Towards affluent proof theory of LOGCFL and related topics

(Proof theoretical study of the structure of logic and computation)

Kuroda, Satoru

数理解析研究所講究録 1635: 77-85

2009-04

URL: http://hdl.handle.net/2433/140461

Type: Departmental Bulletin Paper

Kyoto University
Towards affluent proof theory of LOGCFL and related topics

Satoru Kuroda (黒田覚)
群馬県立女子大学文学部

1 Introduction

LOGCFL is the class of sets which are LOGSPACE reducible to context free languages. This class lies between NL and $AC^1$ and has several alternative characterizations based on different computational concepts.

Such concepts include Boolean circuits (semi-unbounded fan-in logarithmic depth circuits), algebraic structure (word problem for finite groupoids), database theory (acyclic conjunctive queries) and graph theory (acyclic graph embedding problem).

It had been open whether there exists a theory of weak arithmetic which capture this class, that is, a theory whose provably total functions precisely correspond to functions bitwise computable in LOGCFL.

The author gave the first such theory $V-Q^{SAC}(\Sigma^B_0)$ based on the circuit based characterization of LOGCFL, while it is still open to construct similar theories based on other characterizations of the class.

Another aspect of complexity theory is the complexity of propositional proofs. Many natural complexity classes are known to correspond to natural propositional proof systems. So it is likely that LOGCFL also has a natural counterpart in propositional proof systems.

One such system is the Frege system enhanced to manipulate $SAC^1$ circuits in place of Boolean formulae. It is known that Frege with polynomial size circuits is equivalent to polynomial size extended Frege system, which corresponds to P. Similarly, we can prove that $SAC^1$-Frege corresponds to the theory $V-Q^{SAC}(\Sigma^B_0)$.

In connection with propositional complexity, the above characterizations of LOGCFL might give lots of information. For example, the concept needed to define word problem for groupoids can be translated to a family of propositional formulae. So it is an interesting problem which propositional proof system has or does not have polynomial size proof of such a concept.

This note aims to give a guideline for the research of LOGCFL in terms of proof theory as stated above. We give several results obtained so far without mentioning proofs. We will further give some informal ideas for the future research.
2 The class LOGCFL

In this section we overview fundamental properties of the class LOGCFL. We will assume that readers are familiar with basic concepts of computational complexity and formal language theory.

Definition 1 LOGCFL is the class of sets which are logspace reducible to some context free language.

An alternative characterization is known for this class based on Boolean circuits. A semi-unbounded fan-in circuit is a Boolean circuit with basis \{\wedge, \vee, \neg\} where \wedge gates receive at most two inputs and \neg gates appear only at the input level.

Definition 2 SAC\textsuperscript{1} is the class of sets which are decidable by some semi-unbounded fan-in circuit family with \(O(\log n)\) depth and \(n^{O(1)}\) size.

Theorem 1 (Venkateswaran [9]) A set is in LOGCFL if and only if it is decidable by an \(U_{E}^{*}\)-uniform SAC\textsuperscript{1} circuit family.

From this characterization we obtain the following inclusion relation neither of which is known to be proper.

Corollary 1 NL \(\subseteq\) LOGCFL \(\subseteq\) AC\textsuperscript{1}.

Several complete problems are known for LOGCFL.

A groupoid is a set with a (not necessarily be associative) binary operation. For a groupoid \(G\), work problem for \(G\) is the problem of given a finite word \(w = a_1a_2\cdots a_i\) over \(G\), decide whether there exists a bracketing of \(w\) for which it is multiplied out to an element in given \(S \subseteq G\).

Proposition 1 The word problem for finite groupoids is complete for LOGCFL under constant depth reductions.

The following complete problem is also noteworthy.

Theorem 2 (Gottlob et.al. [5]) The acyclic conjunctive query problem is complete for LOGCFL under constant depth reductions.

3 Theories for LOGCFL

Bounded arithmetic gives a logical framework for investigating complexity classes. There are several versions of such theories and two-sort systems gains much popularity recently. In particular, Cook, Kolokolova and Nguyen developed two sort theories which corresponds to complexity classes below \(P\) (see for example [6]).

In this section we give two theories of two sort which correspond to LOGCFL. Let us first give some definitions.

The language \(L_2\) of two sort systems consists of the followings:
- number variables: $x, y, z, \ldots$,
- string variables: $X, Y, Z, \ldots$,
- constant: 0,
- function symbols: $s(x) = x + 1, x + y, x \cdot y, |X|$, 
- predicate symbols: $x < y, x = y, x \in X$.

Note that the intended model of $L_2$ is $(\mathbb{N}, \Sigma^*)$ where $\Sigma = \{0, 1\}$.

Quantifiers of the form $(\forall x)$ and $(\exists x)$ are called number quantifiers and $(\forall X)$ and $(\exists X)$ are called string quantifiers. We define bounded string quantifiers as follows:

$$(\forall X < t)\varphi \equiv (\forall X)(|X| < t \rightarrow \varphi)$$

$$(\exists X < t)\varphi \equiv (\exists X)(|X| < t \wedge \varphi)$$

**Definition 3** $\Sigma_0^B$ is the set of $L_2$ formulae in which all quantifiers are bounded number quantifiers. $\Sigma_1^B$ is the set of $L_2$ formulae whose quantifiers are either bounded number quantifiers, positive appearances of bounded existential string quantifiers or negative appearances of bounded universal string quantifiers.

We are interested in the computational complexity of definable functions of a given $L_2$ theory.

**Definition 4** A number function is a function of the form $f : \mathbb{N}^k \times (\Sigma^*)^l \rightarrow \mathbb{N}$. A string function is a function of the form $f : \mathbb{N}^k \times (\Sigma^*)^l \rightarrow \Sigma^*$.

**Definition 5** Let $T$ be a $L_2$ theory and $\Phi$ be a class of $L_2$ formulae. A string function $F$ is $\Phi$ definable in $T$ if there exists $\varphi \in \Phi$ such that

$$T \vdash (\forall \vec{x})(\forall X)(\exists! Y)\varphi(\vec{x}, X, Y)$$

and $(\mathbb{N}, \Sigma^*) \models (\forall \vec{x})(\forall X)\varphi(\vec{x}, X, F(\vec{x}, X))$.

### 3.1 A theory for SAC$^1$

In [8], the author defined a theory whose provably total functions correspond to SAC$^1$. Here we briefly overview this system.

Given an input to a Boolean circuit, we call a path an alternating path which gives a witness that the circuit accepts the input. Strictly, alternating paths are defined by induction on the depth of circuits.

For an $L_2$-formula $\varphi$ we define the predicate symbol $Q_{\varphi}^{SAC}$ by

$$Q_{\varphi}^{SAC}(n, t) \iff \text{there exists an alternating path with the root } t$$

$$\text{in the circuit defined by } \varphi \text{ with the input length } n.$$ 

and let $V \cdot Q^{SAC}(\Sigma^B_0)$ be the system with bit comprehension axioms for $\{Q_{\varphi}^{SAC}(n, t) : \varphi \in \Sigma^B_0\}$. Then we have
\textbf{Theorem 3 (Kuroda [8])} \( V \cdot Q^{SAC}(\Sigma_0^B) \) captures LOGCFL.

The essential part of the proof of Theorem 3 is to show that \( Q^{SAC}(\Sigma_0^B) \) is strongly closed and, in particular, it is not obvious that \( SAC^1 \) is closed under complementation. Borodin et.al. [2] used the inductive counting to show this and we can formalize their argument inside \( V \cdot Q^{SAC}(\Sigma_0^B) \).

\section{A theory for permutation word problem}

Our new theory for LOGCFL is based on a version of word problem for groupoids. The idea of defining this version of word problem is that we fix trees denoting the bracketing of a given word \( w \) and attach some permutation of \( w' \) to leaves of the tree.

For a node \( v \) of a tree let \( \text{label}(v) \) denote the label assigned to \( v \). We also denote the left and right offsprings of \( v \) by \( \text{left}(v) \) and \( \text{right}(v) \). The Permutation Word Problem (PWP) is the following decision problem:

\textbf{Definition 6} Let \( T = T_1, T_2, \ldots \) be a family of rooted binary trees so that \( T_n \) has \( n \) leaves. Let \( G \) be a finite groupoid and \( S \subseteq G \). The permutation word problem for \( (G; S) \), denoted by \( PWP(T; G, S) \), is the set of finite words \( w \) over \( G \) such that there exists a permutation \( w' \) of \( w \) satisfying the following condition:

\begin{itemize}
  \item if we assign elements of \( w' \) to leaves of \( T_{|w|} \) and compute each label of non-leaf node \( v \) by
  \[ \text{label}(v) = \text{label(left}(v)) \circ_G \text{label(right}(v)) \]
  \item then the label of the root of \( T_{|w|} \) belongs to \( S \).
\end{itemize}

\textbf{Definition 7} BPWP is the problem PWP with the restriction that the family \( T \) of binary trees consists of balanced trees.

\textbf{Theorem 4} BPWP is LOGCFL complete under constant depth reductions.

Our system for LOGCFL is based on the problem BPWP. The essential axiom for the system expresses the computation of the word problem. In order to describe it, we need several functions and predicates which are definable in the base theory.

\textbf{Definition 8} \( V^0 \) is the \( \mathcal{L}_2 \) theory with the following axioms:

\begin{itemize}
  \item \( \text{BASIC}_2 \) : a finite set of axioms which define symbols in \( \mathcal{L}_2 \).
  \item \( \Sigma_0^B \cdot \text{COMP} : (\exists X < a)(\forall y < a)(y \in X \leftrightarrow \varphi(y)), \) where \( \varphi \in \Sigma_0^B \).
\end{itemize}

It is known that we can freely add \( \Sigma_1^B \) definable functions and \( \Delta_1^B \) definable predicates in \( V^0 \). Using this fact we introduce several functions and predicates in \( V^0 \).

A two dimensional array \( X \) defines a finite function so that

\[ X[y] = \min\{z : X(y, z)\}. \]
A finite groupoid is coded by a three dimensional array $n \times n \times n$. We will denote the binary operation $x \circ_G y$ by

$$G[x, y] = \min\{z : G(x, y, z)\}$$

The following formula asserts that any given word can be multiplied out to a single element according to a balanced tree:

$$\delta_{GWP}(a, G, W, Y) \iff \forall x < a \,(Y[x + a] = W[x] \land 0 < x \rightarrow (Y[x] = G[Y[2x], Y[2x + 1]])).$$

Next we define functions and predicates which manipulate permutations. First define

$$Perm(F, n) \iff \forall x < n)(\exists y < n)F(x, y) \land \forall x < n)(\forall y < n)(\forall z < n)(F(x, z) \land F(y, z) \rightarrow x = y).$$

$Perm(X, n)$ asserts that a permutation $X$ of $n$ is a bijection $[n] \rightarrow [n]$ and we have

**Proposition 2** $Perm(X, n)$ is $\Delta^B_1$ definable in $V^0$.

Next we define the application of a permutation to words.

$$Trans(F, X, n) = Y \iff \exists F)(\exists Y)(Perm(F, n) \land \delta_{GWP}(n, G, Trans(F, W, n), Y) \land Y[n - 1] = i).$$

**Proposition 3** $Trans(F, X, n)$ is $\Sigma^B_1$ definable in $V^0$.

Now we can define a single axiom which states that for a given groupoid $G$ and a word $W$ on $G$ there exists a sequence which codes all elements of $G$ so that some permutation of $W$ can be multiplied out to it by a balanced tree. BPWPG (for Balanced Permutation Word Problem for Groupoids) is the following axiom:

$$(\forall G)(\forall W)(\exists S)(\forall i < n) \,(S(i) \rightarrow (\exists F)(\exists Y)(Perm(F, n) \land \delta_{GWP}(n, G, Trans(F, W, n), Y) \land Y[n - 1] = i)).$$

**Definition 9** T-PWP is the $L_2$ theory whose axioms are

- all axioms for $V^0$,
- BPWPG.

**Theorem 5** A function is $\Sigma^B_1$ definable in T-PWP if and only if it is bitwise computable in LOGCFL.

**Problem 1** Show that T-PWP is identical to $V-Q^{SAC}(\Sigma^B_0)$. That is, show that for all $L_2$ formulae $\varphi$ T-PWP $\vdash \varphi$ if and only if $V-Q^{SAC}(\Sigma^B_0) \vdash \varphi$. 
4 Propositional proof system for LOGCFL

It is known that many complexity classes have its counterpart in propositional proofs. A natural candidate for LOGCFL is SAC\(^1\) Frege proof system.

**Definition 10** Circuits \(C\) and \(D\) are similar (written as \(C \simeq D\)) if they are unfolded into the same formula.

**Definition 11** A Circuit Frege (CF) proof system is defined as follows: Let \(B\) be a basis of Boolean formulae and \(R\) be a sound and implicationally complete set of Frege rules over \(B\). A CF proof is a sequence \(A_0, \ldots, A_k\) of \(B\)-circuits such that each \(A_i\) is either

- obtained by some \(R\) rule from \(A_{j_1}, \ldots A_{j_l}\) for some \(j_1, \ldots, j_l < i\), or
- \(A_j \simeq A_i\) for some \(j < i\).

**Definition 12** A SAC\(^1\)-Frege proof is a CF proof where circuits in the proof are restricted to SAC\(^1\) circuits.

**Theorem 6** There exists a translation \(\varphi(\overline{X}) \mapsto \|\varphi(\overline{X})\|_m\) of \(\Sigma_0^B\) formulae into propositional formulae such that if \(V-Q^{SAC}(\Sigma_0^B) \vdash (\forall \overline{X})\varphi(\overline{X})\) where \(\varphi \in \Sigma_0^B\) then \(\|\varphi(\overline{X})\|_m\) has polynomial size SAC\(^1\)-Frege proofs.

**Theorem 7** \(V-Q^{SAC}(\Sigma_0^B)\) proves the reflection principle for SAC\(^1\)-Frege.

5 Future research and open problems

In this note we present two theories for LOGCFL based on different characterization. Another interesting feature of this class is acyclic conjunctive queries defined in section 2. So it is plausible to define another system based on this characterization.

**Problem 2** Define a natural axiom \(A\) which expresses the concept of acyclic conjunctive queries so that \(V^0 + A\) captures LOGCFL.

A similar problem can be considered in the context of propositional proofs. Atserias et.al. [1] defined constraint based proof system called CSP-proofs.

**Definition 13** Let \(\sigma\) be a signature for finite models and \(A\) and \(B\) be \(\sigma\)-structures. The homomorphism problem is to decide whether there exists an homomorphism \(A \rightarrow B\).

It is easily seen that the homomorphism problem is equivalent to constraint satisfaction problem, which is in turn equivalent to conjunctive query problem.

Now the objective of CSP-proof is, given two structures \(A\) and \(B\), show that whether there is no homomorphism \(A \rightarrow B\) in the following manner.
Definition 14 A constraint is a pair \((\overline{x}, R^B)\) where \(\overline{x} = x_1, \ldots, x_l\) is a tuple of elements of \(A\) and \(R^B\) is a relation over \(B\) of the arity \(l\). For two relations \(R, S\) we denote their join by \(R \bowtie S\).

A CSP\((B)\)-proof from \(A\) is a finite sequence of constraints such that each constraint is either one of the followings:

1. \((\overline{x}, R^B) : R \in \sigma\) and \(\overline{x} \in R^A\),
2. \((\overline{x} \cup \overline{y}, R \bowtie S) : (\overline{x}, R)\) and \((\overline{y}, S)\) appears previously,
3. \((\overline{x}, \pi_{\overline{x}-\overline{y}}(R)) : (\overline{x}, R)\) appears previously,
4. \((\overline{x}, S) : (\overline{x}, R)\) appears previously and \(R \subseteq S\).

A CSP\((B)\)-refutation of \(A\) is a CSP\((B)\)-proof from \(A\) whose last constraint has an empty relation.

It is easily seen that CSP-refutation is sound and complete. However, this proof system itself cannot be Cook-Reckhow system as it does not satisfy the P-verifiability condition.

In order to overcome this problem, the notion of representation class is introduced:

Definition 15 Let \(B\) be a finite set. A representation class for Boolean valued functions with domain \(B^k\) is the tuple \(\mathcal{R} = (Q, I, S)\) where

- \(Q\) is the set of representations,
- \(I\) is a mapping \(Q \rightarrow \{f : B^k \rightarrow \{0, 1\}\}\) called interpretation,
- \(S\) is a mapping \(Q \rightarrow \mathbb{N}\) called size function.

Taking some computation model as the set of representation \(Q\) we obtain P-verifiable versions of CSP-proof system.

Definition 16 Let \(\pi\) be an order on Boolean variables. A \(\pi\)-OBDD is a binary decision diagram in which all paths have labels which are consistent with \(\pi\).

Definition 17 Let \(\pi\) be an order on variables. Then \(\pi\)-OBDD proof system is the CSP proof system where the representation class for constraints is \(\pi\) Ordered Binary Decision Diagram (OBDD).

Since basic operations for OBDDs (such as the equivalence of two OBDDs etc) are polynomial time computable (see [3]), this proof system is known to be a Cook-Reckhow system.

However Krajíček [7] proved an exponential lower bounds to OBDD-proofs. So we ask the following question:
Problem 3 Is there a representation class for which CSP-proof corresponds to LOGCFL, \( P \) etc?

Apart from the relation to LOGCFL, CSP proof itself is interesting. We close this note by posing a question about CSP proofs.

**Definition 18** A Free BDD (FBDD) is a BDD such that in each path from the root to an end node, each variable is checked at most once.

Given two FBDDs it is not known to be polynomial time computable to check whether they are identical. However, by introducing the notion of type (see [4]), we can check in polynomial time the equivalence of two given FBDDs of the same type. This leads to the definition of FBDD proofs as follows:

**Definition 19** Let \( \tau \) be a type of Free BDDs. A \( \tau \)-FBDD is a FBDD with type \( \tau \). A \( \tau \)-FBDD proof is a CSP-proof whose representation class is \( \tau \)-FBDDs.

As models of computations, FBDD have exponential speed-up over OBDD. So the author conjectures

**Conjecture 1** FBDD-proofs have exponential speed-up over OBDD-proofs.

参考文献


