$\exists$-InvSat (a.k.a. pp-definability) is *co-NEXPTIME*-complete

Ross Willard

University of Waterloo, Canada

Dagstuhl Seminar 09441
October 30, 2009

Let Γ be a finite constraint language (on a domain D).

Definition

A **pp-formula over** Γ is a logical formula of the form

$$\exists \mathbf{y} \bigwedge_i atomic_i(\mathbf{x}, \mathbf{y})$$

where each $atomic_i(\mathbf{x}, \mathbf{y})$ is a constraint from $\Gamma \cup \{=_D\}$.

Such formulas ...

- are also called $\exists$-$CNF(\Gamma)$ formulas.
- define *implicit constraints* of a CSP(Γ) instance.
- define *conjunctive queries* in database theory.

Relations defined by such formulas are said to be ...

- *expressible* by Γ (by CSP theorists).
- *generated* by Γ (by algebraists).

$\underline{\exists\text{-InvSat}}$ is the following decision problem:

Input:

D – a finite domain

Γ – a finite set of relations on D

R – another relation on D.

Question: is R pp-definable from Γ?

We know that $\exists$-InvSat ...

- is at worst in co-*NExpTime* [from the Galois correspondence ...].
- is locally (i.e., for each *fixed* Γ) in P when $|D| = 2$ [Dalmau '00].
- is globally in P when $|D| = 2$ [Creignou, Kolaitis, Zanuttini '08].

Question: What is the exact complexity of $\exists$-InvSat in general?

Theorem (W)

∃-InvSat is co-NExpTime-complete.

Fine print

- In fact, there exists $k > 2$ such that ∃-InvSat restricted to k-element domains is co-*NExpTime*-complete.
- Remains hard even if, for some tuple $\mathbf{d}$, we know that $R \cup \{\mathbf{d}\}$ is pp-definable from Γ.
- Does not matter whether relations are represented as full truth tables or as lists of tuples.

Outline of proof

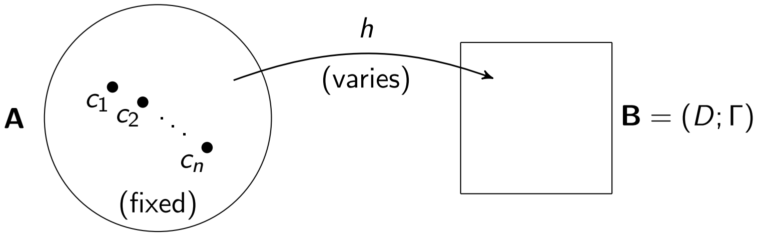
1. Characterize "pp-definability from Γ" (in terms of polymorphisms).
2. Find a nice *NExpTime*-complete problem X.
3. Reduce X to ¬∃-InvSat (via polymorphisms).

Step 1: Characterize pp-definability from Γ.

Write $\mathbf{B} = (D; \Gamma)$. Fix $n \geq 1$.

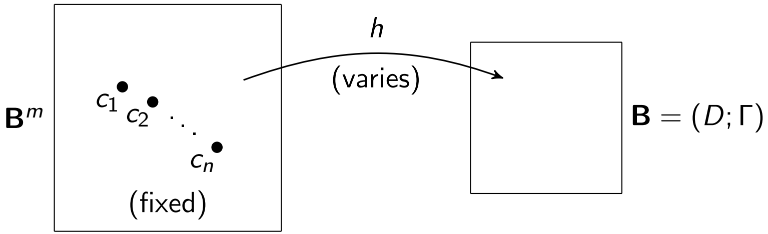
Recipe:

- Let $\mathbf{A} = (A; \ldots)$ be any finite structure (of same type as $\mathbf{B}$).
- Fix an n-tuple $\mathbf{c} = (c_1, \ldots, c_n) \in A^n$.
- Let $Hom_{\mathbf{A},\mathbf{B}}$ be the set of all homomorphisms $\mathbf{A} \xrightarrow{h} \mathbf{B}$.
- Collect all n-tuples $(h(c_1), \ldots, h(c_n)) \in D^n$ as h varies over $Hom_{\mathbf{A},\mathbf{B}}$:



Fact 1: $\{h(\mathbf{c}) \,:\, h \in Hom_{\mathbf{A},\mathbf{B}}\}$ is a typical relation pp-definable from Γ.

Specialize by assuming $\mathbf{A} = \mathbf{B}^m$ (for some $m \geq 1$).



(Thus $Hom_{\mathbf{B}^m,\mathbf{B}} = \{$all m-ary polymorphisms of $\Gamma\}$.)
Define

$$\begin{aligned} H(\mathbf{c}) &= \{h(\mathbf{c}) : h \in Hom_{\mathbf{B}^m,\mathbf{B}}\} \\ P(\mathbf{c}) &= \{p_i^m(\mathbf{c}) : 1 \leq i \leq m\} \subseteq H(\mathbf{c}) \end{aligned}$$

where p_i^m is the i-th dictator function of arity m.

Fact 2: If $P(\mathbf{c}) \subseteq R \subseteq H(\mathbf{c})$, then R is pp-definable from $\Gamma \Leftrightarrow R = H(\mathbf{c})$.

Step 2: Find a nice NExpTime-complete problem.

More precisely, a nice **tiling** problem. Roughly speaking, this involves:

- An unlimited supply of **tiles**, each having a **tile type** $\in \{t_1, \ldots, t_k\}$.
- A positive integer N.
- One then attempts to cover an $N \times N$ grid with tiles,

t_4	t_5	t_5	t_2	t_1	t_1	t_3	t_1
t_2	t_5	t_4	t_1	t_4	t_1	t_4	t_1
t_4	t_4	t_5	t_2	t_4	t_2	t_5	t_4
t_1	t_3	t_4	t_5	t_5	t_3	t_4	t_1
t_5	t_1	t_5	t_5	t_5	t_4	t_5	t_1
t_5	t_1	t_5	t_5	t_5	t_5	t_1	t_2
t_2	t_2	t_3	t_5	t_3	t_2	t_1	t_2
t_3	t_1	t_1	t_3	t_1	t_3	t_1	t_4

  so that horizontally adjacent and vertically adjacent tiles satisfy some given constraints.

More precisely:

Definition

Fix $N \geq 2$.

1. A **domino system** is a finite relational structure $\mathcal{D} = (\Delta; H, V)$ with H, V binary. ($\Delta =$ "tiles," $H =$ "horizontal," $V =$ "vertical.")
2. $[N] = \{0, 1, \dots, N-1\}$.
3. $\mathcal{C}_N$ denotes the structure $([N] \times [N]; \prec_1, \prec_2)$ where

$$\begin{aligned} \prec_1 &= \{(i,j),(i+1,j) : i,j \in [N],\ i < N-1\} \\ \prec_2 &= \{(i,j),(i,j+1) : i,j \in [N],\ j < N-1\}. \end{aligned}$$

4. An $N \times N$ **tiling by** $\mathcal{D}$ is a homomorphism $\tau : \mathcal{C}_N \to \mathcal{D}$.
5. Given $\mathbf{w} = (w_0, w_1, \dots, w_{m-1}) \in \Delta^m$ with $m \leq N$, we say that an $N \times N$ tiling τ **satisfies initial condition w** if $\tau(i, 0) = w_i\ \forall i < m$.

Fix a domino system $\mathcal{D} = (\Delta; H, V)$.

<u>$\text{EXPTILE}(\mathcal{D})$</u>, the *Exponential Tiling-by-$\mathcal{D}$ Problem*, is:

Input:

$\mathbf{w} \in \Delta^m$ for some $m \geq 2$.

Question: does there exist a $2^m \times 2^m$ tiling by $\mathcal{D}$ satisfying initial condition $\mathbf{w}$?

Fact 3: There exists $\mathcal{D}$ such that $\text{EXPTILE}(\mathcal{D})$ is *NEXPTIME*-complete.

Fine print: Can even restrict to inputs where m is a power of 2.

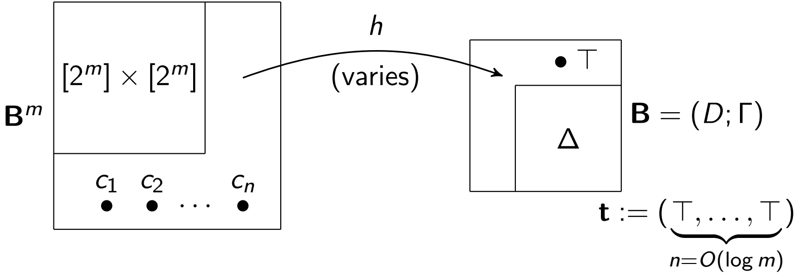
(Very nice.)

Step 3. Reduce ExpTile($\mathcal{D}$) to $\neg\exists$-InvSat.

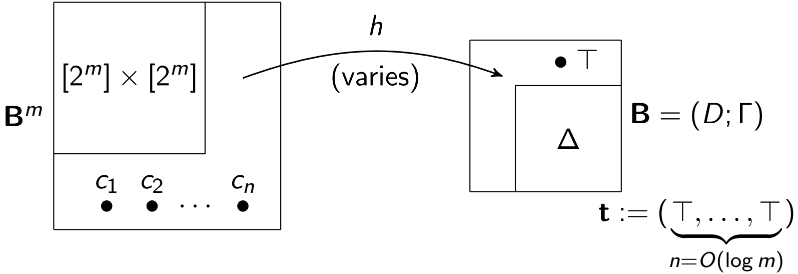
Fix an input $\mathbf{w} \in \Delta^m$ to $\textsc{ExpTile}(\mathcal{D})$, $m \geq 2$. (We must build D, Γ, R.)

- We'll need a set D (small) on which to build a structure $\mathbf{B} = (D; \Gamma)$.
- We'll encode a copy of $[2^m] \times [2^m]$ in D^m and a copy of Δ in D.
- We'll identify auxiliary parameters $c_i \in D^m$ $(1 \leq i \leq n)$ and $\top \in D$.
- We'll define Γ so that no polymorphism $h : \mathbf{B}^m \to \mathbf{B}$ can send $\mathbf{c} = (c_1, \ldots, c_n)$ to $\mathbf{t} = (\top, \ldots, \top)$, unless
  - h sends "$[2^m] \times [2^m]$" to "Δ," and
  - $h\restriction_{[2^m]\times[2^m]}$ is a $2^m \times 2^m$ tiling by $\mathcal{D}$ satisfying initial condition $\mathbf{w}$.
- Constraints:
  - The number of relations in Γ must be $O(m^c)$;
  - The arity of each relation in Γ must all be $O(\log m)$.
  - The number, n, of parameters c_i must be $O(\log m)$.

Recap: given input $\mathbf{w} \in \Delta^m$ to $\textsc{ExpTile}(\mathcal{D})$, we'll build



- We'll examine $\color{red}{H(\mathbf{c})} = \{h(\mathbf{c}) : h \in Hom_{\mathbf{B}^m,\mathbf{B}}\} \subseteq D^n$ and its subset $\color{blue}{P(\mathbf{c})} = \{p_i(\mathbf{c}) : i < m\}$.
- We'll achieve: $\forall h : \mathbf{B}^m \to \mathbf{B}$, $\quad h(\mathbf{c}) = \mathbf{t} \;\Leftrightarrow\; h$ encodes a $[2^m] \times [2^m]$ tiling by $\mathcal{D}$ satisfying i.c. $\mathbf{w}$. (Conversely, $\exists$ a tiling $\Rightarrow \exists$ such h.)
- Hence: $\mathbf{t} \in \color{red}{H(\mathbf{c})} \;\Leftrightarrow\;$ there exists a $2^m \times 2^m$ tiling by $\mathcal{D}$ satisfying the input initial condition $\mathbf{w}$.



Moreover, We'll find that there is an easily constructed n-ary relation R (not depending on $\mathbf{w}$), satisfying the following:

- $\mathbf{t} \notin R$.
- $R \cup \{\mathbf{t}\}$ is (easily) pp-definable from Γ.
- $P(\mathbf{c}) \subseteq R \subseteq H(\mathbf{c}) \subseteq R \cup \{\mathbf{t}\}$
- Hence

$$H(\mathbf{c}) = \begin{cases} R \cup \{\mathbf{t}\} & \text{if } \exists\ 2^m \times 2^m \text{ tiling by } \mathfrak{D} \text{ with i.c. } \mathbf{w} \\ R & \text{otherwise.} \end{cases}$$

R is pp-df$/\Gamma$ $\Leftrightarrow$ there does **not** exist such a tiling.

Details

1. Construct D.

$D =$

<table>
<tr><td colspan="2">00 01 10 11</td><td>∞</td></tr>
<tr><td>$\top$ $\bot$</td><td colspan="2" rowspan="3">Δ</td></tr>
<tr><td>a b</td></tr>
<tr><td>0 1</td></tr>
</table>

2. Encode Δ and $\top$ in D. ✓

3. Encode $[2^m] \times [2^m]$ in D^m.
   - (Assume $m = 8$.) To encode e.g. $(53, 188)$ we do:

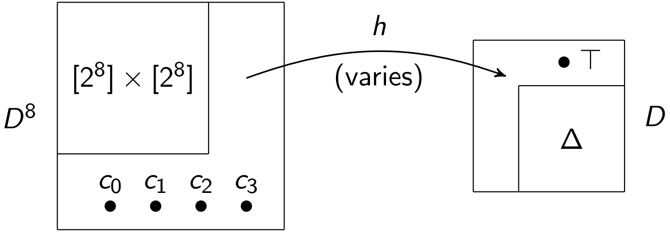
$$
\begin{aligned}
53 &= 10101100 \quad \text{(least significant bit at left)} \\
188 &= 00111101 \\
(53, 188) &\doteq (10, 00, 11, 01, 11, 11, 00, 01) \in D^8.
\end{aligned}
$$

4. Define auxiliary parameters $c_i \in D^m$.

- We'll use $\log_2 m + 1$ of them. When $m = 8$:

$$\begin{aligned} c_0 &= (0,1,0,1,0,1,0,1) \\ c_1 &= (0,0,1,1,0,0,1,1) \\ c_2 &= (0,0,0,0,1,1,1,1) \\ c_3 &= (\mathtt{b},\mathtt{b},\mathtt{a},\mathtt{b},\mathtt{a},\mathtt{a},\mathtt{a},\mathtt{b}). \end{aligned}$$

(Rule: $c_3(j) = \mathtt{a}$ iff $c_0(j), c_1(j), c_2(j)$ contains a subsequence 0,1.)



Next: add structure (Γ) to impose "tiling requirement" on h.

5. Enforcing the horizontal adjacency constraints.

Example ($m = 8$): if $h : \mathbf{B}^8 \to \mathbf{B}$ and

$$\begin{array}{rcll}
(151, 54) & \doteq & (10, 11, 11, 00, 11, 01, 00, 01) & x \\
(152, 54) & \doteq & (00, 01, 01, 10, 11, 01, 00, 01) & y \\
c_0 & = & (\,0\,,\,1\,,\,0\,,\,1\,,\,0\,,\,1\,,\,0\,,\,1\,) \;\xrightarrow{h} & \top \\
c_1 & = & (\,0\,,\,0\,,\,1\,,\,1\,,\,0\,,\,0\,,\,1\,,\,1\,) & \top \\
c_2 & = & (\,0\,,\,0\,,\,0\,,\,0\,,\,1\,,\,1\,,\,1\,,\,1\,) & \top \\
c_3 & = & (\,\mathtt{b}\,,\,\mathtt{b}\,,\,\mathtt{a}\,,\,\mathtt{b}\,,\,\mathtt{a}\,,\,\mathtt{a}\,,\,\mathtt{a}\,,\,\mathtt{b}\,) & \top
\end{array}$$

then we want this to imply $x, y \in \Delta$ and moreover $(x, y) \in H$.

Note that the "carry" in $151 + 1 = 152$ occurs in column 4.

We can build a 6-ary relation H_4 which will enforce this implication for all horizontally adjacent pairs $u, v \in [2^8] \times [2^8]$ where the carry in the x increment occurs in column 4.

Here it is:

$$
\begin{aligned}
H_4 \;=\; & \{(1y, 0y, c_0(j), c_1(j), c_2(j), c_3(j)) \;:\; y \in \{0,1\},\, 1 \le j < 4\} \\
\cup \;& \{(0y, 1y, 1, 1, 0, \mathtt{b}) \;:\; y \in \{0,1\}\} \\
\cup \;& \{(xy, xy, c_0(j), c_1(j), c_2(j), c_3(j)) \;:\; x, y \in \{0,1\},\, 4 < j \le 8\} \\
\cup \;& \{(x, y, \top, \top, \top, \top) \;:\; x, y \in \Delta \text{ and } (x,y) \in H\} \\
\cup \;& \{(x, y, b_0, b_1, b_2, b_3) \in \Delta^2 \times \{\top, \bot\}^4 \;:\; \bot \in \{b_0, b_2, b_2, b_3\}\} \\
\cup \;& \{(\infty, \infty, \infty, \infty, \infty, \infty)\}.
\end{aligned}
$$

Similarly, for each i we build a relation H_i to recognize carries at coordinate i.

Do the same thing for vertical adjacencies.

Ultimately, Γ will include:

- m relations $H_1, \ldots, H_m$ of arity $\log_2 m + 3$ for horizontal constraints.
- m relations $V_1, \ldots, V_m$ of arity $\log_2 m + 3$ for vertical constraints.
- m relations $I_1^{\mathbf{w}}, \ldots, I_m^{\mathbf{w}}$ of arity $\log_2 m + 2$ for the initial condition $\mathbf{w}$.

R will be the $n := (\log_2 m + 1)$-ary relation

$$\begin{aligned} R \;=\; & \quad P(\mathbf{c}) \\ & \cup \;\; \{\top, \bot\}^n \;\setminus\; \{(\top, \top, \ldots, \top)\} \\ & \cup \;\; \{(\infty, \infty, \ldots, \infty)\}. \end{aligned}$$

Open problems

In my construction of (D, Γ, R), D is independent of the input $\mathbf{w}$.

Hence there exists $k > 2$ such that $\exists$-InvSat(k) (i.e., restricted k-element domains) is co-*NExpTime*-complete.

Question 1. Can k be reduced to 3?

Question 2. Does there exist a fixed k and fixed Γ such that $\exists$-InvSat(k, Γ) is co-*NExpTime*-complete?

If "yes," then this would complement the corresponding result of M. Kozik [TCS 2008] on the algebraic side.

Thank you!