5. Fagin, R., Kolaitis, P.G.: Local transformations and conjunctive-query equivalence. In: PODS, pp. 179–190 (2012)
6. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: semantics and query answering. *Theor. Comput. Sci.* **336**(1), 89–124 (2005)
7. Fagin, R., Kolaitis, P.G., Nash, A., Popa, L.: Towards a theory of schema-mapping optimization. In: PODS, pp. 33–42 (2008)
8. Fagin, R., Kolaitis, P.G., Popa, L., Tan, W.C.: Composing schema mappings: second-order dependencies to the rescue. *ACM Trans. Database Syst.* **30**(4), 994–1055 (2005)
9. Fagin, R., Kolaitis, P.G., Popa, L., Tan, W.C.: Schema mapping evolution through composition and inversion. In: *Schema Matching and Mapping*, pp. 191–222. Springer (2011)
10. Feinerer, I., Pichler, R., Sallinger, E., Savenkov, V.: On the undecidability of the equivalence of second-order tuple generating dependencies. *Inf. Syst.* **48**, 113–129 (2015)
11. Kolaitis, P.G.: Schema mappings, data exchange, and metadata management. In: PODS, pp. 61–75 (2005)
12. Lenzerini, M.: Data integration: a theoretical perspective. In: PODS, pp. 233–246 (2002)
13. Lovász, L.: Large networks and graph limits, volume 60 of colloquium publications. American Mathematical Society (2012)
14. Madhavan, J., Halevy, A.Y.: Composing mappings among data sources. In: VLDB, pp. 572–583 (2003)
15. Nešetřil, J., de Mendez, P.O.: Sparsity - graphs, structures, and algorithms, volume 28 of algorithms and combinatorics. Springer, Berlin (2012)
16. Nešetřil, J., de Mendez, P.O.: A unified approach to structural limits, and limits of graphs with bounded tree-depth. [arXiv:1303.6471](https://arxiv.org/abs/1303.6471) (2013)
17. Pin, J.: Profinite methods in automata theory. In: STACS, pp. 31–50 (2009)
18. Shirali, S., Vasudeva, H.: *Metric spaces*. Springer, Berlin (2006)
19. ten Cate, B., Kolaitis, P.G.: Structural characterizations of schema-mapping languages. In: ICDT, pp. 63–72 (2009)
20. Torunczyk, S.: Languages of profinite words and the limitedness problem. In: ICALP, pp. 377–389 (2012)