The equational theory of the natural join and of the inner union is decidable[1]

Luigi Santocanale
LIF, Aix-Marseille Université

Meeting TICAMORE@Marseille, November 15-17, 2017

[1] Preprint available on HAL:
https://hal.archives-ouvertes.fr/hal-01625134/

Plan

Algebra from databases

Some undecidable theories

The structure of relational lattices

Decidability of the equational theory of the relational lattices

Plan

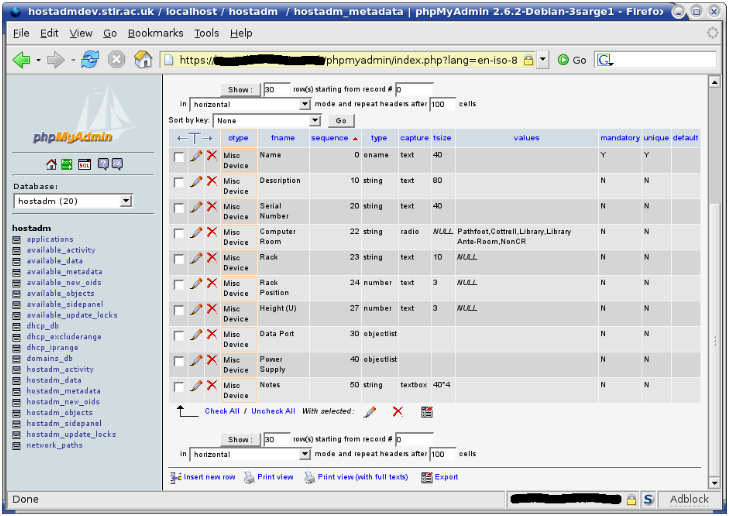
Algebra from databases

Some undecidable theories

The structure of relational lattices

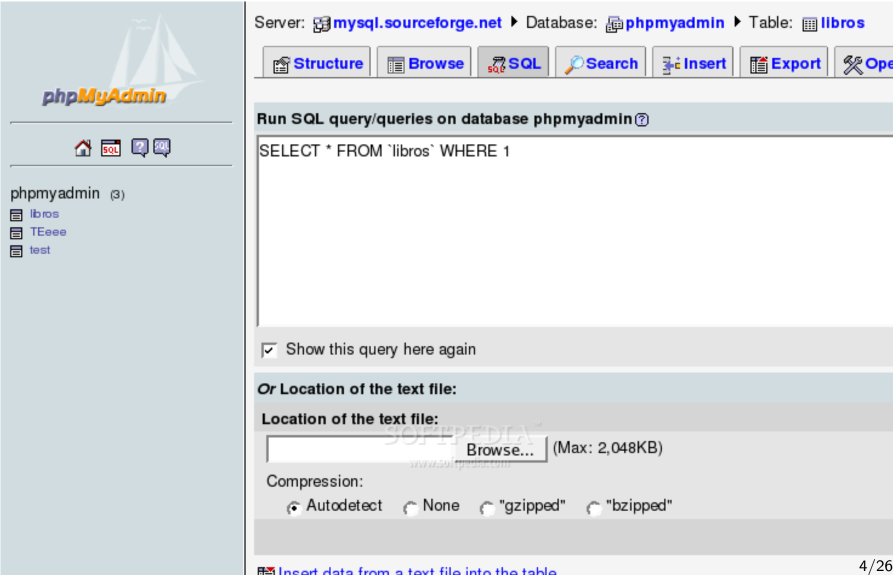
Decidability of the equational theory of the relational lattices

Databases, tables, sqls . . .



otype	fname	sequence	type	capture	tsize	values	mandatory	unique	default
Misc Device	Name	0	oname	text	40		Y	Y	
Misc Device	Description	10	string	text	80		N	N	
Misc Device	Serial Number	20	string	text	40		N	N	
Misc Device	Computer Room	22	string	radio	NULL	Pathfoot,Cottrell,Library,Library Ante-Room,NonCR	N	N	
Misc Device	Rack	23	string	text	10	NULL	N	N	
Misc Device	Rack Position	24	number	text	3	NULL	N	N	
Misc Device	Height (U)	27	number	text	3	NULL	N	N	
Misc Device	Data Port	30	objectlist				N	N	
Misc Device	Power Supply	40	objectlist				N	N	
Misc Device	Notes	50	string	textbox	40*4		N	N	

Databases, tables, sqls . . .



Operations on tables: the natural join

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.

Plan

Algebra from databases

Some undecidable theories

The 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.

n-dimensional diagonal free cylindric algebras, aka the multidimensional modal logic $\mathbf{S5}^n$

- n-dimensional cylindric algebras:
  algebraic modelling of first order logic with no more than n variables.
  Diagonal free: no equality.
- n-multimodal logic **S5**: we have n modal operators $\langle i \rangle$, $i = 1, \ldots, n$, each one is **S5**.
- $\mathbf{S5}^n$ is the n-multimodal logic determined by the universal product frames. These are product sets

  $$X_1 \times \ldots \times X_n$$

  with accessibility given by:

  $$(x_1, \ldots, x_n) R_i (y_1, \ldots, y_n) \text{ iff } x_j = y_j, \text{ for all } j \neq i.$$

- For $n \geq 3$, $\mathbf{S5}^n$ has the finite model property, it is recursively enumerable, yet it is not decidable.

Quasiequations, equations

- A *quasiequation* (definite Horn clause) is the universal closure of a formula of the form

$$s_1 = t_1 \wedge \ldots \wedge s_n = t_n \implies s_0 = t_0 \,,$$

with s_i, t_i, $i = 0, \ldots, n$, terms build over a fixed signature.

- The *quasiequational theory* of a class $\mathcal{K}$: the set of quasiequations holding in all elements of $\mathcal{K}$.

Quasiequations, equations

- A *quasiequation* (definite Horn clause) is the universal closure of a formula of the form

$$s_1 = t_1 \wedge \ldots \wedge s_n = t_n \implies s_0 = t_0\,,$$

with s_i, t_i, $i = 0, \ldots, n$, terms build over a fixed signature.

- The *quasiequational theory* of a class $\mathcal{K}$: the set of quasiequations holding in all elements of $\mathcal{K}$.

- An *equation* is a quasiequation as above with $n = 0$.

- The *equational theory* of a class $\mathcal{K}$: the set of equations holding in all elements of $\mathcal{K}$.
  See the standard Birkhoff's theorem.

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 (Santocanale, 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 (Santocanale 2017)

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

Plan

Algebra from databases

Some undecidable theories

The 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 $Y \subseteq D^{\alpha}$.

We have

$$(\alpha_1, Y_1) \leq (\alpha_2, Y_2) \text{ iff } \alpha_2 \subseteq \alpha_1 \text{ and } Y_1{\upharpoonright\!\upharpoonright}_{\alpha_2} \subseteq Y_2 \,.$$

NB :

- $\upharpoonright\!\upharpoonright$ is restriction:

$$Y{\upharpoonright\!\upharpoonright}_{\alpha} = \{\, f_{\upharpoonright_{\alpha}} \mid f \in Y \,\} \,.$$

Meet and join

$$
\begin{aligned}
(\alpha_1, Y_1) \wedge (\alpha_2, Y_2) &:= (\alpha_1 \cup \alpha_2, Y) \\
\text{where } Y &= \{ f \mid f_{\restriction \alpha_i} \in Y_i, i = 1, 2 \} \\
&= i_{\alpha_1 \cup \alpha_2}(Y_1) \cap i_{\alpha_1 \cup \alpha_2}(Y_2)\,,
\end{aligned}
$$

$$
\begin{aligned}
(\alpha_1, Y_1) \vee (\alpha_2, Y_2) &:= (\alpha_1 \cap \alpha_2, Y) \\
\text{where } Y &= \{ f \mid \exists i \in \{1, 2\}, \exists g \in Y_i \text{ s.t. } g_{\restriction \alpha_1 \cap \alpha_2} = f \} \\
&= Y_1 {\Vert\!\restriction}_{\alpha_1 \cap \alpha_2} \cup Y_2 {\Vert\!\restriction}_{\alpha_1 \cap \alpha_2}\ .
\end{aligned}
$$

NB :

- i is cylindrification:

$$
i_\alpha(Y) = \{ f \mid f_{\restriction \alpha} \in Y \}\,.
$$

Representation via closure operators

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 via closure operators

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 $(P(A), \emptyset, \cup)$.

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

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

Representation via closure operators

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 $(P(A), \emptyset, \cup)$.

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

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

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

Lattices from generalized ultrametric spaces

In a similar way, we can construct a lattice from any generalized ultrametric space (X, δ) over some $P(A)$.

A subset $Z \in P(A + X)$ is *closed* if

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

Lattices from generalized ultrametric spaces

In a similar way, we can construct a lattice from any generalized ultrametric space (X, δ) over some $P(A)$.

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

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

Lattices from generalized ultrametric spaces

In a similar way, we can construct a lattice from any generalized ultrametric space (X, δ) over some $P(A)$.

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

$$\begin{pmatrix} \delta(f,g) \subseteq \alpha \\ g \in Y \end{pmatrix} \text{ implies } f \in Z\,.$$

Thus we put

$$\mathsf{L}(X,\delta) := \{\, (\alpha, Y) \mid \langle\alpha\rangle Y \subseteq Y \,\}\,,$$

where

$$\langle\alpha\rangle Y = \{\, f \in X \mid \exists g \in Y \,\text{s.t.}\, \delta(f,g) \subseteq \alpha \,\}\,.$$

Lattices from generalized ultrametric spaces

In a similar way, we can construct a lattice from any generalized ultrametric space (X, δ) over some $P(A)$.

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

$$\begin{pmatrix} \delta(f,g) \subseteq \alpha \\ g \in Y \end{pmatrix} \text{ implies } f \in Z\,.$$

Thus we put

$$\mathsf{L}(X, \delta) := \{\, (\alpha, Y) \mid \langle \alpha \rangle Y \subseteq Y \,\}\,,$$

where

$$\langle \alpha \rangle Y = \{\, f \in X \mid \exists g \in Y \text{ s.t. } \delta(f,g) \subseteq \alpha \,\}\,.$$

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

Universal product spaces as injective generalized ultrametric spaces

Lemma

TFAE :

- (X, δ) *is injective in the category of generalized ultrametric spaces over* $P(A)$*,*
- (X, δ) *is, up to isomorphism, a universal product space:*

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

Remark : intuitively, injective means complete.

Relational lattices as modal logic

We can interpret the theory of the lattices $\mathsf{L}(X, \delta)$ in a sort of multidimensional $\mathbf{S5}^n$ modal logic. Modal operators are indexed by subsets of A:

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

If (X, δ) is injective, then we have the following logical equivalence:

$$\langle \alpha_1 \cup \alpha_2 \rangle Y = \langle \alpha_1 \rangle \langle \alpha_2 \rangle Y .$$

Meet is conjunction, where the join is:

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

Plan

Algebra from databases

Some undecidable theories

The structure of relational lattices

Decidability of the equational theory of the relational lattices

Strategy

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

Strategy

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.
- We show that if an inclusion $t \leq s$ fails in a lattice $\mathrm{R}(D, A)$, then it fails in a lattice $\mathrm{R}(E, B)$ of size $O(2^{2^{2^n}})$, with $n = size(t, s)$.

Strategy

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.
- We show that if an inclusion $t \leq s$ fails in a lattice $\mathrm{R}(D, A)$, then it fails in a lattice $\mathrm{R}(E, B)$ of size $O(2^{2^{2^n}})$, with $n = \mathit{size}(t, s)$.
- The method is reminiscent of Gabbay's selective filtration in modal logic.

Strategy

- Every lattice equation $t = s$ is equivalent to a pair of "inclusions", $t \leq s$ and $s \leq t$.
- We show that if an inclusion $t \leq s$ fails in a lattice $\mathrm{R}(D, A)$, then it fails in a lattice $\mathrm{R}(E, B)$ of size $O(2^{2^{2^n}})$, with $n = \mathit{size}(t, s)$.
- The method is reminiscent of Gabbay's selective filtration in modal logic.
- Thanks to TICAMORE: I would have not found this, if had not wondered about semantics vs syntactic methods for decidability in modal logic.

The tableau of a failure

Suppose $t \leq s$ is not vaild in all the $\mathsf{R}(D, A)$s.

- For some A, D and $v : \mathsf{X} \to \mathsf{R}(D, A)$, $[\![t]\!]_v \not\subseteq [\![s]\!]_v$.
- We can suppose that $f \in [\![t]\!]_v \setminus [\![s]\!]_v$ for some $f \in D^A$.

Lemma (preservation of failures)

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

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

- Above, (T, δ) is the subspace of (D^A, δ) induced by T.

Failures in a finite lattice

- The lattice $\mathsf{L}(T, \delta)$ might still be infinite, even if T is finite.
- This is because we have a copy of $P(A)$ inside it.

Failures in a finite lattice

- The lattice $\mathsf{L}(T, \delta)$ might still be infinite, even if T is finite.
- This is because we have a copy of $P(A)$ inside it.

Let T be finite.

- If $\mathcal{B}_T$ is the Boolean algebra generated by

$$\{\, \delta(g, h) \mid g, h \in T \,\} \cup \{\, A \cap v(x) \mid x \in Vars(t, s) \,\},$$

  then $\mathcal{B}_T \simeq P(B_T)$ for some finite subset B_T.
- We consider T as a generalized ultrametric space (T, δ_{B_T}) over $P(B_T)$.

Failures in a finite lattice

- The lattice $\mathsf{L}(T, \delta)$ might still be infinite, even if T is finite.
- This is because we have a copy of $P(A)$ inside it.

Let T be finite.

- If $\mathcal{B}_T$ is the Boolean algebra generated by

$$\{\, \delta(g,h) \mid g,h \in T \,\} \cup \{\, A \cap v(x) \mid x \in \mathit{Vars}(t,s) \,\}\,,$$

  then $\mathcal{B}_T \simeq P(B_T)$ for some finite subset B_T.
- We consider T as a generalized ultrametric space (T, δ_{B_T}) over $P(B_T)$.

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

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

Failures in a universal product frame

- The $\mathsf{L}(T, \delta_{B_T})$ is finite lattice.
- Yet, it does not need to be in the variety of the $\mathsf{R}(D, A)$s.

Lemma

For each finite T there is a finite $G(T) \subseteq D^A$ such that

- $T \subseteq G(T)$,
- $\mathcal{B}_T = \mathcal{B}_{G(T)}$,
- $(G(T), \delta_{B_{G(T)}})$ *is injective relative to* $P(B_T)$.

Corollary. Let $T_0 := T(f, t)$. Then the lattice $L(G(T_0), \delta_{B_{T_0}})$ is finite and

$$\mathsf{L}(G(T_0), \delta_{B_{T_0}}) \not\models t \leq s\,.$$

Summing up

- If $(G(T_0), \delta_{B_{T_0}})$ is injective then (up to isomorphism)

$$G(T_0) = \prod_{b \in B_{T_0}} X_b \,.$$

- Taking b_0 so X_{b_0} is of maximal cardinality, we can embed $G(T_0)$ into ${X_{b_0}}^{B_{T_0}}$:

$$\prod_{b \in B_{T_0}} X_b \subseteq \prod_{b \in B_{T_0}} X_{b_0} = {X_{b_0}}^{B_{T_0}} \,.$$

- Using injectivity, functoriality, and a bit of lattice theoretic tricks, we can show $\mathrm{L}(G(T_0), \delta_{B_{T_0}})$ is a homomorphic image of a sublattice of $\mathrm{R}(X_{b_0}, B_{T_0})$.
- Then

$$\mathrm{R}(X_{b_0}, B_{T_0}) \not\models t \leq s$$

otherwise we would have $\mathrm{L}(G(T_0), \delta_{B_{T_0}}) \models t \leq s$.

TICAMORE challenges

- complexity issues ;
- axiomatizations and automated tools ;
- completeness ;
- from labeled calculi to pure equational axioms ?