The equational theory of relational lattices (natural join and inner union) is decidable [1]

Luigi Santocanale
LIS, Aix-Marseille Université

TACL@Nice, June 18, 2019

[1] Appeared in FOSSCAS 2018, preprint on HAL:
https://hal.archives-ouvertes.fr/hal-01625134/

Plan

Algebra and lattices from (and for) databases

Many undecidable theories

Structure of relational lattices

Decidability of the equational theory of the relational lattices

Plan

Algebra and lattices from (and for) databases

Many undecidable theories

Structure of relational lattices

Decidability of the equational theory of the relational lattices

Operations on tables: the natural join (pullback)

Name	Surname	Item
Luigi	Santocanale	33
Alan	Turing	21

$\bowtie$

Item	Description
33	Book
33	Livre
21	Machine

$=$

Name	Surname	Item	Description
Luigi	Santocanale	33	Book
Luigi	Santocanale	33	Livre
Alan	Turing	21	Machine

Operations on tables: the inner union

Name	Surname	Item
Luigi	Santocanale	33
Alan	Turing	21

$\cup$

Name	Surname	Sport
Diego	Maradona	Football
Usain	Bolt	Athletics

$=$

Name	Surname
Luigi	Santocanale
Alan	Turing
Diego	Maradona
Usain	Bolt

Lattices from databases

Proposition. [Spight & Tropashko, 2006] The set of tables, whose columns are indexed by a subset of A and values are from a set D, is a lattice, with natural join as meet and inner union as join.

$\mathrm{R}(D, A)$ shall denote the lattice whose elements are tables, with columns indexed a subset of A and cells' values are from a set D.

A project (Tropashko): Rebuild Codd's relational algebra out of lattice theoretic building blocks.
See QBQL.

Lattices from databases

Proposition. [Spight & Tropashko, 2006] The set of tables, whose columns are indexed by a subset of A and values are from a set D, is a lattice, with natural join as meet and inner union as join.

R(D, A) shall denote the lattice whose elements are tables, with columns indexed a subset of A and cells' values are from a set D.

A project (Tropashko): Rebuild Codd's relational algebra out of lattice theoretic building blocks.
See QBQL.

For lattices of tables (the relational lattices):

$$\wedge \text{ is } \bowtie, \qquad \vee \text{ is } \cup .$$

Lattice terms = queries.

Plan

Algebra and lattices from (and for) databases

Many undecidable theories

Structure of relational lattices

Decidability of the equational theory of the relational lattices

A family of undecidable theories and problems

Theorem (Maddux)

The equational theory of 3*-dimensional diagonal free cylindric algebras is undecidable.*

Theorem (Hirsch and Hodkinson)

It is not decidable whether a finite simple relation algebra embeds into a concrete one (a powerset of a binary product).

Theorem (Hirsch, Hodkinson and Kurucz)

It is not decidable whether a finite mutimodal frame has a surjective p-morphism from a universal product frame.

Undecidable quasiequational theories of relational lattices

Theorem (Litak, Mikulás and Hidders, 2015)

The set of quasiequations in the signature $(\wedge, \vee, H)$ *that are valid on relational lattices is undecidable.*

Undecidable quasiequational theories of relational lattices

Theorem (Litak, Mikulás and Hidders, 2015)

The set of quasiequations in the signature $(\wedge, \vee, H)$ *that are valid on relational lattices is undecidable.*

This was refined to:

Theorem (S., RAMICS 2017)

The set of quasiequations in the signature $(\wedge, \vee)$ *that are valid on relational lattices is undecidable.*

where we actually proved a stronger result:

Theorem (S., RAMICS 2017)

It is undecidable whether a finite subdirectly irreducible lattice embeds into some $\mathsf{R}(D, A)$.

Main result

Theorem (S., FOSSACS 2018)

The equational theory of the relational lattices is decidable.

Plan

Algebra and lattices from (and for) databases

Many undecidable theories

Structure of relational lattices

Decidability of the equational theory of the relational lattices

The relational lattices $\mathsf{R}(D, A)$

A a set of attributes, D a set of values.

An element of $\mathsf{R}(D, A)$:

- a pair (α, Y) with $\alpha \subseteq A$ and $Y \subseteq D^{\alpha}$.

The relational lattices $\mathsf{R}(D, A)$

A a set of attributes, D a set of values.

An element of $\mathsf{R}(D, A)$:

- a pair (α, Y) with $\alpha \subseteq A$ and $Y \subseteq D^{\alpha}$.

The ordering:

- $(\alpha_1, Y_1) \leq (\alpha_2, Y_2)$ iff $\alpha_2 \subseteq \alpha_1$ and $Y_1 \upharpoonright^{\alpha_1}_{\alpha_2} \subseteq Y_2$

where:

- $\upharpoonright$ is direct image of restriction:

$$Y \upharpoonright^{\alpha_1}_{\alpha_2} = \{\, f_{\restriction_{\alpha_2}} \mid f : \alpha_1 \to D \,,\, f \in Y \,\} \,.$$

The relational lattices $\mathsf{R}(D, A)$

A a set of attributes, D a set of values.

An element of $\mathsf{R}(D, A)$:

- a pair (α, Y) with $\alpha \subseteq A$ and $Y \subseteq D^{\alpha}$.

The ordering:

- $(\alpha_1, Y_1) \leq (\alpha_2, Y_2)$ iff $\alpha_2 \subseteq \alpha_1$ and $Y_1 {\upharpoonright\!\upharpoonright}^{\alpha_1}_{\alpha_2} \subseteq Y_2$
  iff $\dots$ $Y_1 \subseteq i^{\alpha_2}_{\alpha_1}(Y_2)$

where:

- ${\upharpoonright\!\upharpoonright}$ is direct image of restriction:

$$Y {\upharpoonright\!\upharpoonright}^{\alpha_1}_{\alpha_2} = \{ f_{\upharpoonright_{\alpha_2}} \mid f : \alpha_1 \longrightarrow D \,,\, f \in Y \} \,.$$

- i is cylindrification (inverse image of restriction):

$$i^{\alpha_2}_{\alpha_1}(Y) = \{ f : \alpha_1 \longrightarrow D \mid f_{\upharpoonright_{\alpha_2}} \in Y \} \,.$$

Meet and join

$$(\alpha_1, Y_1) \wedge (\alpha_2, Y_2) = i^{\alpha_1}_{\alpha_1 \cup \alpha_2}(Y_1) \cap i^{\alpha_2}_{\alpha_1 \cup \alpha_2}(Y_2)\,,$$

$$(\alpha_1, Y_1) \vee (\alpha_2, Y_2) = Y_1 \restriction^{\alpha_1}_{\alpha_1 \cap \alpha_2} \cup Y_2 \restriction^{\alpha_2}_{\alpha_1 \cap \alpha_2}\;.$$

NB :

- $\mathrm{R}(D, A)$ is the Grothendieck construction of the functor

$$P(D^{(\,\cdot\,)}) : P(A)^{op} \longrightarrow \mathrm{Latt}_{\vee}\,.$$

Representation of R(D, A) via a closure operator

The Hamming/Priess_Crampe-Ribenboim ultrametric distance on D^A:

$$\delta(f, g) := \{ x \in A \mid f(x) \neq g(x) \} .$$

NB: this distance takes values in the join-semilattice $(P(A), \emptyset, \cup)$.

Representation of R(D, A) via a closure operator

The Hamming/Priess_Crampe-Ribenboim ultrametric distance on D^A:

$$\delta(f, g) := \{ x \in A \mid f(x) \neq g(x) \} .$$

NB: this distance takes values in the join-semilattice $(P(A), \emptyset, \cup)$.

Proposition. [Litak, Mikulás and Hidders 2015] R(D, A) is isomorphic to the lattice of closed subsets of $A + D^A$, where ...

Representation of R(D, A) via a closure operator

The Hamming/Priess_Crampe-Ribenboim ultrametric distance on D^A:

$$\delta(f, g) := \{ x \in A \mid f(x) \neq g(x) \} .$$

NB: this distance takes values in the join-semilattice $(P(A), \emptyset, \cup)$.

Proposition. [Litak, Mikulás and Hidders 2015] R(D, A) is isomorphic to the lattice of closed subsets of $A + D^A$, where ...

... a subset Z of $A + D^A$ is *closed* if

$$\left(\begin{array}{r} g \in D^A \cap Z \\ \delta(f, g) \subseteq A \cap Z \end{array} \right) \text{ implies } f \in Z .$$

Plan

Algebra and lattices from (and for) databases

Many undecidable theories

Structure of relational lattices

Decidability of the equational theory of the relational lattices

Ingredients

- Duality, for non-distributive lattices.
- Generalized ultrametric spaces, injectivity.
- Modal logic, (selective) filtration techniques, tableaux.
- A finite model theorem with bounding of size.

Generalized ultrametric spaces

A *generalized ultrametric space* over $P(A)$ is a pair (X, δ) with

- X a set,
- $\delta : X \times X \to P(A)$,

and s.t.

- $\delta(f, g) = \emptyset$ iff $f = g$,
- $\delta(f, g) \subseteq \delta(f, h) \cup \delta(h, g)$,
- $\delta(f, g) = \delta(g, f)$.

Lattices from generalized ultrametric spaces

Let (X, δ) be a generalized ultrametric space over some $P(A)$.

Lattices from generalized ultrametric spaces

Let (X, δ) be a generalized ultrametric space over some $P(A)$.

A pair $(\alpha, Y) \in P(A) \times P(X)$ is *closed* if

$$\left(\begin{array}{r} g \in Y \\ \delta(f, g) \subseteq \alpha \end{array} \right) \text{ implies } f \in Y\,.$$

Lattices from generalized ultrametric spaces

Let (X, δ) be a generalized ultrametric space over some $P(A)$.

A pair $(\alpha, Y) \in P(A) \times P(X)$ is *closed* if

$$\left(\begin{array}{r} g \in Y \\ \delta(f, g) \subseteq \alpha \end{array} \right) \text{ implies } f \in Y \,.$$

Let

$$\mathsf{L}(X, \delta) := \{\, (\alpha, Y) \mid (\alpha, Y) \text{ is closed } \}\,,$$

then $\mathsf{L}(X, \delta)$ is a lattice (w.r.t. $\subseteq$).

Lattices from generalized ultrametric spaces

Let (X, δ) be a generalized ultrametric space over some $P(A)$.

A pair $(\alpha, Y) \in P(A) \times P(X)$ is *closed* if

$$\left(\begin{array}{r} g \in Y \\ \delta(f, g) \subseteq \alpha \end{array} \right) \text{ implies } f \in Y\,.$$

Let

$$\mathsf{L}(X, \delta) := \{\, (\alpha, Y) \mid (\alpha, Y) \text{ is closed } \}\,,$$

then $\mathsf{L}(X, \delta)$ is a lattice (w.r.t. $\subseteq$).

Notice that $\mathsf{R}(D, A) = \mathsf{L}(D^A, \delta)$.

Injective generalized ultrametric spaces

Consider

$$X = \prod_{a \in A} X_a \,, \qquad \delta(x,y) = \{\, a \in A \mid x_a \neq y_a \,\} \,. \qquad (**)$$

Injective generalized ultrametric spaces

Consider

$$X = \prod_{a \in A} X_a\,, \qquad \delta(x,y) = \{\, a \in A \mid x_a \neq y_a \,\}\,. \qquad (**)$$

These are :

- Hamming graphs,
- Dependent product types,
- Partial products, sections, $\forall_!$,
- Universal product frames,
- ...

Injective generalized ultrametric spaces

Consider

$$X = \prod_{a \in A} X_a \,, \qquad \delta(x,y) = \{\, a \in A \mid x_a \neq y_a \,\}\,. \qquad (**)$$

These are :

- Hamming graphs,
- Dependent product types,
- Partial products, sections, $\forall_!$,
- Universal product frames,
- ...

Proposition. Spaces as in (**) are, up to iso, the injective (read: complete) spaces in the category of generalized ultrametric spaces.

Relational lattices as modal logic

The theory of the lattices $\mathsf{L}(X, \delta)$ is interpreted in a multidimensional $\mathbf{S5}^n$ modal logic:

$$\langle \alpha \rangle Y := \{ f \in D^A \mid \exists g \in Y \text{ s.t. } \delta(f, g) \subseteq \alpha \}, \text{ where } \alpha \subseteq A$$

Relational lattices as modal logic

The theory of the lattices $\mathsf{L}(X, \delta)$ is interpreted in a multidimensional $\mathbf{S5}^n$ modal logic:

$$\langle \alpha \rangle Y := \{ f \in D^A \mid \exists g \in Y \text{ s.t. } \delta(f, g) \subseteq \alpha \}, \text{ where } \alpha \subseteq A$$

If (X, δ) is injective, then:

$$\langle \alpha_1 \cup \alpha_2 \rangle Y \equiv \langle \alpha_1 \rangle \langle \alpha_2 \rangle Y$$

(Beck-Chevalley, Malcev, injectiveness, pairwise completeness)

Relational lattices as modal logic

The theory of the lattices $\mathsf{L}(X, \delta)$ is interpreted in a multidimensional $\mathbf{S5}^n$ modal logic:

$$\langle \alpha \rangle Y := \{ f \in D^A \mid \exists g \in Y \text{ s.t. } \delta(f, g) \subseteq \alpha \}, \text{ where } \alpha \subseteq A$$

If (X, δ) is injective, then:

$$\langle \alpha_1 \cup \alpha_2 \rangle Y \equiv \langle \alpha_1 \rangle \langle \alpha_2 \rangle Y$$

(Beck-Chevalley, Malcev, injectiveness, pairwise completeness)

Meet is conjunction, join is:

$$\begin{aligned}
(\alpha_1, Y_1) \vee (\alpha_2, Y_2) &= (\alpha_1 \cup \alpha_2, \langle \alpha_1 \cup \alpha_2 \rangle (Y_1 \cup Y_2)) \\
&= (\alpha_1 \cup \alpha_2, \langle \alpha_1 \cup \alpha_2 \rangle Y_1 \cup \langle \alpha_1 \cup \alpha_2 \rangle Y_2) \\
&= (\alpha_1 \cup \alpha_2, \langle \alpha_2 \rangle \langle \alpha_1 \rangle Y_1 \cup \langle \alpha_1 \rangle \langle \alpha_2 \rangle Y_2) \\
&= (\alpha_1 \cup \alpha_2, \langle \alpha_2 \rangle Y_1 \cup \langle \alpha_1 \rangle Y_2) \,.
\end{aligned}$$

A finite model theorem of bounded size

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.

A finite model theorem of bounded size

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.

- If $\mathrm{R}(D, A) \not\models t \leq s$, then $\mathrm{R}(E, B) \not\models t \leq s$, where

$$\mathtt{size}(\,\mathrm{R}(E, B)\,) = O(2^{2^{2^{\mathtt{size}(t,s)}}})\,.$$

A finite model theorem of bounded size

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.

- If $\mathrm{R}(D, A) \not\models t \leq s$, then $\mathrm{R}(E, B) \not\models t \leq s$, where

$$\mathtt{size}(\,\mathrm{R}(E, B)\,) = O(2^{2^{2^{\mathtt{size}(t,s)}}})\,.$$

- Construction reminiscent of Gabbay's selective filtration in modal logic.

Failure in a finite generalized ultrametric space

Suppose $\mathrm{R}(D, A), v \not\models t \leq s$, so there is $f \in A \cup D^A$ such that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$.

Failure in a finite generalized ultrametric space

Suppose $\mathrm{R}(D, A), v \not\models t \leq s$, so there is $f \in A \cup D^A$ such that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$.

- We can assume that $f \in D^A$.

Failure in a finite generalized ultrametric space

Suppose $\mathrm{R}(D, A), v \not\models t \leq s$, so there is $f \in A \cup D^A$ such that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$.

- We can assume that $f \in D^A$.
- There is a finite subset $T(f, t) \subseteq D^A$ witnessing that $f \in [\![t]\!]_v$.

Failure in a finite generalized ultrametric space

Suppose $\mathrm{R}(D, A), v \not\models t \leq s$, so there is $f \in A \cup D^A$ such that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$.

- We can assume that $f \in D^A$.
- There is a finite subset $T(f, t) \subseteq D^A$ witnessing that $f \in [\![t]\!]_v$.
- Consider the subspace induced by $T \subseteq D^A$:

$$(T, \delta) \hookrightarrow (D^A, \delta)$$

Failure in a finite generalized ultrametric space

Suppose $\mathsf{R}(D, A), v \not\models t \leq s$, so there is $f \in A \cup D^A$ such that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$.

- We can assume that $f \in D^A$.
- There is a finite subset $T(f, t) \subseteq D^A$ witnessing that $f \in [\![t]\!]_v$.
- Consider the subspace induced by $T \subseteq D^A$:

$$(T, \delta) \hookrightarrow (D^A, \delta)$$

Lemma (preservation of failures)

If $T(f, t) \subseteq T \subseteq D^A$, then

$$\mathsf{L}(T, \delta) \not\models t \leq s \,.$$

Failure with a finite Boolean algebra

Let T be finite.

- The lattice $\mathsf{L}(T, \delta)$ might still be infinite,
- …since it contains a copy of $P(A)$.
- We can find a finite Boolean sub-algebra $P(B)$ of $P(A)$ and
- consider T as a generalized ultrametric space (T, δ_B) over $P(B)$.

Failure with a finite Boolean algebra

Let T be finite.

- The lattice $\mathsf{L}(T, \delta)$ might still be infinite,
- ... since it contains a copy of $P(A)$.
- We can find a finite Boolean sub-algebra $P(B)$ of $P(A)$ and
- consider T as a generalized ultrametric space (T, δ_B) over $P(B)$.

Lemma (preservation of failures in the finite)

There is a finite subset $T(f,t) \subseteq D^A$ such that, if $T(f,t) \subseteq T \subseteq D^A$ and T is finite, then

$$\mathsf{L}(T, \delta_B) \not\models t \leq s\,.$$

Failures in an injective

- $\mathrm{L}(T, \delta_B)$ is a finite lattice.
- $\mathrm{L}(T, \delta_B)$ does not belong to the variety of the $\mathrm{R}(D, A)s$.
- We expand T to its (finite) injective hull $\overline{T}$.
- Then $\mathrm{L}(\overline{T}, \delta_B)$ belongs to the variety of the $\mathrm{R}(D, A)s$.

Theorem

Let $T_0 := T(f, t)$. Then the lattice $L(\overline{T_0}, \delta_B)$

- *is finite,*
- $\mathrm{L}(\overline{T_0}, \delta_B) \not\models t \leq s$,
- *satisfies all the equations satisfied by all the* $\mathrm{R}(D, A)s$.

Thanks for your attention !!!