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

Ross Willard

University of Waterloo, Canada

Dagstuhl Seminar 09441
October 30, 2009
Let $\Gamma$ be a finite constraint language (on a domain $D$).

**Definition**

A **pp-formula over** $\Gamma$ is a logical formula of the form

$$\exists y \bigwedge_{i} \text{atomic}_i(x, y)$$

where each $\text{atomic}_i(x, y)$ is a constraint from $\Gamma \cup \{=D\}$.

Such formulas . . .

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

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

- *expressible* by $\Gamma$ (by CSP theorists).
- *generated* by $\Gamma$ (by algebraists).
$\exists$-InvSat is the following decision problem:

**Input:**
- $D$ – a finite domain
- $\Gamma$ – a finite set of relations on $D$
- $R$ – another relation on $D$.

**Question:** is $R$ pp-definable from $\Gamma$?

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

- is at worst in co-\textit{NExpTime} [from the Galois correspondence . . .].
- is locally (i.e., for each fixed $\Gamma$) 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)

$\exists$-InvSat is co-NEXPTIME-complete.

Fine print

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

Outline of proof

1. Characterize “pp-definability from $\Gamma$” (in terms of polymorphisms).
2. Find a nice NEXPTIME-complete problem $X$.
3. Reduce $X$ to $\neg \exists$-InvSat (via polymorphisms).
Step 1: Characterize pp-definability from $\Gamma$.

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

Recipe:

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

**Fact 1:** $\{h(c) : h \in \text{Hom}_{A,B}\}$ is a typical relation pp-definable from $\Gamma$. 

Ross Willard (Waterloo)
Specialize by assuming $A = B^m$ (for some $m \geq 1$).

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

Define

\[
H(c) = \{h(c) : h \in Hom_{B^m, B}\}
\]

\[
P(c) = \{p^m_i(c) : 1 \leq i \leq m\} \subseteq H(c)
\]

where $p^m_i$ is the $i$-th dictator function of arity $m$.

**Fact 2:** If $P(c) \subseteq R \subseteq H(c)$, then $R$ is pp-definable from $\Gamma \iff R = H(c)$.  

Ross Willard (Waterloo)
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 \( t_1, \ldots, t_k \).
- A positive integer \( N \).
- One then attempts to cover an \( N \times N \) grid with tiles,

```
<table>
<thead>
<tr>
<th>t_4</th>
<th>t_5</th>
<th>t_5</th>
<th>t_2</th>
<th>t_1</th>
<th>t_1</th>
<th>t_3</th>
<th>t_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>t_2</td>
<td>t_5</td>
<td>t_4</td>
<td>t_1</td>
<td>t_4</td>
<td>t_1</td>
<td>t_4</td>
<td>t_1</td>
</tr>
<tr>
<td>t_4</td>
<td>t_4</td>
<td>t_5</td>
<td>t_2</td>
<td>t_4</td>
<td>t_2</td>
<td>t_5</td>
<td>t_4</td>
</tr>
<tr>
<td>t_1</td>
<td>t_3</td>
<td>t_4</td>
<td>t_5</td>
<td>t_5</td>
<td>t_3</td>
<td>t_4</td>
<td>t_1</td>
</tr>
<tr>
<td>t_5</td>
<td>t_1</td>
<td>t_5</td>
<td>t_5</td>
<td>t_4</td>
<td>t_5</td>
<td>t_1</td>
<td>t_2</td>
</tr>
<tr>
<td>t_5</td>
<td>t_1</td>
<td>t_5</td>
<td>t_5</td>
<td>t_5</td>
<td>t_5</td>
<td>t_1</td>
<td>t_2</td>
</tr>
<tr>
<td>t_2</td>
<td>t_2</td>
<td>t_3</td>
<td>t_5</td>
<td>t_3</td>
<td>t_2</td>
<td>t_1</td>
<td>t_2</td>
</tr>
<tr>
<td>t_3</td>
<td>t_1</td>
<td>t_1</td>
<td>t_3</td>
<td>t_1</td>
<td>t_3</td>
<td>t_1</td>
<td>t_4</td>
</tr>
</tbody>
</table>
```

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. ($\Delta = \text{“tiles,” } H = \text{“horizontal,” } V = \text{“vertical.”}$)

2. $[N] = \{0, 1, \ldots, N - 1\}$.

3. $\mathcal{C}_N$ denotes the structure $([N] \times [N]; \prec_1, \prec_2)$ where
   
   $\prec_1 = \{(i, j), (i + 1, j) : i, j \in [N], i < N - 1\}$
   
   $\prec_2 = \{(i, j), (i, j + 1) : i, j \in [N], j < N - 1\}$.

4. An $N \times N$ **tiling by** $\mathcal{D}$ is a homomorphism $\tau : \mathcal{C}_N \to \mathcal{D}$.

5. Given $w = (w_0, w_1, \ldots, w_{m-1}) \in \Delta^m$ with $m \leq N$, we say that an $N \times N$ tiling $\tau$ **satisfies initial condition** $w$ if $\tau(i, 0) = w_i \ \forall i < m$. 
Fix a domino system $\mathcal{D} = (\Delta; H, V)$.

$\textbf{ExpTile}(\mathcal{D})$, the \textit{Exponential Tiling-by-$\mathcal{D}$ Problem}, is:

**Input:**

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

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

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

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

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

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

- We’ll need a set \( D \) (small) on which to build a structure \( \mathcal{B} = (D; \Gamma) \).
- We’ll encode a copy of \( [2^m] \times [2^m] \) in \( D^m \) and a copy of \( \Delta \) in \( D \).
- We’ll identify auxiliary parameters \( c_i \in D^m \) (\( 1 \leq i \leq n \)) and \( \top \in D \).
- We’ll define \( \Gamma \) so that no polymorphism \( h : \mathcal{B}^m \to \mathcal{B} \) can send \( \mathbf{c} = (c_1, \ldots, c_n) \) to \( \mathbf{t} = (\top, \ldots, \top) \), unless
  - \( h \) sends \( [2^m] \times [2^m] \) to \( \Delta \), and
  - \( h \upharpoonright_{[2^m] \times [2^m]} \) is a \( 2^m \times 2^m \) tiling by \( \Delta \) satisfying initial condition \( \mathbf{w} \).

Constraints:
- The number of relations in \( \Gamma \) must be \( O(m^c) \);
- The arity of each relation in \( \Gamma \) must all be \( O(\log m) \).
- The number, \( n \), of parameters \( c_i \) must be \( O(\log m) \).
Recap: given input $w \in \Delta^m$ to $\text{EXP TILE}(\mathcal{D})$, we’ll build $\Delta^{2m \times 2m}$

- We’ll examine $H(c) = \{ h(c) : h \in \text{Hom}_{\mathcal{B}^m, \mathcal{B}} \} \subseteq D^n$
  and its subset $P(c) = \{ p_i(c) : i < m \}$.

- We’ll achieve: $\forall h : \mathcal{B}^m \rightarrow \mathcal{B}, \ h(c) = t \iff h$ encodes a $[2^m \times 2^m]$ tiling by $\mathcal{D}$ satisfying i.c. $w$. (Conversely, $\exists$ a tiling $\Rightarrow \exists$ such $h$.)

- Hence: $t \in H(c) \iff$ there exists a $2^m \times 2^m$ tiling by $\mathcal{D}$ satisfying the input initial condition $w$. 
Moreover, we’ll find that there is an easily constructed $n$-ary relation $R$ (not depending on $w$), satisfying the following:

- $t \not\in R$.
- $R \cup \{t\}$ is (easily) pp-definable from $\Gamma$.
- $P(c) \subseteq R \subseteq H(c) \subseteq R \cup \{t\}$
- Hence

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

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

1. Construct $D$.

   
   \[
   D =
   \begin{array}{cccc}
   00 & 01 & 10 & 11 \\
   \top & \bot & \Delta & \infty \\
   a & b \\
   0 & 1
   \end{array}
   \]

2. Encode $\Delta$ 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:

   \[
   \begin{align*}
   53 &= 10101100 \text{ (least significant bit at left)} \\
   188 &= 00111101 \\
   (53, 188) &= (10, 00, 11, 01, 11, 11, 00, 01) \in D^8.
   \end{align*}
   \]
4. Define auxiliary parameters $c_i \in D^m$.

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

  \[
  c_0 = (0, 1, 0, 1, 0, 1, 0, 1) \\
  c_1 = (0, 0, 1, 1, 0, 0, 1, 1) \\
  c_2 = (0, 0, 0, 0, 1, 1, 1, 1) \\
  c_3 = (b, b, a, b, a, a, a, b).
  \]

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

Next: add structure $(\Gamma)$ to impose “tiling requirement” on $h$. 
5. Enforcing the horizontal adjacency constraints.

Example \((m = 8)\): if \(h : B^8 \to B\) and

\[
\begin{align*}
(151, 54) & \equiv (10, 11, 11, 00, 11, 01, 00, 01) & x \\
(152, 54) & \equiv (00, 01, 01, 10, 11, 01, 00, 01) & y \\
c_0 & = (0, 1, 0, 1, 0, 1, 0, 1) & h \\
c_1 & = (0, 0, 1, 1, 0, 0, 1, 1) & \top \\
c_2 & = (0, 0, 0, 0, 1, 1, 1, 1) & \top \\
c_3 & = (b, b, a, b, a, a, a, b) & \top
\end{align*}
\]

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:

\[ H_4 = \{(1y, 0y, c_0(j), c_1(j), c_2(j), c_3(j)) : y \in \{0, 1\}, 1 \leq j < 4\} \]
\[ \cup \{(0y, 1y, 1, 1, 0, b) : y \in \{0, 1\}\} \]
\[ \cup \{(xy, xy, c_0(j), c_1(j), c_2(j), c_3(j)) : x, y \in \{0, 1\}, 4 < j \leq 8\} \]
\[ \cup \{(x, y, T, T, T, T) : x, y \in \Delta \text{ and } (x, y) \in H\} \]
\[ \cup \{(x, y, b_0, b_1, b_2, b_3) \in \Delta^2 \times \{\bot\}^4 : \bot \in \{b_0, b_2, b_2, b_3\}\} \]
\[ \cup \{(\infty, \infty, \infty, \infty, \infty, \infty)\}. \]
Similarly, for each $i$ we build a relation $H_i$ to recognize carries at coordinate $i$.

Do the same thing for vertical adjacencies.

Ultimately, $\Gamma$ 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 $I^w_1, \ldots, I^w_m$ of arity $\log_2 m + 2$ for the initial condition $w$.

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

$$R = P(c) \cup \{\top, \bot\}^n \setminus \{(\top, \top, \ldots, \top)\} \cup \{\langle \infty, \infty, \ldots, \infty \rangle\}.$$
Open problems

In my construction of $(D, \Gamma, R)$, $D$ is independent of the input $w$.

Hence there exists $k > 2$ such that $\exists$-InvSat$(k)$ (i.e., restricted $k$-element domains) is co-$\text{NEXP}$$\text{TIME}$-complete.

**Question 1.** Can $k$ be reduced to 3?

**Question 2.** Does there exist a fixed $k$ and fixed $\Gamma$ such that $\exists$-InvSat$(k, \Gamma)$ is co-$\text{NEXP}$$\text{TIME}$-complete?

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

Thank you!

Ross Willard (Waterloo)