Lower Bounds for Formula Size

Philipp Weis

Department of Computer Science
University of Massachusetts Amherst, USA

pweis@cs.umass.edu
http://www.cs.umass.edu/~pweis/

Logic and Algorithms, University of Edinburgh, July 2008
Outline

1. Motivation: Succinctness and Complexity
2. Expressibility with Bounded Number of Variables
3. Lower Bound Techniques
Succinctness

- Succinctness is a model-theoretic concept orthogonal to expressiveness.

**Definition**

The *succinctness* of a logic $L_1$ w.r.t. (in) a logic $L_2$ is $F$ if there is a function $f \in F$ such that for every sentence $\varphi \in L_1$ there is a sentence $\psi \in L_2$ with $\psi \equiv \varphi$ and $|\psi| \leq f(|\varphi|)$.

- The succinctness of $CTL^+$ w.r.t. $CTL$ is not $o(n!)$ (Adler and Immerman, 2003), it is (exactly) $O(n!)$.

- Alternatively:

$$\text{CTL}^+\text{-SIZE}[n] \subsetneq \text{CTL\text{-SIZE}}[o(n!)]$$
$$\text{CTL}^+\text{-SIZE}[n] \subseteq \text{CTL\text{-SIZE}}[O(n!)]$$
Computational Complexity

- Fundamental units in computational complexity: sequential time, parallel time, hardware (space, number of processors)
- Trade-off between parallel time and number of processors is equivalent to trade-off between formula size and number of variables.

**Definition**

Let $\text{FOQB}[t(n)]$ be the set of all formulas of the form $[QB]^{t(n)} M_0$, where $QB = Q_1 x_1 . M_1 \ldots Q_k x_k . M_k$ and $M_0, \ldots, M_k$ are quantifier-free.

**Theorem (Immerman)**

For all $k > 1$, $\text{FOQB}^k[\star] = \text{DSpace}[n^{k-1}]$. 

Philipp Weis  Lower Bounds for Formula Size
Repeated Quantifier Blocks

\[ \text{FOQB}[t(n)] : [ Q_1 x_1 . M_1 \ldots Q_k x_k . M_k ]^{t(n)} M_0 \]

\[
\begin{align*}
\text{FOQB}[O(1)] & \quad = \quad \text{AC}^0 \\
\cap I \quad \cap I \\
\text{FOQB}[O(\frac{\log n}{\log \log n})] & \quad \supseteq \quad \text{NC}^1 \quad \subseteq \quad L \\
\cap I \quad \cap I \\
\text{FOQB}[O(\log n)] & \quad = \quad \text{AC}^1 \quad \supseteq \quad s\text{AC}^1 \\
\cap I \quad \cap I \\
\text{FOQB}[n^{O(1)}] & \quad = \quad \text{P} \\
\cap I \quad \cap I \\
\text{FOQB}[2^{n^{O(1)}}] & \quad = \quad \text{PSPACE} = \text{SOQB}[n^{O(1)}] \\
\end{align*}
\]

- First two lines require BIT.
- Number of quantifier block repetitions is parallel time.
- Number of variables corresponds to number of processors / space.
Our Goals

- Develop more sophisticated tools for succinctness arguments.
- Settle some open questions about the succinctness of the finite-variable fragments of first-order logic.
- Gain insights into the trade-off between the number of variables and formula size.
- Better understanding of the power of repeated quantifier block classes.
1. Motivation: Succinctness and Complexity

2. Expressibility with Bounded Number of Variables

3. Lower Bound Techniques
Some Basic Properties

Let $\text{LO}_n$ be the logical structure with universe $\{1, \ldots, n\}$, a linear order relation $\leq$, and constants min, max. Depending upon context, $\text{LO}_n$ may also interpret a binary successor relation Suc.

\[
\text{LENGTH}_n = \{\text{LO}_i \mid i \leq n\}
\]
\[
\text{EVEN–LENGTH}_n = \{\text{LO}_i \mid i \leq n \text{ and } i \text{ is even}\}
\]

We identify a binary string $w \in \{0, 1\}^n$ with a logical structure $W_w$ that extends $\text{LO}_n$ with a predicate $P$ such that $i \in P^w$ iff $w_i = 1$.

\[
\text{PARITY}_n = \{W_w \mid w \in \{0, 1\}^n \text{ and } \#_1(w) \text{ is even}\}
\]
Formulas for $\text{LENGTH}_n$

\[ \text{LENGTH}_n = \{ \text{LO}_i \mid i \leq n \} \]

\[ \text{length}_{\geq n}^2(x) = \exists y. (y > x) \text{length}_{\geq n-1}^2(y) \]

$\text{LENGTH}_n \in \text{FOQB}^2[n] \subseteq \text{FOSIZE}^2[n]$}

\[ \text{len}_{\geq n}^3(x, y) = \exists z \left( \text{length}_{\geq \lfloor n/2 \rfloor}^3(x, z) \land \text{length}_{\geq \lfloor n/2 \rfloor}^3(z, y) \right) \]

\[ \text{length}_{\geq n}^4(x, y) = \exists z \forall w. (w = x \lor w = y) \text{length}_{\geq n/2}^4(w, z) \]

$\text{LENGTH}_n \in \text{FOQB}^4[\log n] \subseteq \text{FOSIZE}^4[\log n]$
Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

<table>
<thead>
<tr>
<th>$k$</th>
<th>LENGTH</th>
<th>EVEN–LENGTH</th>
<th>PARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$O(n)$</td>
<td>$\Omega(n^2)\dagger$</td>
<td>$O(2^{n+\log n})\dagger$</td>
</tr>
<tr>
<td></td>
<td>$\Theta(n)$</td>
<td>$O(n^2)$</td>
<td>$O(n)^B$</td>
</tr>
<tr>
<td>$2S$</td>
<td>$\Omega(n)$</td>
<td>$\Theta(n)$</td>
<td>$O(n)^B$</td>
</tr>
<tr>
<td>3</td>
<td>$\Omega(n)\dagger$</td>
<td>$\Omega(\sqrt{n})$</td>
<td>$\Omega(n)\dagger$</td>
</tr>
<tr>
<td></td>
<td>$\Omega(\sqrt{n})$</td>
<td>$\Omega(\sqrt{n})$</td>
<td>$\Omega(n)^{\dagger}$</td>
</tr>
<tr>
<td>4</td>
<td>$\Theta(\log n)$</td>
<td>$\Theta(\log n)$</td>
<td>$\Theta(\log n)^B$</td>
</tr>
</tbody>
</table>

Philipp Weis Lower Bounds for Formula Size
Formulas for $\text{EVEN–LENGTH}_n$

$$\text{EVEN–LENGTH}_n = \{LO_i \mid i \leq n \text{ and } i \text{ is even}\}$$

$$\text{even}^2_n = \bigvee_{i \leq n \text{ and } i \text{ even}} \text{length}^2_{\leq i}$$

$\text{EVEN–LENGTH}_n \in \text{FOSIZE}^2[n^2]$?

$$\text{even}^r_n = \text{length}^2_{\geq 2} \text{ length}^2_{< 3} \text{ length}^2_{\geq 4} \ldots$$

$$= \exists x.(x > \text{min}) \forall x.(x > \text{min}) \forall y.(y > x) \ldots$$

$\text{EVEN–LENGTH}_n \in \text{FOQB}^2[n^2]$?

$$\text{even}_n^{2S}(x) = \exists y.(\text{Suc}(x, y)) \exists x.(\text{Suc}(y, x)) (x = \text{max} \bigvee \text{even}^{2S}_{n-2}(x))$$

$\text{EVEN–LENGTH}_n \in \text{FOQB}^{2S}[n]$?

$$\text{elen}_n^4(x, y) = \exists z \exists w.(w = x \lor w = y) \text{even}^4_{\lfloor n/2 \rfloor}(w, z)$$
Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

<table>
<thead>
<tr>
<th></th>
<th>LENGTH</th>
<th>EVEN–LENGTH</th>
<th>PARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$O(n)$</td>
<td>$\Omega(n^2) \dagger$</td>
<td>$O(2^{n+\log n}) \dagger$</td>
</tr>
<tr>
<td>2S</td>
<td>$\Omega(n)$</td>
<td>$\Theta(n)$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\Omega(n) \dagger$, $\Omega(\sqrt{n})$</td>
<td>$\Omega(n) \dagger$, $\Omega(\sqrt{n})$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$\Theta(\log n)$</td>
<td>$\Theta(\log n)$</td>
<td>$\Theta(\log n)^B$, $\Omega(n) \dagger$</td>
</tr>
</tbody>
</table>
Formulas for PARITY

\[ \text{PARITY}_n = \{ W_w \mid w \in \{0, 1\}^n \text{ and } \#_1(w) \text{ is even} \} \]

\[ \text{PARITY}_n \in \text{FOSIZE}^2[2^{n+\log n}] \]

\[ \text{PARITY}_n \notin \text{FOQB}^2[*] \quad \text{(conjectured)} \]

\[ \text{parity}^{2B}_n(x, b) = \exists b'.((b' = b \land \neg P(x)) \lor (b' \neq b \land P(x))) \]

\[ \exists x'.(x < x') \quad \text{parity}^{2B}_{n-1}(x', b') \]

\[ \text{PARITY}_n \in \text{FOQB}^{2B}[n] \]

\[ \text{parity}^{4B}_n(x, y, b) = \exists z \exists b' \exists b''.((b = 0 \land b' = b'') \lor (b = 1 \land b' \neq b'')) \]

\[ \forall w \forall b.((w = x \land b = b') \lor (w = y \land b = b'')) \]

\[ \text{parity}^{4B}_{n/2}(w, z, b) \]

\[ \text{PARITY}_n \in \text{FOQB}^{4B}[\log n] \]
Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>EVEN–LENGTH</th>
<th>PARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$O(n)$</td>
<td>$\Omega(n^2)^\dagger$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td>2S</td>
<td>$\Omega(n)$</td>
<td>$\Theta(n)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\Omega(n)^\dagger$</td>
<td>$\Omega(n)^\dagger$</td>
</tr>
<tr>
<td></td>
<td>$\Omega(\sqrt{n})$</td>
<td>$\Omega(\sqrt{n})$</td>
</tr>
<tr>
<td>4</td>
<td>$\Theta(\log n)$</td>
<td>$\Theta(\log n)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Open Problems

- Improve succinctness lower bound for $\text{FO}^2$ vs. $\text{FO}^3$ from $\Omega(\sqrt{n})$ to $\Omega(n)$.

- Succinctness hierarchy: $\text{FO}^k$ vs. $\text{FO}^{k+1}$ (compare to expressibility hierarchy result of Rossman)

- Succintness results for words, trees, graphs? (results from (Weis and Immerman, 2007) useful for succinctness on words?)
Outline

1 Motivation: Succinctness and Complexity

2 Expressibility with Bounded Number of Variables

3 Lower Bound Techniques
The Adler-Immerman Game

Adler and Immerman (2003)

- Standard Ehrenfeucht-Fraïssé games are used for lower bounds on quantifier depth. The game is played on a pair of structures, and corresponds to a first-order formula separating the two structures.
- For bounds on size, we need to consider all possible moves of Delilah.
- Adler-Immerman game is played on two sets of structures. The game tree corresponds to a formula that separates all possible pairs of structures.
The Adler-Immerman Game

- During the course of the game, a labeled tree is constructed that corresponds to a first-order formula.
- Initially, the tree consists only of the root node, labeled \((A, B)\).
- Samson can close a leaf node if the corresponding sets of structures disagree on an atomic formula, or he can play an OR, NOT or existential move on an open leaf node.
- Samson wins the game if he can close all leaves.

\[
\exists x (A, B) \lor (A', B') \\
(\forall (A'_1, B') \lor (A'_2, B')
\]
Rules of the Adler-Immerman Game

Samson plays one of the following three moves on an open leaf.

- In an OR move, he picks $A'_1, A'_2 \subseteq A'$ such that $A' = A'_1 \cup A'_2$.
- In an existential move with variable $x$, he places $x$ on an element $f(A)$ of $A$ for every $A \in A$, and adds a child node labeled $\left( \left\{ A \frac{f(A)}{x} \mid A \in A \right\}, \left\{ B \frac{i}{x} \mid B \in B, i \in |B| \right\} \right)$.
Observations about the Adler-Immerman Game

- AND and universal moves can be added in the obvious way for convenience.

- Extensions for temporal logic, transitive closure operators, etc. are relatively straightforward.

- With only boolean moves, this game is exactly the communication complexity game introduced by Karchmer and Wigderson.
More uniform version of the Adler-Immerman game that corresponds to iterated quantifier block formulas.

- All $M_i$ are quantifier-free.
- Here depth and size coincide.
Lower Bounds on Size

How can we prove size lower bounds with the Adler-Immerman game?

- Similar to communication complexity, but in those games we only have boolean moves available.
- With extended moves of the Adler-Immerman game (quantifiers, transitive closure, etc.), the structures change throughout the game.

Available techniques:

- **Incompatible Pairs**: Two pairs of structures are incompatible if they need to appear separately on at least one leaf. The number of incompatible pairs yields a lower bound on the number of leaves. (Adler and Immerman, 2003)
- **Weight Function**: Define a function $w : \mathcal{P}(A) \times \mathcal{P}(B) \rightarrow \mathbb{R}$, argue bottom-up about the maximum increase of $w$ in each move, and bound $w(A, B)$. (Grohe and Schweikardt, 2005)
Grohe and Schweikardt (2005) define a weight function for the 3-variable game on $\text{LENGTH}_n$ based on separators. A separator for two sets of structures $A, B$ is a function

$$\delta : \mathcal{P}_2(\{\text{min}, x, y, z, \text{max}\}) \rightarrow \mathbb{N}$$

such that for all $A \in A, B \in B$, there are $u, v \in \{\text{min}, x, y, z, \text{max}\}$ with

- $\text{ord}(u^A, v^A) \neq \text{ord}(u^B, v^B)$ and $\delta(\{u, v\}) > 0$

or

- $\text{dist}(u^A, v^A) \neq \text{dist}(u^B, v^B)$ and

$$\delta(\{u, v\}) \geq \min\{\text{dist}(u^A, v^A), \text{dist}(u^B, v^B)\}$$

This is the distance required to walk to distinguish the structures.
The weight \( w(A, B) \) is defined as a function of a minimal separator for \((A, B)\), involving summing up several distances and taking square roots.

With this particular weight function, able to prove that \( \text{LENGTH}_n \notin \text{FOSIZE}^3[o(\sqrt{n})] \).

**Conjecture**

\[
\text{LENGTH}_n \notin \text{FOSIZE}^3[o(n)]
\]

**Proposition**

\[
\text{LENGTH}_n \in \text{FOSIZE}^3[O(n)].
\]
Conclusion

- Formula size game for lower bounds.
- Develop new techniques / refine existing techniques for succinctness arguments.
- Gain insight into trade-off between formula size and number of variables, first-order characterizations of complexity classes.

Main Open Problems:
- Succinctness hierarchy for $\text{FO}^k$?
- Succinctness of $\text{FO}^k$ on words, trees, graphs, ...
- First step (?): Show that $\text{EVEN–LENGTH}_n \notin \text{FOQB}^2[o(n^2)]$.
- Improve (Grohe and Schweikardt, 2005) to show that $\text{LENGTH}_n \notin \text{FOSIZE}^3[o(n)]$. 

Philipp Weis  
Lower Bounds for Formula Size
