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:

$$\text{CTL}^+\text{-SIZE}[n] \not\subseteq \text{CTL-SIZE}[o(n!)]$$
$$\text{CTL}^+\text{-SIZE}[n] \subseteq \text{CTL-SIZE}[O(n!)]$$

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 $\mathrm{FOQB}[t(n)]$ be the set of all formulas of the form $[QB]^{t(n)}M_0$, where $QB = Q_1x_1.M_1 \ldots Q_kx_k.M_k$ and $M_0, \ldots, M_k$ are quantifier-free.

Theorem (Immerman)

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

Repeated Quantifier Blocks

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

$$\begin{array}{ccccccc}
\mathrm{FOQB}[O(1)] & = & \mathrm{AC}^0 & & \\
\cap| & & \cap| & & \\
\mathrm{FOQB}[O(\frac{\log n}{\log\log n})] & \supseteq & \mathrm{NC}^1 & \subseteq & \mathrm{L} \\
\cap| & & \cap| & & \mathrm{NL} \\
\mathrm{FOQB}[O(\log n)] & = & \mathrm{AC}^1 & \supseteq & s\mathrm{AC}^1 \\
\cap| & & \cap| & & \\
\mathrm{FOQB}[n^{O(1)}] & = & \mathrm{P} & & \\
\cap| & & \cap| & & \\
\mathrm{FOQB}[2^{n^{O(1)}}] & = & \mathrm{PSPACE} & = & \mathrm{SOQB}[n^{O(1)}]
\end{array}$$

- 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.

Outline

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

Some Basic Properties

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

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

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

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

Formulas for LENGTH_n

$$\mathrm{LENGTH}_n = \{\mathrm{LO}_i \mid i \leq n\}$$

$$\mathrm{length}^2_{\geq n}(x) = \exists y.(y > x)\,\mathrm{length}^2_{\geq n-1}(y)$$

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

$$len^3_{\geq n}(x, y) = \exists z\,(\mathrm{length}^3_{\geq \lceil n/2 \rceil}(x, z) \wedge \mathrm{length}^3_{\geq \lfloor n/2 \rfloor}(z, y))$$

$$\mathrm{length}^4_{\geq n}(x, y) = \exists z\,\forall w.(w = x \vee w = y)\,\mathrm{length}^4_{\geq n/2}(w, z)$$

$$\mathrm{LENGTH}_n \in \mathrm{FOQB}^4[\log n] \subseteq \mathrm{FOSIZE}^4[\log n]$$

Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

	LENGTH		EVEN–LENGTH		PARITY	
2	$O(n)$		$\Omega(n^2)^\dagger$	$O(n^2)$	$O(2^{n+\log n})^\dagger$	$O(n)^B$
	$\vert$					$\vert$
2S	$\Omega(n)$		$\Theta(n)$			$\vert$
	$\vert$		$\vert$			$\vert$
3	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$		$\vert$
						$\vert$
4	$\Theta(\log n)$		$\Theta(\log n)$		$\Theta(\log n)^B$	$\Omega(n)^\dagger$

Formulas for EVEN–LENGTH$_n$

$$\text{EVEN–LENGTH}_n = \{\text{LO}_i \mid i \le n \text{ and } i \text{ is even}\}$$

$$\text{even}^2_n = \bigvee\nolimits_{i \le n \text{ and } i \text{ even}} \text{length}^2_{=i}$$

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

$$\begin{aligned} \text{even}'^2_n &= \text{length}^2_{\ge 2}\ \text{length}^2_{<3}\ \text{length}^2_{\ge 4}\ \dots \\ &= \exists x.(x > \min)\,\forall x.(x > \min)\,\forall y.(y > x)\ \dots \end{aligned}$$

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

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

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

$$\mathit{elen}^4_n(x,y) = \exists z\,\exists w.(w = x \vee w = y)\,\text{even}^4_{\lceil n/2 \rceil}(w,z)$$

Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

	LENGTH		EVEN–LENGTH		PARITY	
2	$O(n)$		$\Omega(n^2)^\dagger$	$O(n^2)$	$O(2^{n+\log n})^\dagger$	$O(n)^B$
	│					│
2S	$\Omega(n)$		$\Theta(n)$			│
	│		│			│
3	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$		│
						│
4	$\Theta(\log n)$		$\Theta(\log n)$		$\Theta(\log n)^B$	$\Omega(n)^\dagger$

Formulas for PARITY

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

$$\mathrm{PARITY}_n \in \mathrm{FOSIZE}^2[2^{n+\log n}]$$
$$\mathrm{PARITY}_n \notin \mathrm{FOQB}^2[\star] \qquad \text{(conjectured)}$$

$$\mathrm{parity}_n^{2B}(x,b) = \exists b'.((b' = b \wedge \neg P(x)) \vee (b' \neq b \wedge P(x)))$$
$$\exists x'.(x < x')\ \mathrm{parity}_{n-1}^{2B}(x', b')$$
$$\mathrm{PARITY}_n \in \mathrm{FOQB}^{2B}[n]$$

$$\mathrm{parity}_n^{4B}(x,y,b) = \exists z\, \exists b'\, \exists b''.((b = 0 \wedge b' = b'') \vee (b = 1 \wedge b' \neq b''))$$
$$\forall w\, \forall b.((w = x \wedge b = b') \vee (w = y \wedge b = b''))$$
$$\mathrm{parity}_{n/2}^{4B}(w,z,b)$$
$$\mathrm{PARITY}_n \in \mathrm{FOQB}^{4B}[\log n]$$

Bounds on Formula Size

Conjectured or suspected bounds are marked with †.

	LENGTH		EVEN–LENGTH		PARITY	
2	$O(n)$		$\Omega(n^2)^\dagger$	$O(n^2)$	$O(2^{n+\log n})^\dagger$	$O(n)^B$
	$\vert$					$\vert$
2S	$\Omega(n)$		$\Theta(n)$			$\vert$
	$\vert$		$\vert$			$\vert$
3	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$	$\Omega(n)^\dagger$	$\Omega(\sqrt{n})$		$\vert$
						$\vert$
4	$\Theta(\log n)$		$\Theta(\log n)$		$\Theta(\log n)^B$	$\Omega(n)^\dagger$

Open Problems

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

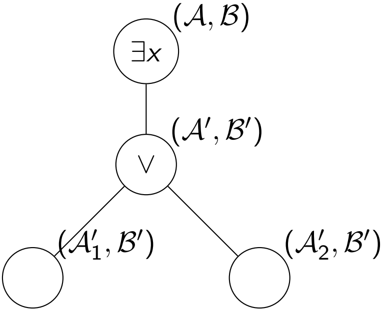
The Adler-Immerman Game

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.

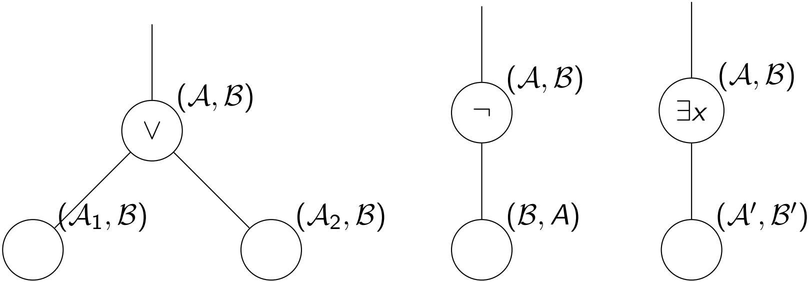
The Adler-Immerman Game

- 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 $(\mathcal{A}, \mathcal{B})$.
- 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.



Rules of the Adler-Immerman Game

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

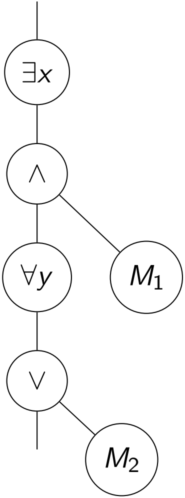


- In an OR move, he picks $\mathcal{A}'_1, \mathcal{A}'_2 \subseteq \mathcal{A}'$ such that $\mathcal{A}' = \mathcal{A}'_1 \cup \mathcal{A}'_2$.
- In an existential move with variable x, he places x on an element $f(A)$ of A for every $A \in \mathcal{A}$, and adds a child node labeled $\left(\left\{A\,\frac{f(A)}{x}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

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.

Stubby Line Game



- 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 $w : \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B}) \rightarrow \mathbb{R}$, argue bottom-up about the maximum increase of w in each move, and bound $w(\mathcal{A}, \mathcal{B})$. (Grohe and Schweikardt, 2005)

Grohe-Schweikardt Separator Weight Technique I

Grohe and Schweikardt (2005) define a weight function for the 3-variable game on LENGTH_n based on *separators*. A separator for two sets of structures $\mathcal{A}, \mathcal{B}$ is a function

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

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

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

or

- $\mathrm{dist}(u^A, v^A) \neq \mathrm{dist}(u^B, v^B)$ and $\delta(\{u, v\}) \geq \min\{\mathrm{dist}(u^A, v^A), \mathrm{dist}(u^B, v^B)\}$

This is the distance required to walk to distinguish the structures.

Grohe-Schweikardt Separator Weight Technique II

- The weight $w(\mathcal{A}, \mathcal{B})$ is defined as a function of a minimal separator for $(\mathcal{A}, \mathcal{B})$, involving summing up several distances and taking square roots.
- With this particular weight function, able to prove that $\mathrm{LENGTH}_n \notin \mathrm{FOSIZE}^3[o(\sqrt{n})]$.

Conjecture

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

Proposition

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

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

References

Micah Adler and Neil Immerman. An *n*! lower bound on formula size. *ACM Transactions on Computational Logic*, 4(3):296–314, 2003.

Martin Grohe and Nicole Schweikardt. The succinctness of first-order logic on linear orders. *Logical Methods in Computer Science*, 1(1:6), 2005.

Philipp Weis and Neil Immerman. Structure theorem and strict alternation hierarchy for FO^2 on words. In *Computer Science Logic*, 2007.