RELATIONAL LATTICES: AN INTRODUCTION

TADEUSZ LITAK, JAN HIDDERS, AND SZABOLCS MIKULÁS

We study an interpretation of lattice connectives as *natural join* and *inner union* between database relations with non-uniform headers. To the best of our knowledge, this interpretation was proposed first by database researchers in [Tropashko, 2005, Spight and Tropashko, 2006]. It does not seem to have attracted the attention of algebraists even though connections between algebraic logic and relational databases have been discussed in the literature (see for example [Imieliński and Lipski, 1984] or [Düntsch and Mikulás, 2001]). This interpretation seems also related to an alternative algebraization of first-order logic proposed by [Craig, 1974].

The connective *natural join* (not to be confused with *lattice join*—in fact, in our setting it corresponds to *lattice meet*) is one of the basic operations of Codd's *(named) relational algebra* (see [Abiteboul et al., 1995, Codd, 1970] for basic information). Incidentally, it is also one of the few genuine algebraic operations—i.e., defined for all arguments. Codd's "algebra" from a mathematical point of view is only a *partial algebra*: some operations are defined only between relations with suitable headers. This includes another Codd's operation potentially dual to natural join: *(set) union*. Apart from the issues of mathematical elegance and generality, this partial nature of operations has also some unpleasant practical consequences. For example, queries which do not observe constraints on headers can *crash* (see, e.g., [Van den Bussche et al., 2007]).

It turns out, however, that it is possible to generalize the union operation to *inner union* defined on all elements of the algebra and lattice-dual to natural join. By adding suitable constants, this signature allows also to capture remaining basic database operations on queries like *projection*, *selection* and *renaming*, although this issue is not pursued here. This approach appears more natural and has several advantages over the embedding of relational "algebras" in cylindric algebras proposed in [Imieliński and Lipski, 1984]. For example, we avoid an artificial uniformization of headers and hence one can prove that queries formed with the use of proposed connectives enjoy the *domain independence property* (see, e.g., [Abiteboul et al., 1995] for a discussion of its importance in databases).

Let $\mathcal{A}$ be a set of *attribute names* $\mathcal{A}$ and let $\mathcal{D}$ be a set of *domain values*. A *header* is a subset of $\mathcal{A}$. A *H-sequence from $\mathcal{D}$* is a function $x : H \to \mathcal{D}$, i.e., an element of ${}^{H}\mathcal{D}$. The *restriction of a sequence x to a header H* is defined as $x[H] := \{(a, v) \in x \mid a \in H\}$ with the convention that $x[H] = \emptyset$ if $H \cap h(x) = \emptyset$. We generalize this to the *projection of a set of sequences X to a header H* which is $X[H] := \{x[H] \mid x \in X\}$. A *relation* is a pair $r = (H, B)$ where $H \subseteq \mathcal{A}$ is the *header of the relation* and $B \subseteq {}^{H}\mathcal{D}$ the *body of the relation*. The collections of all relations over $\mathcal{D}$ whose headers are contained in $\mathcal{A}$ will be denoted as $R(\mathcal{D}, \mathcal{A})$. For the relations $r = (H_r, B_r)$ and $s = (H_s, B_s)$ we define the *natural join* $r \otimes s$, and *inner union* $r \oplus s$:

$$\begin{aligned} r \otimes s &:= (H_r \cup H_s, \{x \in {}^{H_r \cup H_s}\mathcal{D} \mid x[H_r] \in B_r \quad \text{AND} \quad x[H_s] \in B_s\}) \\ r \oplus s &:= (H_r \cap H_s, \{x \in {}^{H_r \cap H_s}\mathcal{D} \mid x \in B_r[H_s] \quad \text{OR} \quad x \in B_s[H_r]\}) \end{aligned}$$

The element $\mathbf{H} = (\emptyset, \emptyset)$ plays a special role, as $(H, B) \otimes \mathbf{H} = (H, \emptyset)$. Hence, r_1 and r_2 have the same headers iff $\mathbf{H} \otimes r_1 = \mathbf{H} \otimes r_2$. Note also that the projection of r_1 to the header of r_2 can be defined as $r_1 \oplus (\mathbf{H} \otimes r_2)$. In fact, we can take $\mathbf{H} \otimes r$ to represent the header of r. We denote $(R(\mathcal{D}, \mathcal{A}), \otimes, \oplus, \mathbf{H})$ as $\mathfrak{R}^{\mathbf{H}}(\mathcal{D}, \mathcal{A})$ and call *the relational lattice over ($\mathcal{D}$,$\mathcal{A}$)*. $\mathfrak{R}(\mathcal{D}, \mathcal{A})$ denotes its *lattice reduct* $(R(\mathcal{D}, \mathcal{A}), \otimes, \oplus)$. These names are justified by the following

Lemma 1. *For any domain $\mathcal{D}$ and any collection of attributes $\mathcal{A}$, $\mathfrak{R}(\mathcal{D}, \mathcal{A})$ is a lattice.*

Proof. Define $Dom = \mathcal{A} \cup {}^{\mathcal{A}}\mathcal{D}$ and for any $X \subseteq Dom$ let

$$Cl(X) = X \cup \{x \in {}^{\mathcal{A}}\mathcal{D} \mid \exists y \in X \cap {}^{\mathcal{A}}\mathcal{D}.\forall a \in \mathcal{A} - X.x(a) = y(a)\}$$

Date: March 22, 2009.

Table 1. Proposed Axioms for Abstract Classes of Relational Lattices

Class $\underline{R}^{\mathbf{H}}$ in the full signature of L (with constant H)

all lattice axioms

AxRH1:	$\mathsf{H}\cdot x\cdot(y+z)+y\cdot z = (\mathsf{H}\cdot x\cdot y+z)\cdot(\mathsf{H}\cdot x\cdot z+y)$
AxRH2:	$x\cdot(y+z) = x\cdot(z+\mathsf{H}\cdot y)+x\cdot(y+\mathsf{H}\cdot z)$
AxRL1:	$x\cdot y+x\cdot z = x\cdot(y\cdot(x+z)+z\cdot(x+y))$

Class $\underline{R}$ in the lattice signature (without H)

all lattice axioms

AxRL1	$x\cdot y+x\cdot z = x\cdot(y\cdot(x+z)+z\cdot(x+y))$
AxRL2:	$t\cdot((x+y)\cdot(x+z)+(u+w)\cdot(u+v)) =$
	$= t\cdot((x+y)\cdot(x+z)+u+w\cdot v)+t\cdot((u+w)\cdot(u+v)+x+y\cdot z)$
	(in the full language L, AxRL2 is derivable from AxRH1 and AxRH2 above)

In other words, $Cl(X)$ is the sum of $X\cap\mathcal{A}$ (the set of attributes contained in X) with the cylindrification of $X\cap{}^{\mathcal{A}}\mathcal{D}$ along the axes in $X\cap\mathcal{A}$. It is straightforward to verify Cl is a closure operator and hence Cl-closed sets form a lattice. It remains to observe $\mathfrak{R}(\mathcal{D},\mathcal{A})$ is isomorphic to this lattice. □

The lattice order given by these operations is $(H_r, B_r) \sqsubseteq (H_s, B_s)$ iff $(H_s \subseteq H_r$ AND $B_r[H_s] \subseteq B_s)$. We propose an alternative name *Tropashko lattices* to honor the inventor of these structures. Note that a Tropashko lattice contain a top element $\top = (\emptyset, \{\emptyset\})$ and a bottom element $\bot = (\mathcal{A}, \emptyset)$. We do not include them in the signature as expressions containing $\bot$ would not enjoy the domain independence property mentioned above. For the same reason, we do not include a constant element $\mathbf{U} = (\mathcal{A}, {}^{\mathcal{A}}\mathcal{D})$ proposed by the inventor.

Let $\mathcal{R}^{\mathbf{H}}_{fin}$ denote the subalgebra closure of the class of relational lattices of the form $\mathfrak{R}(\mathcal{D},\mathcal{A},\mathbf{H})$ where both $\mathcal{D}$ and $\mathcal{A}$ are assumed to be finite and let $\mathcal{R}^{\mathbf{H}}_{inf}$ denote the subalgebra closure of the class of all relational lattices (with no restrictions on $\mathcal{D}$ and $\mathcal{A}$). Similarly, let $\mathcal{R}_{fin}$ and $\mathcal{R}_{inf}$ denote the lattice reducts of respective classes. Let $L = \{\cdot, +, \mathsf{H}\}$ be an algebraic language of the signature suitable for $\mathcal{R}^{\mathbf{H}}_{inf}$— i.e., the first two connectives are interpreted as $\otimes$ and $\oplus$, respectively, and H as $\mathbf{H}$. It is natural to ask how Tropashko lattices relate to already investigated varieties of lattices and weak forms of distributivity (see [Jipsen and Rose, 1992, Jipsen and Rose, 1998, Stern, 1999]). As it turns out, the results are mostly negative.

Theorem 2. *$\mathcal{R}_{fin}$ (and hence $\mathcal{R}_{inf}$) does not have any of the following properties (see the above references for definitions): semidistributivity (and hence also almost distributivity and neardistributivity), upper- or lower- semimodularity (and hence also modularity), local distributivity/local modularity, the Jordan-Dedekind chain condition, supersolvability.*

Many such properties can be disproved by observing that $\mathcal{R}(\{0,1\},\{a\}))$ is isomorphic to L_4, one of the covers of the non-modular lattice N_5 in [McKenzie, 1972] (see also [Jipsen and Rose, 1998]): a routine counterexample for most of them. Nevertheless, the variety generated by $\mathcal{R}_{fin}$ or $\mathcal{R}_{inf}$ is not the variety of all lattices.

Theorem 3. *AxRL1 and AxRL2 in Table 1 are valid in $\mathcal{R}_{inf}$ and consequently in $\mathcal{R}_{fin}$. Moreover, neither equation implies the other in the class of all lattices and consequently none is valid in this class.*

AxRL1 comes from [Padmanabhan et al., 2007] as an example of an equation which forces *the Huntington property* (distributivity under unique complementation). As for equations in the full signature of L, we have

Theorem 4. *AxRH1 and AxRH2 in Table 1 are valid equations of $\mathcal{R}^{\mathbf{H}}_{inf}$ and both are independent from each other even in presence of AxRL1. Taken together with lattice axioms, AxRH1 and AxRH2 jointly allow to derive AxRL2 and the following equations and quasi-equations:*

Quasi1: $x + y = x + z \quad \Rightarrow \quad x\cdot(y + z) = x\cdot y + x\cdot z$
Quasi2: $\mathsf{H}\cdot(x + y) = \mathsf{H}\cdot(x + z) \quad \Rightarrow \quad x\cdot(y + z) = x\cdot y + x\cdot z$
Eq1: $\mathsf{H}\cdot x\cdot(y + z) = \mathsf{H}\cdot x\cdot y + \mathsf{H}\cdot x\cdot z$
Eq2: $\mathsf{H}\cdot x + x\cdot y = x\cdot(y + \mathsf{H}\cdot x)$

We conjecture that the equational theory of $\mathcal{R}^{\mathbf{H}}_{inf}$ (and $\mathcal{R}^{\mathbf{H}}_{fin}$, as we conjecture both theories are equal) is either axiomatized by the equations AxRH1, AxRH2 and possibly AxRL1 (the last one may still turn out to be derivable as well) or not finitely axiomatizable due to some standard rainbow-style argument from algebraic logic. Even if the latter is the case, we propose these three axioms for the abstract, axiomatic class of relational lattices $\underline{R}^{\mathbf{H}}$—just as the abstract axiomatic classes of cylindric or relation algebras parallel the concrete ones.

From a database point of view, axiomatization and decidability results for the quasi-equational theory would be of even more interest. Many database constraints can be formulated as algebraic equations and hence reasoning over them can be reduced to quasi-equational reasoning. Besides, while it is doubtful whether the SP-closure of the class of Tropashko lattices is a variety, it does form a recursively axiomatizable quasivariety.

Theorem 5. *$\mathcal{R}^{\mathbf{H}}_{inf}$ and $\mathcal{R}_{inf}$ are pseudoelementary classes and hence are closed under ultraproducts. Moreover, the axiomatization in the extended elementary language is finite.*

Corollary 6. *The SP-closures of $\mathcal{R}^{\mathbf{H}}_{inf}$ and $\mathcal{R}_{inf}$ are quasi-equational classes. The quasi-equational theories of $\mathcal{R}^{\mathbf{H}}_{inf}$ and $\mathcal{R}_{inf}$ are recursively enumerable (the same applies to universal and elementary theories of these classes).*

There were some indications that the considered choice of connectives may lead to positive results concerning decidability even for quasi-equational theories. There is an elegant procedure known as *the chase* (see [Abiteboul et al., 1995]) applicable for certain classes of queries and database constraints similar to those that can be expressed with natural join and inner union. To our surprise, however, it turned out that as far as quasi-equational decidability results are concerned, relational lattices belong in the same category as Tarski's relation algebras, cylindric algebras or other "untamed" structures from algebraic logic. Moreover, the proof does not even require all the axioms introduced in Table 1. Let $\underline{RH1}$ be the variety of L-algebras axiomatized by the lattice axioms and AxRH1. It is straightforward to verify that $\mathcal{R}^{\mathbf{H}}_{fin} \subset \mathcal{R}^{\mathbf{H}}_{inf} \subset SP(\mathcal{R}^{\mathbf{H}}_{inf}) \subseteq \underline{R}^{\mathbf{H}} \subset \underline{RH1}$.

Definition 7. *Let $\overline{e} = (u_0, u_1, u_2, e_0, e_1)$ be an arbitrary 5-tuple of variables. We abbreviate $u_0\cdot u_1\cdot u_2$ as u. For arbitrary L-terms s, t define $\mathbf{c}^{\overline{e}}_0 \langle t\rangle = u\cdot(\mathsf{H}\cdot u_1\cdot u_2 + u\cdot t)$, $\mathbf{c}^{\overline{e}}_1 \langle t\rangle = u\cdot(\mathsf{H}\cdot u_0\cdot u_2 + u\cdot t)$, $\mathbf{c}^{\overline{e}}_2 \langle t\rangle = u\cdot(\mathsf{H}\cdot u_0\cdot u_1 + u\cdot t)$ and $s \circ^{\overline{e}} t = \mathbf{c}^{\overline{e}}_2 \left\langle \mathbf{c}^{\overline{e}}_1 \left\langle e_0\cdot \mathbf{c}^{\overline{e}}_2 \langle s\rangle\right\rangle \cdot \mathbf{c}^{\overline{e}}_0 \left\langle e_1\cdot \mathbf{c}^{\overline{e}}_2 \langle s\rangle\right\rangle\right\rangle$. Let $T_n(x_1, \ldots, x_n)$ be the collection of all semigroup terms in n variables. Let $\overline{e} = (x_{n+1}, \ldots, x_{n+5})$ and define the translation $\tau^{\overline{e}}$ of semigroup terms as follows: $\tau^{\overline{e}}(x_i) = x_i$ for $i \leq n$ and $\tau^{\overline{e}}(s \circ t) = s \circ^{\overline{e}} t$ for any $s, t \in T_n(x_1, \ldots, x_n)$.*

Theorem 8. *Tfae for any $p_0, \ldots, p_m, r_0, \ldots, r_m, s, t \in T_n(x_1, \ldots, x_n)$:*

(I) The quasiequation $\forall x_1, \ldots, x_n.(p_0 = r_0 \,\&\, \ldots \,\&\, p_m = r_m \Rightarrow s = t)$ holds in all semigroups (finite semigroups)

(II) For $\overline{e} = (x_{n+1}, \ldots, x_{n+5})$ as in Definition 7, the quasiequation

Quasi3: $\forall x_0, x_1, \ldots, x_{n+5}.(\tau^{\overline{e}}(p_0) = \tau^{\overline{e}}(r_0) \,\&\, \ldots \,\&\, \tau^{\overline{e}}(p_m) = \tau^{\overline{e}}(r_m) \,\&$
$\& \, x_{n+4} = \mathbf{c}^{\overline{e}}_0 \langle x_{n+4}\rangle \,\&\, x_{n+5} = \mathbf{c}^{\overline{e}}_1 \langle x_{n+5}\rangle) \quad \Rightarrow \quad \tau^{\overline{e}}(s) \circ^{\overline{e}} \mathbf{c}^{\overline{e}}_1 \langle x_0\rangle = \tau^{\overline{e}}(t) \circ^{\overline{e}} \mathbf{c}^{\overline{e}}_1 \langle x_0\rangle)$

holds in every member of $\mathcal{R}^{\mathbf{H}}_{inf}$ (every member of $\mathcal{R}^{\mathbf{H}}_{fin}$).

(III) Quasi3 above holds in every member of $\underline{RH1}$ (finite member of $\underline{RH1}$).

This can be proved following a strategy employed in [Maddux, 1980] for CA_3, the need to simulate cylindrifications being the main additional complication in our case. Note that Maddux' proof already implied undecidability of the quasi-equational theory of negation-free cylindric algebras, i.e., *cylindric lattices* (see [Düntsch and Mikulás, 2001] for more on these structures; note that cylindric lattices are distributive and relational lattices are *not* subreducts of cylindric lattices!). Negation was only needed to reduce quasi-equations to equations using the fact that finitely-dimensional cylindric algebras have a discriminator term.

Corollary 9. *The quasi-equational theory of any class of algebras between $\mathcal{R}^{\mathbf{H}}_{fin}$ and $\underline{RH1}$ is undecidable.*

The corollary follows from results in [Gurevich, 1966] (see also [Gurevich and Lewis, 1984]) and [Post, 1947] (for finite and arbitrary semigroups, respectively). We leave as an open question whether the quasi-equational theories of $\mathcal{R}_{inf}$ and $\mathcal{R}_{fin}$ (i.e., of lattice reducts) are decidable and whether the quasi-equational theory of $\mathcal{R}^{\mathbf{H}}_{inf}$ is finitely axiomatizable (cf. Corollary 6 above). Note that by Harrop's criterion, the quasi-equational theory of $\mathcal{R}^{\mathbf{H}}_{fin}$ cannot be finitely axiomatizable.

Acknowledgements. The authors would like to thank Vadim Tropashko, Marshall Spight, Maarten Marx, Balder ten Cate and Jan Paredaens. The first author happily acknowledges the usefulness of programs Prover9 and Mace4 [McCune, , McCune, 2003] in the search for derivations and counterexamples.

References

[Abiteboul et al., 1995] Abiteboul, S., Hull, R., and Vianu, V. (1995). *Foundations of Databases*. Addison-Wesley.

[Codd, 1970] Codd, E. F. (1970). A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387.

[Craig, 1974] Craig, W. (1974). *Logic in Algebraic Form. Three Languages and Theories*, volume 72 of *Studies in Logic and the Foundations of Mathematics*. North Holland.

[Düntsch and Mikulás, 2001] Düntsch, I. and Mikulás, S. (2001). Cylindric structures and dependencies in relational databases. *Theor. Comput. Sci.*, 269(1-2):451–468.

[Gurevich, 1966] Gurevich, Y. (1966). The word problem for certain classes of semigroups. *Algebra and Logic*, 5:25–35.

[Gurevich and Lewis, 1984] Gurevich, Y. and Lewis, H. R. (1984). The word problem for cancellation semigroups with zero. *The Journal of Symbolic Logic*, 49(1):184–191.

[Imieliński and Lipski, 1984] Imieliński, T. and Lipski, W. (1984). The relational model of data and cylindric algebras. *J. Comput. Syst. Sci.*, 28(1):80–102.

[Jipsen and Rose, 1992] Jipsen, P. and Rose, H. (1992). *Varieties of Lattices*, volume 1533 of *Lecture Notes in Mathematics*. Springer Verlag.

[Jipsen and Rose, 1998] Jipsen, P. and Rose, H. (1998). Varieties of lattices. In Grätzer, G., editor, *General Lattice Theory*, pages 555–574. Brikhäuser Verlag. Appendix F to the second edition.

[Maddux, 1980] Maddux, R. (1980). The equational theory of CA_3 is undecidable. *The Journal of Symbolic Logic*, 45(2):311–316.

[McCune,] McCune, W. Prover9 and Mace4 website. `http://www.cs.unm.edu/~mccune/`.

[McCune, 2003] McCune, W. (2003). Mace4 reference manual and guide. *CoRR*, cs.SC/0310055.

[McKenzie, 1972] McKenzie, R. (1972). Equational bases and non-modular lattice varieties. *Trans. Amer. Math. Soc.*, 174:1–43.

[Padmanabhan et al., 2007] Padmanabhan, R., McCune, W., and Veroff, R. (2007). Lattice laws forcing distributivity under unique complementation. *Houston Journal of Mathematics*, 33(2):391–401.

[Post, 1947] Post, E. L. (1947). Recursive unsolvability of a problem of Thue. *The Journal of Symbolic Logic*, 12(1):1–11.

[Spight and Tropashko, 2006] Spight, M. and Tropashko, V. (2006). First steps in relational lattice. `http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0603044`.

[Stern, 1999] Stern, M. (1999). *Semimodular Lattices*, volume 73 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press.

[Tropashko, 2005] Tropashko, V. (2005). Relational algebra as non-distributive lattice. `http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0501053`.

[Van den Bussche et al., 2007] Van den Bussche, J., Van Gucht, D., and Vansummeren, S. (2007). A crash course on database queries. In *PODS '07: Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of Database Systems*, pages 143–154, New York, NY, USA. ACM.

(Tadeusz Litak and Szabolcs Mikulás) School of Computer Science and Information Systems, Birkbeck, University of London, WC1E 7HX London, UK

E-mail address: `{tadeusz,szabolcs}@dcs.bbk.ac.uk`

(Jan Hidders) Delft University of Technology, Elektrotechn., Wisk. and Inform., Mekelweg 4, 2628CD Delft, The Netherlands

E-mail address: `A.J.H.Hidders@tudelft.nl`