COMPLEXITY OF FINITELY PRESENTED ALGEBRAS*

Dexter Kozen

TR 76-294

Department of Computer Science
Cornell University
Ithaca, New York 14853

COMPLEXITY OF FINITELY PRESENTED ALGEBRAS*

Dexter Kozen
Department of Computer Science
Cornell University
Ithaca, New York 14853

Abstract

An algebra $\mathcal{A}$ is finitely presented if there is a finite set G of generator symbols, a finite set O of operator symbols, and a finite set Γ of defining relations $x \equiv y$ where x and y are well-formed terms over G and O, such that $\mathcal{A}$ is isomorphic to the free algebra on G and O modulo the congruence induced by Γ.

The uniform word problem, the finiteness problem, the triviality problem (whether $\mathcal{A}$ is the one element algebra), and the subalgebra membership problem (whether a given element of $\mathcal{A}$ is contained in a finitely generated subalgebra of $\mathcal{A}$) for finitely presented algebras are shown to be $\leq^{m}_{\log}$-complete for P. The schema satisfiability problem and schema validity problem are shown to be $\leq^{m}_{\log}$-complete for NP and co-NP, respectively. Finally, the problem of isomorphism of finitely presented algebras is shown to be polynomial time many-one equivalent to the problem of graph isomorphism.

*This research has been supported in part by National Science Foundation Grant DCR75-09433.

1. Introduction

In this paper we study the complexity of some decision problems of finitely presented algebras, a class of simple algebraic structures.

An algebra $\mathcal{A}$ is finitely presented if there is a finite set G of generator symbols, a finite set O of operator symbols of various finite arities, and a finite set Γ of axioms or defining relations of the form $x \equiv y$, where x and y are well-formed terms over G and O, such that $\mathcal{A}$ is isomorphic to the free algebra on G and O modulo the congruence induced by Γ. That is, if τ is the free algebra (algebra of terms) over G and O, and $\equiv_\Gamma$ is the smallest congruence relation satisfying the relations Γ, then $\mathcal{A}$ is isomorphic to the quotient algebra ${}^{\tau}/{\equiv_\Gamma}$ with domain $\{ [x] \mid x \in \tau,\ [x] \text{ is the } \equiv_\Gamma\text{-congruence class of } x\}$.

For example, the two element Boolean algebra is presented by

$$G = \{0,1\}$$

$$O = \{\wedge, \vee, \neg\}$$

$$\Gamma = \{0\wedge 0\equiv 0,\ 0\wedge 1\equiv 0,\ 1\wedge 0\equiv 0,\ 1\wedge 1\equiv 1,\ 0\vee 0\equiv 0,\ 0\vee 1\equiv 1,\ 1\vee 0\equiv 1,\ 1\vee 1\equiv 1,\ \neg 0\equiv 1,\ \neg 1\equiv 0\}.$$

All algebras with finite domains are finitely presented. Finitely presented algebras may be infinite, but infinite groups and semigroups are never finitely presented, since an axiom schema (a rule representing infinitely many axioms) is needed to postulate associativity.

There is a strong relationship between finitely presented algebras and the finite tree automata of Thatcher and Wright[12] and Doner[13]. This relationship is summed up in the following theorem, analogous to a theorem of Nerode[9] regarding the representation of regular sets over a semigroup.

Theorem

L is a regular tree language (accepted by a finite tree automaton) over τ iff L is a union of congruence classes of a finitely generated congruence relation on τ of finite index. ■

Moreover, all congruence classes of any finitely generated congruence, finite index or not, are regular tree languages. The set of terms representing a single element of a finitely presented algebra is such a class.

Finite tree automata appear in diverse settings. Not only do they have a substantial theory of their own (see [10,11] for a good bibliography), but they have also been used in logic to show the decidability of some second order theories[12,13,14] and in formal language theory to study derivation trees of context free grammars (see [10,11]). In view of the above theorem, the complexity results presented here should apply to those areas.

Most of the decision problems addressed herein can be restated as problems of tree replacement systems, hence our complexity results carry over into that area.

Finally, very recent results, notably [1,2], have pegged down the complexities of various decision problems in different algebras. The present results fill a large gap here, and so would be essential to a general theory of the complexity of algebraic decision problems.

In spite of the above, the results presented here are most interesting not for any of these reasons. Their real interest lies in the generality and expressive power of the language of universal algebra. The finite structures that interest computer scientists, e.g. graphs, are easily represented as finitely presented algebras, and many known complete problems for P, NP, etc., can be reformulated easily as natural questions about finitely presented algebras, as evidenced by the trivial (often gsm) reductions from known complete problems to the problems discussed in this paper. Thus finitely presented algebras should be viewed as a unifying framework in which many of the interesting questions of low-level complexity can be reformulated.

In §3 we give several natural problems of finitely presented algebras which are $\leq^m_{\log}$-complete for P. These problems generalize known problems complete for P. In §4 we look at axiom schemata of the form $x \equiv y$ where x and y are terms with variables, and show that the schema satisfiability problem is $\leq^m_{\log}$-complete for NP and the schema validity problem is $\leq^m_{\log}$-complete for co-NP. In §5 we show that the problem of isomorphism of finitely presented algebras is polynomial time many-one equivalent to the problem of graph isomorphism.

2. Preliminaries

Definition

Let <M,ARITY> be a ranked alphabet, i.e., M is a finite set of symbols and ARITY: M → N, where N is the set of nonnegative integers. Partition M into two sets:

$G = \{a \in M \mid \text{ARITY}(a) = 0\}$ are generator symbols,

$O = \{a \in M \mid \text{ARITY}(a) > 0\}$ are the operator symbols. ■

We will use variables a,b to denote elements of G and θ, ϕ to denote elements of O.. Let M* be the set of finite length strings over M.

Definition

The set of terms over M is the smallest subset of M* such that

i) all elements of G are terms;

ii) if θ is m-ary (i.e. ARITY(θ)=m) and $x_1,\ldots,x_m$ are terms, then $\theta x_1 \ldots x_m$ is a term. ■

Denote the set of terms by τ . Variables w,x,y,z will range over terms.

τ may be viewed as the domain of an algebra with operations O defined by

$$\theta(x_1,\ldots,x_m) = \theta x_1 \ldots x_m \text{ if } \theta \text{ is m-ary.}$$

In this light we will refer to τ as the free algebra on M.

Definition

$x \blacktriangleleft y$ if x is a (not necessarily proper) subterm of y.

$x[y\backslash z]$ is the term x with all occurrences of the term y in x replaced by z.

If y' is a particular occurrence of a term y as a subterm of x, then $x[y'\backslash z]$ is the term x with that occurrence only replaced by z. ■

Definition

A binary relation $\eqsim$ on τ is a congruence provided

i) $\eqsim$ is an equivalence relation

ii) if θ is m-ary and $x_1,\ldots,x_m,y_1,\ldots y_m$ are terms such that $x_i \eqsim y_i$, $1\le i\le m$, then $\theta x_1\ldots x_m \eqsim \theta y_1\ldots y_m$. ■

In the above definition, ii) guarantees that the operations O are well-defined on $\eqsim$-congruence classes, thus we can form the quotient algebra $^{\tau}/_{\eqsim}$ with domain

$\{[x] \mid x \in \tau,\ [x]$ is the $\eqsim$-congruence class of $x\}$

and operations O.

Definition

Let Γ be a set of unordered pairs of terms. These pairs will be written $x\equiv y$ and will be called axioms or defining relations. Define $\equiv_{\Gamma}$ to be the smallest congruence on τ satisfying the axioms of Γ, and let

$$[x]_\Gamma = \{y \in \tau \mid x \equiv_\Gamma y\}.$$

We will omit the subscript Γ from $\equiv_\Gamma$ and $[x]_\Gamma$ when it is understood. It is straightforward to show that $x \equiv_\Gamma y$ iff it can be deduced from:

i) $x \equiv_\Gamma x$

ii) $\dfrac{x \equiv_\Gamma y}{y \equiv_\Gamma x}$

iii) $\dfrac{x \equiv_\Gamma y,\ y \equiv_\Gamma z}{x \equiv_\Gamma z}$

iv) $\dfrac{x_1 \equiv_\Gamma y_1, \ldots, x_m \equiv_\Gamma y_m,\ \mathrm{ARITY}(\theta)=m}{\theta x_1 \ldots x_m \equiv_\Gamma \theta y_1 \ldots y_m}$

v) $x \equiv_\Gamma y$ for all axioms $x \equiv y$ of Γ.

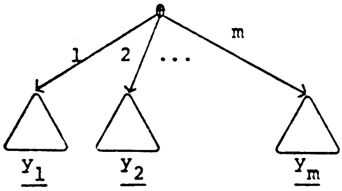
Definition

An algebra $\mathscr{A}$ is presented by $\langle M, \mathrm{ARITY}, \Gamma \rangle$ if $\mathscr{A}$ is (isomorphic to) $^{\tau}/_{\equiv_\Gamma}$. The triple $\langle M, \mathrm{ARITY}, \Gamma \rangle$ is called a presentation of $\mathscr{A}$. $\mathscr{A}$ is finitely presented if a presentation can be found with Γ a finite set. ■

It is convenient to represent terms as labeled trees, as follows:

i) if $a \in G$ then a is represented by a single vertex with label a.

ii) $\theta y_1 \ldots y_m$ is represented by

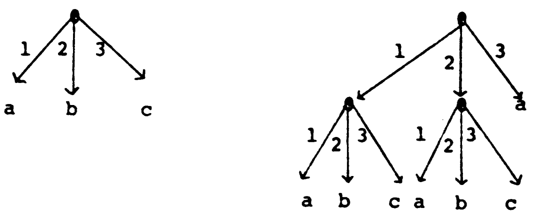


where the root has label ● and $\underline{y_i}$ is the tree representation of term y_i.

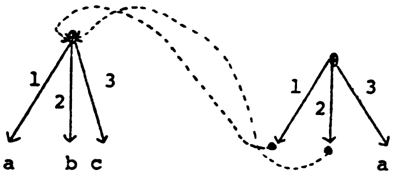
This representation has three immediate advantages:

1) We can give a presentation consisting of a finite set of trees labeled as above, with an extra undirected edge set AXIOM such that, if $\underline{y}$ is the tree representation of term y,
$x \equiv y$ is an axiom ↔ the roots of $\underline{x}$ and $\underline{y}$ are connected with an AXIOM edge.
We no longer need to specify M and ARITY, since these are implicit in the new representation.

2) We can represent terms with common subterms more concisely, by "factoring out" the common subterm, i.e. representing a set of trees as a dag.
E.g. terms ●abc, ●●abc●abca
may be represented as trees by



and then as dags, after factoring, by



3) We have a conceptually simpler deductive system for proving congruence of terms:

Define $\underline{x} \to \underline{y}$ ($\underline{x}$ derives $\underline{y}$ in one step) if there is an axiom $z \equiv w$ in Γ and an occurrence of $\underline{z}$ as a subtree of $\underline{x}$ such that if that occurrence of $\underline{z}$ is replaced by $\underline{w}$, then the result is $\underline{y}$.

Let $\overset{*}{\to}$ be the reflexive transitive closure of $\to$. The following is proved by an easy induction:

Theorem

$$x \equiv y \quad \text{iff} \quad \underline{x} \overset{*}{\to} \underline{y}. \qquad \blacksquare$$

For these reasons we will henceforth adopt the new representation, and use the words "term" and "tree" interchangeably. We will allow trees to be represented as dags by factoring out common subterms, and we will consider a presentation to be given by a dag with AXIOM edges, as outlined above. We will reuse the symbol Γ for the dag representing presentation <M,ARITY,Γ>.

Definition

A proof of $x \equiv y$ is a sequence $x_1 \ldots x_n$ of terms such that $x = x_1 \to x_2 \to \ldots \to x_n = y$.

The root of the transformation in $x \to y$ is the root of the subtree replaced. We say the entire tree is replaced in $x \to y$ if the subtree of x that is replaced is the whole tree x. ■

Definition

Let Γ be given. The following sets will be used throughout:

$$R_\Gamma = \{x \mid \text{there is an axiom } y \equiv z \text{ in } \Gamma \text{ and } x \prec y\}.$$
$$r_\Gamma = \{[x]_\Gamma \mid x \in R_\Gamma\}.$$ ■

Thus R_Γ is the set of terms appearing in the presentation, and r_Γ are the elements of the presented algebra represented by terms in R_Γ. The subscript Γ will be omitted when understood.

It is assumed the reader is familiar with the complexity classes P,NP, etc., the reducibilities $\leq^m_{\log}$ and $\leq^m_p$, and the notion of completeness.

3. Problems complete for P

The obvious first task is to determine the complexity of deciding whether two terms represent the same element of the algebra.

Definition

The word problem is the set

$$WP = \{\langle\Gamma,x,y\rangle \mid x \equiv_\Gamma y\}. \qquad ■$$

WP would more accurately be called the uniform word problem, since the presentation Γ is an input parameter.

Theorem 1

$WP \in P$.

Proof

Let $\langle\Gamma,x,y\rangle$ be input. Let Γ^+ be the graph Γ plus the vertices and edges of x and y. Let $R^+ = R_{\Gamma^+}$. We will describe a polynomial time algorithm to construct a new undirected edge set E on Γ^+ so that $\forall z,w \in R^+$, the roots of z and w are connected by an E edge (henceforth zEw) iff $z \equiv w$.

Step 0. Add edges wEw, $w \in R^+$.

Add edges wEz for all axioms $w \equiv z$.

Step n. If $u,v,w \in R^+$ and uEv, vEw, then add edge uEw. If $\theta y_1 \dots y_m$, $\theta x_1 \dots x_m \in R^+$ and $x_i E y_i$, $1 \le i \le m$, then add edge $\theta x_1 \dots x_m E \theta y_1 \dots y_m$. If no new edges were added at Step n, then stop.

The algorithm is clearly polynomial, since at most n^2 edges can be added.

Claim

$\forall w,z \in R^+$ wEz iff $w \equiv z$.

Proof of claim

(→)Easy induction on the step at which the edge was added, since no edge is added unless forced to be by the properties of $\equiv$.

(←)Define a relation $\sqsubset$ on congruent pairs of terms, as follows:

$\langle x,y \rangle \sqsubset \langle z,w \rangle$ iff either

i) the shortest proof $x \overset{*}{\to} y$ is shorter than the shortest proof $z \overset{*}{\to} w$; or

ii) the shortest proofs $x \overset{*}{\to} y$ and $z \overset{*}{\to} w$ are the same length and $x \underset{\neq}{\prec} z$, $y \underset{\neq}{\prec} w$.

Clearly $\sqsubset$ is a well-founded relation, so we proceed by induction on $\sqsubset$.

Let $x,y \in R^+$.

Basis: length of proof $x \overset{*}{\to} y = 0$.

Then $x=y$ and xEy at step 0.

Induction step: $x \overset{*}{\to} y$ is a nonzero-length shortest proof.

One of the following two cases must hold:

Case 1: The entire tree is replaced somewhere in $x \overset{*}{\to} y$. In this case $\exists z,w$ $x \overset{*}{\to} z$, $z \equiv w$ is an axiom, $w \overset{*}{\to} y$. But $z,w \in R^+$ hence xEz and wEy by induction hypothesis, since they are congruent via shorter proofs; and zEw in step 0. Thus xEy within two more steps.

<u>Case 2:</u> The entire tree is never replaced in $x \overset{*}{\to} y$. Then $x = \theta x_1 \ldots x_m$, $y = \theta y_1 \ldots y_m$, and $x_i \overset{*}{\to} y_i$ via a proof of length shorter than or equal to $x \overset{*}{\to} y$ (the proof $x_i \overset{*}{\to} y_i$ is given by the transformations applied to the interior of x_i in the proof $x \overset{*}{\to} y$) and $x_i \lneq x$, $y_i \lneq y$, $1 \le i \le m$, hence $\langle x_i, y_i \rangle \sqsubset \langle x, y \rangle$, $1 \le i \le m$. But all $x_i, y_i \in R^+$, hence by the induction hypothesis, $x_i E y_i$, $1 \le i \le m$. Then in the next step of the algorithm, xEy. ■

<u>Definition</u>

An instance of the <u>circuit value problem</u> (CVP) is a list $\mathscr{B}$ of assignments to variables $C_1, C_2, \ldots, C_n$ of the form

$$C_i = 0,$$
$$C_i = 1,$$
$$C_i = C_j \wedge C_k,\ j,k<i,$$
$$\text{or } C_i = C_j \vee C_k,\ j,k<i,$$

such that each C_i appears on the left side of an assignment exactly once. $\mathscr{B}$ is in CVP provided $\mathrm{val}(C_n) = 1$, where $\mathrm{val}(C_i)$ is the Boolean value of C_i computed from the list of assignments in the obvious way. ■

As demonstrated by Ladner[3], CVP is $\le^m_{\log}$-complete for P.

<u>Theorem 2</u>

CVP $\le^m_{\log}$ WP.

Proof

Given the above instance of CVP, take

$G = \{C_1,\ldots,C_n,0,1\}$

$O = \{\wedge,\vee\}$

$\Gamma = \mathscr{B} \cup \{0\vee 0\equiv 0,\ 0\vee 1\equiv 1,\ 1\vee 0\equiv 1,\ 1\vee 1\equiv 1,$
$0\wedge 0\equiv 0,\ 0\wedge 1\equiv 0,\ 1\wedge 0\equiv 0,\ 1\wedge 1\equiv 1\}.$

The restrictions on $\mathscr{B}$ and the eight extra axioms guarantee that the algebra presented by Γ is the two element lattice, and $\mathscr{B} \in$ CVP iff $C_n \equiv 1$ iff $\langle\Gamma,C_n,1\rangle \in$ WP. ■

Observe that CVP is really a special case of the word problem, as shown by the trivial (gsm) reduction.

Corollary 3

WP is $\le^m_{\log}$-complete for P. ■

We now wish to show the following three problems complete for P.

Definition

TRIV = $\{\Gamma \mid \Gamma$ presents the trivial (one element) algebra$\}$.

FIN = $\{\Gamma \mid \Gamma$ presents a finite algebra$\}$.

GEN = $\{\langle\Gamma, x_1,\ldots,x_n,y\rangle \mid [y]$ is contained in the subalgebra of $^{\tau}/\equiv$ generated by $[x_1],\ldots,[x_n]\}$.

Theorem 4

TRIV $\in$ P.

Proof

Use the algorithm of Theorem 1 to decide for all a,b ϵ G whether $a \equiv b$, then for all $\theta \,\epsilon\, O$ whether $\theta a \ldots a \equiv a$. ■

Theorem 5

CVP $\le^{m}_{\log}$ TRIV.

Proof

Let $\mathscr{B}$ be an instance of CVP. Construct Γ as in Theorem 2 and let $\Gamma' = \Gamma \cup \{C_n \equiv 0\}$. Then

$\mathscr{B} \,\epsilon\,$ CVP iff $C_n \equiv_\Gamma 1$ iff $1 \equiv_{\Gamma'} 0$ iff ${}^{\tau}/\!\equiv_{\Gamma'}$ is trivial. ■

Corollary 6 ■

TRIV is complete for P.

We now wish to show GEN is complete for P. GEN is a more general formulation of the problem of the same name of Jones and Laaser.[4]

Theorem 7

GEN ϵ P.

Proof

Given $\langle\Gamma, x_1, \ldots, x_n, y\rangle$, let $\mathscr{A}$ be the subalgebra of ${}^{\tau}/\!\equiv$ generated by $[x_1], \ldots, [x_n]$, and let Σ be the subalgebra of τ generated by $x_1, \ldots, x_n$. Then ${}^{\Sigma}/\!\equiv$ is isomorphic to $\mathscr{A}$, and $[y] \,\epsilon\, \mathscr{A}$ iff $\exists x \,\epsilon\, \Sigma$ such that $y \equiv x$.

Let $\Gamma^{+} = \Gamma \cup \{x_1, \ldots, x_n, y\}$ and let $R^{+} = R_{\Gamma^{+}}$. Consider the following algorithm to mark elements of R^{+} (vertices of Γ^{+}):

Step 0. Run the algorithm of Theorem 1 on Γ^+ to determine for all $x,w \in R^+$ whether $x \equiv w$. Mark each x_i.

Step n. If $\theta y_1 \ldots y_m \in R^+$ and $y_1, \ldots, y_m$ are marked, mark $\theta y_1 \ldots y_m$.

If $x,w \in R^+$, $x \equiv w$, and x is marked, mark w.

If no new terms are marked, stop.

The algorithm is clearly polynomial. The following claim establishes the result.

Claim

$[y] \in \mathcal{A}$ iff y is marked by the above algorithm.

Proof of claim

($\leftarrow$) clear.

($\rightarrow$) let $C_0 = \{x_1, \ldots, x_n\}$

$$C_{k+1} = \{\theta y_1 \ldots y_m \mid y_1, \ldots, y_m \in C_k\} \cup C_k .$$

Then $\cup_k C_k = \Sigma$. Let $y \in R^+$ such that $[y] \in \mathcal{A}$. Then $\exists x \in \Sigma \; y \equiv x$. We prove the result by induction on the least k such that $\exists x \in C_k \; y \equiv x$.

Basis: $k=0$. Then $y \equiv x_i$, and x_i is marked at step 0, hence y is marked at step 1.

Induction step: $k>0$. Then $y \equiv \theta y_1 \ldots y_m$ and $y_1, \ldots, y_m \in C_{k-1}$. We have no guarantee that $y_1, \ldots, y_m \in R^+$ however, hence we cannot claim that $y_1, \ldots, y_m$ are marked. But let us consider a shortest proof $y \overset{*}{\rightarrow} \theta y_1 \ldots y_m$. One of the following two cases occurs:

<u>Case 1</u>: The entire tree is never replaced in $y \overset{*}{\rightarrow} \theta y_1 \ldots y_m$. Then $y = \theta z_1 \ldots z_m$ and each $z_i \equiv y_i$. But each $y_i \in C_{k-1}$ hence $[z_i] \in \mathcal{A}$, thus since $z_i \in R^+$, by the induction hypothesis z_i is eventually marked. Then when each z_i is marked, y is marked in the next step.

<u>Case 2</u>: $y \overset{*}{\rightarrow} z$, $z \equiv w$ is an axiom, $w = \theta w_1 \ldots w_m$ and $w_i \equiv y_i$, $1 \le i \le m$. I.e., $z \rightarrow w$ is the last time in the proof $y \overset{*}{\rightarrow} \theta y_1 \ldots y_m$ that the entire tree is replaced. Then $w_i \in R^+$ and $y_i \in C_{k-1}$, hence by the induction hypothesis, w_i is eventually marked, and $w \in R^+$ hence w is marked; but $y \equiv w$, thus y is marked in the next step. ■

<u>Corollary 8</u>

GEN is complete for P.

<u>Proof</u>

There is a trivial reduction from Jones' and Laaser's GEN to our GEN. The details are left to the reader. ■

We now turn to the finiteness problem. Let Γ be a presentation of $\mathcal{A}$.

<u>Lemma 9</u>

$\mathcal{A}$ is finite iff $\mathcal{A} = r_\Gamma$.

<u>Proof</u>

(←) Trivial.

(→) Clearly $r_\Gamma \subseteq \mathcal{A}$. Now assume there is an x such that

$[x] \in \mathcal{A} - r_\Gamma$, i.e. $\forall y \in R_\Gamma\ x \not\equiv y$. Let x be $\prec$-minimal in its congruence class. Define a set of terms

$$x_0 = x$$
$$x_{n+1} = \theta x_n \ldots x_n$$

where θ is any operator. For all $j>0$, if $x_0 \equiv x_j$, then $x \not\preceq x_j$ and $x_j \overset{*}{\rightarrow} x$. But in the proof $x_j \overset{*}{\rightarrow} x$, no ancestor of x in x_j can ever be the root of a transformation, else x is congruent to a term in R_Γ, hence $x \overset{*}{\rightarrow} y$ in the proof $x_j \overset{*}{\rightarrow} x$, where y is a proper subterm of x, contradicting the assumption of $\prec$-minimality.

Proceeding by induction, if $x_{i+1} \equiv x_{j+1}$, $i \neq j$, then $\theta x_i \ldots x_i \overset{*}{\rightarrow} \theta x_j \ldots x_j$. But since $x \prec x_i$ and no transformation can be applied to an ancestor of x in x_i, it must be that $x_i \equiv x_j$, contradicting the induction hypothesis.

Hence $\{[x_0],[x_1],\ldots\}$ is an infinite subset of $\mathcal{A}$, contradicting the finiteness of $\mathcal{A}$. ■

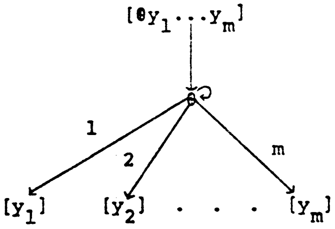
The above lemma uncovers a surprising fact: a finite algebra can be no bigger than its smallest finite presentation. Thus to give a finite presentation of a finite group, say, we might as well write down its multiplication table.

We wish to show a ptime algorithm to decide finiteness. The straightforward approach would be to show that if $\mathcal{A}$ is infinite then there is a small term (i.e. one of the form $\theta y_1 \ldots y_m$ where $y_1, \ldots, y_m \in R_\Gamma$) with $[\theta y_1 \ldots y_m] \notin r_\Gamma$, thus we would need only to run the algorithm of Theorem 1 on each small term y and each

$x \in R_\Gamma$ to see if $y \equiv x$. However, the arity of the operators is an input parameter and can grow as much as linearly with the size of the presentation, so there may be too many $@y_1 \ldots y_m$ to try; hence this approach works only when the arity of the operators is bounded. We circumvent this problem with the following construction.

Definition

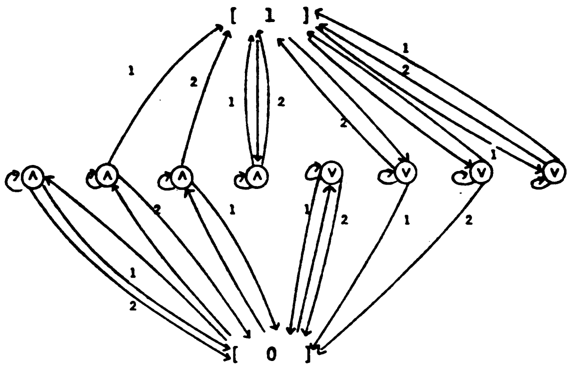
The characteristic graph of a presentation Γ, denoted χ_Γ, is a labeled directed graph with vertex labels O and edge labels {1,2,...,k} where k is the maximum arity of any operator in O. The vertex set of χ_Γ consists primarily of unlabeled vertices r_Γ, plus other labeled vertices and labeled and unlabeled edges such that χ_Γ is the smallest graph in which



appears as a subgraph of χ_Γ for every $@y_1 \ldots y_m \in R_\Gamma$. ■

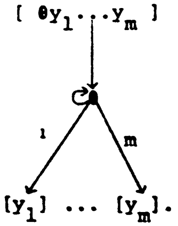
Example

The Γ in Theorem 2 presents the two element lattice. Its characteristic graph is:



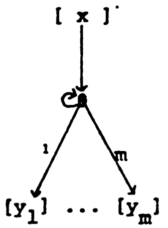
■

χ_Γ is meant to represent the interaction of the elements of r_Γ under the operations $\mathcal{O}$. χ_Γ is constructible in polynomial time: we can just run the algorithm of Theorem 1 to determine $\equiv_\Gamma$-classes (they appear as cliques of E-edges), then for each $\theta y_1 \ldots y_m$ in R_Γ add a vertex θ and edges



Lemma 10

For $x, y_1, \ldots, y_m \in R_\Gamma$, $x \equiv \theta y_1 \ldots y_m$ iff

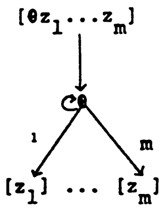


appears in χ_Γ.

Proof

(←) If the subgraph pictured appears in χ_Γ, then by construction of χ_Γ it must be that $\exists z_1,\ldots,z_m \;\; x \equiv \theta z_1 \ldots z_m$ and $z_i \equiv y_i$, $1 \le i \le m$. Then $x \equiv \theta z_1 \ldots z_m \equiv \theta y_1 \ldots y_m$.

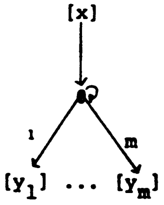
(→) Consider a proof $x \overset{*}{\to} \theta y_1 \ldots y_m$. If the entire tree is ever replaced, $\exists z_1 \ldots z_m \;\; x \overset{*}{\to} w \to \theta z_1 \ldots z_m$ and $z_i \overset{*}{\to} y_i$, $1 \le i \le m$, and $w \equiv \theta z_1 \ldots z_m$ is an axiom. If the entire tree is never replaced, then $\exists z_1 \ldots z_m \;\; x = \theta z_1 \ldots z_m \overset{*}{\to} \theta y_1 \ldots y_m$ and $z_i \overset{*}{\to} y_i$, $1 \le i \le m$. In either case,



appears in χ_Γ, and $[\theta z_1 \ldots z_m] = [x]$ and $[z_i] = [y_i]$, $1 \le i \le m$. ■

Lemma 11

$\mathcal{A}$ is finite iff for all m, for all m-ary $\bullet$ and $[y_1],\ldots[y_m] \in r_\Gamma$, there is an $[x] \in r_\Gamma$ such that



appears in χ_Γ.

Proof

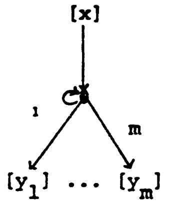
By the previous lemma, for all $\bullet$ and $[y_1],\ldots[y_m]$ there is an $[x] \in r_\Gamma$ such that the above graph appears in χ_Γ iff r_Γ is closed under all the operations O. But since $\mathcal{A}$ is generated by $\{[a] \mid a \in G\} \subseteq r_\Gamma$, this occurs iff $\mathcal{A}=r_\Gamma$. By Lemma 9, this occurs iff $\mathcal{A}$ is finite. ■

Theorem 12

FIN $\in$ P.

Proof

Construct χ_Γ, and for each m-ary $\bullet$, cycle through all $[y_i],\ldots,[y_m]$, rejecting if we ever find a $[y_1],\ldots,[y_m]$ such that for no $x \in R_\Gamma$ does



appear in χ_Γ. For each distinct $[y_1],\dots,[y_m]$ tested such that an [x] is found, there must be a distinct vertex ● (i.e. the one appearing in the above subgraph), hence the number of steps of the algorithm is polynomially bounded to the size of χ_Γ, which is polynomial in the input Γ. ■

Theorem 13

CVP $\leq^{m}_{\log}$ FIN.

Proof

We use the presentation Γ' constructed in Theorem 5, such that Γ' presents either the trivial algebra or the two element lattice, and

$$^{\tau}/\equiv_{\Gamma'} \text{ is trivial} \leftrightarrow \mathscr{B} \in \text{CVP}.$$

Append another generator symbol b to G, and the axioms $\{b\wedge b\equiv 0,\ b\wedge 0\equiv 0,\ 0\wedge b\equiv 0,\ b\vee b\equiv 0,\ b\vee 0\equiv 0,\ 0\vee b\equiv 0\}$ to Γ' to get Γ''. It is left to the reader to verify that Γ'' presents the trivial algebra if Γ' does, an infinite algebra otherwise. ■

Corollary 14

FIN is complete for P. ■

4. Two schema problems complete for NP and co-NP

Definition

Let τ = {terms over G and O}, as usual.

Let $V = \{v_1,\ldots,v_m\}$ be a set of variable symbols,

$G^+ = G \cup V$,

τ^+ = {terms over G^+ and O}.

Thus τ^+ is the set of terms with occurrences of variables V.

An assignment to variables is a map I: $V \to \tau$. If we take $I(a) = a$ for $a \in G$, then I extends uniquely to a homomorphism $\tau^+ \to \tau$, which we will also denote by I.

A schema is a formula $x \equiv y$, where $x,y \in \tau^+$. Given Γ, a schema $x \equiv y$ is satisfiable in ${}^{\tau}/{\equiv_\Gamma}$ if there is an assignment I such that $I(x) \equiv_\Gamma I(y)$. I.e., $x \equiv y$ is satisfiable if there is an interpretation of terms with variables over ${}^{\tau}/{\equiv_\Gamma}$ such that x and y represent the same element of ${}^{\tau}/{\equiv_\Gamma}$.

A schema $x \equiv y$ is valid in ${}^{\tau}/{\equiv_\Gamma}$ if for all assignments I, $I(x) \equiv_\Gamma I(y)$. ■

Definition

The schema satisfiability problem is the set

SATIS = $\{<\Gamma, x, y> \mid$ schema $x \equiv y$ is satisfiable$\}$.

The schema validity problem is the set

VALID = $\{<\Gamma,x,y> \mid$ schema $x \equiv y$ is valid$\}$. ■

Observe that the Boolean satisfiability problem of Cook[5] is a special case of SATIS, where the algebra presented by Γ is the two element Boolean algebra. It will be left to the reader to verify that there are very easy (gsm) reductions from the Boolean satisfiability problem to SATIS and from the Boolean tautology problem to VALID.

Fix $< \Gamma,x,y >$. $V = \{v_1,\ldots,v_m\}$ are the variables occurring in x and y. Define Γ^+ to be the graph Γ plus x and y, and let $R^+ = R_{\Gamma^+}$.

In the following, if $X = \{v_{i_1},\ldots,v_{i_k}\}$ is a set of variables, and $I\colon X \to \tau$ is a partial assignment, we will write

$$z[X \setminus I[X]]$$

in place of

$$z[v_{i_1}\setminus I(v_{i_1}) ,\ldots, v_{i_k}\setminus I(v_{i_k})].$$

Definition

An assignment $I\colon V \to \tau$ is well-specified on $X \subseteq V$ if

$$\exists f\colon X \to R^+ \text{ such that if } v \in X \text{ then}$$

$$I(v) = f(v)[X \setminus I[X]]. \qquad \blacksquare$$

I.e. the values of I on members of X are uniquely determined by assigning terms in R^+ to variables. We wish to show that if $x \equiv y$ is satisfiable, then it is satisfiable via an assignment I well-specified on V. This will allow us to guess the map f and use the word problem algorithm to verify that $I(x) \equiv_\Gamma I(y)$, thereby showing SATIS $\in$ NP.

Lemma 15

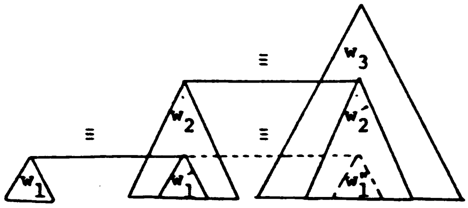
If $x \equiv y$ is satisfiable, then it is satisfiable via an assignment well-specified on V.

Proof

Suppose $x \equiv y$ is satisfiable, and $X \neq V$ is a maximal subset of V such that there is an assignment I well-specified on X satisfying $x \equiv y$.

Let us call a term $z \in \tau$ well-specified if it is congruent to some $w[X \setminus I[X]]$, where $w \in R^+$; ill-specified otherwise. We observe that if a term $w \in R^+$ has variables exclusively in X, then any subterm of $w[X \setminus I[X]]$ is well-specified. In particular, any term in R^+ without variables is well-specified, and every $I(v)$ for $v \in X$ is well specified.

Define a binary relation $\prec$ on ill-specified terms by: $w \prec z$ iff w is congruent to a proper subterm of z. The fact that $\prec$ is defined only on ill-specified terms guarantees transitivity. For, suppose $w_1 \prec w_2 \prec w_3$. Then $w_1 \equiv w_1 \lneq w_2$, and $w_2 \equiv w_2' \lneq w_3$, as shown.



But in a proof $w_2 \overset{*}{\rightarrow} w_2'$, no transformation can be applied to an ancestor of w_1', else w_1 would be congruent to a term in R, hence an occurrence of v in x or y. Thus

$$x'' = x[V \setminus I'[V]] \text{ and}$$

$$y'' = y[V \setminus I'[V]]$$

where $I'(v)$ is $I(v)$ with all occurrences of subterms congruent to $I(v^*)$ replaced by $I(v^*)$ itself. But since there is a proof $x'' \overset{*}{\to} y''$ leaving occurrences of $I(v^*)$ intact, any term in place of $I(v^*)$, including any well-specified one, will not alter the proof, so let $a \in G$ and define

$$I''(v) = I'(v)[I(v^*) \setminus a] \text{ for all } v \in V,$$

$$x''' = x''[I(v^*) \setminus a]$$

$$= x[V \setminus I''[V]],$$

$$y''' = y''[I(v^*) \setminus a]$$

$$= y[V \setminus I''[V]].$$

Then $x''' \equiv y'''$, so I'' satisfies $x \stackrel{?}{\equiv} y$, and $I''(v) = I(v)$ for $v \in X$, and $I''(v^*) = a$, hence I'' is well-specified on $X \cup \{v^*\}$, contradicting the maximality of X. ■

Let $\mathscr{A} = {}^{\tau}/{\equiv_{\Gamma}}$ and let $x \stackrel{?}{\equiv} y$ be a schema with variables V.

Lemma 16

$x \stackrel{?}{\equiv} y$ is satisfiable in $\mathscr{A}$ iff there is a map $f : V \to R^+$ such that if $\Gamma^+ = \Gamma \cup \{v \equiv f(v) \mid v \in V\}$ then

i) every variable is $\equiv_{\Gamma^+}$ -congruent to a term in τ,

ii) $x \equiv_{\Gamma^+} y$.

well-specified. Thus $w_1' \equiv w_1'' \not\preceq w_2'$ and $w_1 \prec w_3$ follows. By a similar argument, we can show $\prec$ is well-defined on $\equiv$- classes: i.e. if $w_1 \equiv w_2 \prec w_3 \equiv w_4$ then $w_1 \prec w_4$.

Moreover, $\succ$ is acyclic, since if $w_1 \prec w_2 \prec \cdots \prec w_m \succ w_1$ then by transitivity $w_1 \prec w_1$. If w_0 is ◂-minimal in the congruence class of w_1, then $w_0 \equiv w_1 \succ w_1 \equiv w_0$, and hence w_0 is congruent to a proper subterm of itself, contradicting ◂-minimality.

Thus $\prec$ restricted to $\{I(v) \mid v \in V\text{-}X\}$ is well-founded. Let $v^* \in V\text{-}X$ such that $I(v^*)$ is $\prec$-minimal, and let

$$x' = x[V \setminus I[V]]$$

$$y' = y[V \setminus I[V]].$$

Since $\prec$ is acyclic and $w_1 \not\preceq w_2 \rightarrow w_1 \prec w_2$, we have that any two distinct occurrences of subterms of x' congruent to $I(v^*)$ are ◂-incomparable; thus replacing all occurrences of terms congruent to $I(v^*)$ with $I(v^*)$ itself is a well-defined notion. Let x'' and y'' be formed from x' and y' in this way. Then $x'' \equiv x' \equiv y' \equiv y''$. But consider a proof $x'' \overset{*}{\rightarrow} y''$. No transformation can ever be applied to an ancestor of $I(v^*)$, else $I(v^*)$ was well-specified. Thus all occurrences of $I(v^*)$ in x'' go to occurrences of $I(v^*)$ in y'' in the corresponding positions, and vice-versa. Thus there is a proof $x'' \overset{*}{\rightarrow} y''$ leaving each occurrence of $I(v^*)$ intact.

Finally, it follows from the $\prec$-minimality of $I(v^*)$ that each occurrence of a subterm of x' or y' congruent to $I(v^*)$ occurs as a subterm of some ill-specified $I(v)$ which replaced

Proof

($\rightarrow$) Let $X_0 = \{v \in V \mid f(v) \in \tau\}$. X_0 must be nonempty, else i) is contradicted. Similarly if X_n has been defined and $X_n \subsetneq V$, then $X_{n+1} = \{v \in V \mid f(v)$ contains only variables in $X_n\}$ must contain a variable not in X_n. Thus $\cup_n X_n = V$. Let $p(v) =$ least n such that $v \in X_n$. Define $I(v) = f(v)$ for $v \in X_0$, and for $v \in X_{n+1} - X_n$ define $I(v) = f(v)[X_n \setminus I[X_n]]$. An easy induction on p(v) gives $v \equiv_{\Gamma}+ I(v)$ for all $v \in V$.

Now we claim that I satisfies $x \equiv y$. Let

$$x' = x[V \setminus I[V]],$$
$$y' = y[V \setminus I[V]].$$

We have $x' \equiv_{\Gamma}+ y'$, by ii). We need to show $x' \equiv_\Gamma y'$. Consider a proof $x' \overset{*}{\rightarrow} y'$. If no variable ever appears in the proof, then done. Otherwise let n be the largest number such that $v \in V$ appears in the proof with $p(v) = n$. v appears only via application of the axiom $v \equiv f(v)$. But the only way an ancestor of v can be the root of a subsequent transformation is if a variable with a larger p appears. But v eventually disappears, since y' has no variables, and this can occur only if $v \equiv f(v)$ is applied. Thus we can eliminate the two transformations $f(v) \rightarrow v$ and $v \rightarrow f(v)$. This process may be repeated until no variables appear in the proof $x' \overset{*}{\rightarrow} y'$; then $x \equiv_\Gamma y$.

($\leftarrow$) By the previous lemma, there is an assignment I well-specified on V satisfying $x \equiv y$, i.e. $\exists f : V \rightarrow R^+$ such that $I(v) = f(v)[V \setminus I[V]]$. It is left to the reader to verify that if $\Gamma^+ = \Gamma \cup \{v \equiv f(v) \mid v \in V\}$ then

i) $v \equiv_{\Gamma^+} I(v)$, and

ii) $x \equiv_{\Gamma^+} y$. ■

Theorem 17

SATIS $\in$ NP.

Proof

Guess the map $f{:}V \to R^+$, append axioms $\{v \equiv f(v) \mid v \in V\}$ to Γ to get Γ^+, and verify using the word problem algorithm that $x \equiv_{\Gamma^+} y$. ■

Theorem 18

Boolean satisfiability $\leq^m_{\log}$ SATIS.

Proof

The proof is left to the reader. ■

Corollary 19

SATIS is $\leq^m_{\log}$-complete for NP. ■

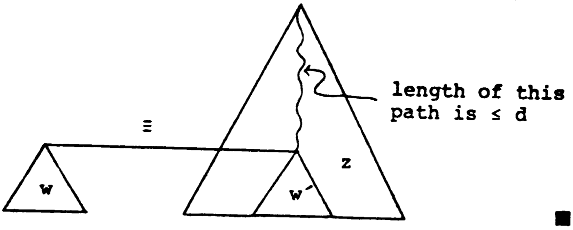
We have immediately that $\overline{\text{SATIS}}$ is complete for co-NP. Unfortunately, this does not help us with the problem VALID.

Let $\mathcal{A}$ be presented by Γ.

Definition

Let $d \in N$. Define $w \leq_d z$ iff

$\exists w'\ w \equiv w' \lessdot z$ and the length of the path from the root of w' to the root of z is at most d, as illustrated below.



Lemma 20

If $\mathscr{A}$ is infinite, then for any d,k, there exists a set of k arbitrarily large terms $\{w_1,\dots,w_k\}$ such that if $i \neq j$ then w_i and w_j are $\preceq_d$-incomparable.

Proof

If k = 1, then since $\mathscr{A}$ is infinite, an arbitrarily large w_1 can be chosen not congruent to any smaller term. Now suppose $\{w_1,\dots,w_{k-1}\}$ have been constructed so that $w_i \not\preceq_d w_j$ for $i \neq j$. Take w_k^0 to be larger than all terms in R_Γ and larger than all of $w_1,\dots,w_{k-1}$ such that w_k^0 is not congruent to any smaller term. Let θ be an operator, and define $w_k^{m+1} = \theta w_k^m \cdots w_k^m$. Then $\forall m,n\ \forall i < k\ w_k^m \not\preceq_n w_i$; if not, then for some m, $w_k^0 \lessdot w_k^m \equiv w' \lessdot w_i$, and an analysis of a proof $w_k^m \overset{*}{\rightarrow} w'$ shows that w_k^0 would necessarily be congruent to a smaller term. Moreover, $\forall m\ \forall i < k$, $w_i \not\preceq_m w_k^m$, by induction on m: certainly $w_i \preceq_0 w_k^0$ implies $w_i \equiv w_k^0$, contrary to our choice of

w_k^0; and if $w_i \not\uparrow_m w_k^m$ then $w_i \underline{\lambda}_{m+1} w_k^{m+1}$ implies $w_i \equiv w_k^{m+1}$, which by analysis of a proof $w_i \overset{*}{\rightarrow} w_k^{m+1}$ can be shown impossible, as above. Now let $w_k = w_k^d$. Then $\{w_1,\dots,w_k\}$ is desired set. ■

Theorem 21

VALID ϵ co-NP.

Proof

Let Γ present $\mathscr{A}$ and let $x \equiv y$ be a schema with variables V. Use the algorithm of Theorem 12 to determine whether $\mathscr{A}$ is finite. If so then $\mathscr{A} = r_\Gamma$, so verify in parallel for each assignment $I:V \rightarrow R_\Gamma$ that $x[V \setminus I[V]] \equiv y[V \setminus I[V]]$, using the word problem algorithm. If $\mathscr{A}$ is infinite, let $\Gamma^+ = \Gamma$ but consider Γ^+ as inducing a congruence over τ^+ instead of τ, where $\tau^+ = \{\text{terms over } G \cup V, 0\}$. Let $R^+ = R_{\Gamma^+}$.

Claim

If $\mathscr{A}$ is infinite, $x \equiv y$ is valid iff $x \equiv_{\Gamma^+} y$.

The claim establishes the result, since if $\mathscr{A}$ is infinite we can use the word problem algorithm to check $x \equiv_{\Gamma^+} y$, and the parallel procedure given above for the finite case is a Π_p^1 computation[6], thus in co-NP.

Proof of claim

($\leftarrow$) is clear. To show ($\rightarrow$), suppose $x \equiv y$ is valid. Let $d > \max\{\text{height}(z) \mid z \epsilon R^+\}$ and let $V = \{v_1,\dots,v_k\}$. The previous lemma guarantees arbitrarily large $w_1,\dots,w_k$ such that if $i \neq j$ then $w_i \not\uparrow_d w_j$. Take such $w_1,\dots,w_k$ to be each of height > d and let $I(v_i) = w_i$, $1 \le i \le k$. Let

$$x' = x[V \setminus I[V]],$$
$$y' = y[V \setminus I[V]].$$

Then $x' \overset{*}{\to} y'$, but no ancestor of any w_i is ever the root of a transformation, thus $x' \overset{*}{\to} y'$ via a proof in which all transformations to terms $\trianglelefteq$-incomparable to any w_i are done first, followed by transformations to the interiors of the w_i. Thus $\exists z \;\; z' = z[V \setminus I[V]]$, $x \equiv_{\Gamma}+ z$, and $z' \overset{*}{\to} y'$ via transformations only to the interiors of the w_i. But since the w_i are so big, each w_i in z' must go to either a subterm or superterm of some w_j in y' in the proof $z' \overset{*}{\to} y'$. But this says that either $w_i \trianglelefteq_d$ or w_j or $w_j \trianglelefteq_d w_i$, hence it must be that $i = j$. But then $z' = y'$. Since for $i \neq j$ we have $w_i \not\trianglelefteq_d w_j$, all occurrences of w_i in z' or y' replaced an occurrence of v_i in z' or y', thus $z = y$. Therefore $x \equiv_{\Gamma}+ y$, and the claim is verified. ■

Theorem 22

VALID is $\le^{m}_{\log}$-hard for co-NP.

Proof

The reduction from the Boolean tautology problem is left to the reader. ■

Corollary 23

VALID is $\le^{m}_{\log}$-complete for co-NP. ■

It should be noted that if we allow quantification over variables, deciding membership in SATIS is equivalent to deciding truth of closed formulas (those in which all variables are quantified) of the form

$$\exists v_1 \ldots \exists v_n \; x \equiv y,$$

and deciding membership in VALID is equivalent to deciding truth of closed formulas of the form

$$\forall v_1 \ldots \forall v_n \; x \equiv y,$$

in the algebra presented by Γ (equivalently, in all algebras satisfying the axioms of Γ). This is quite remarkable in view of the fact that the quantified variables range unboundedly over a possibly infinite set. In other results of this type,[5,7] either the structure is finite, or the quantifiers are bounded.

If we define S_n, V_n by

$S_n(V_n) = \{<\Gamma, \bar{Q}\; x\equiv y > \mid$ $\bar{Q}\; x\equiv y$ is a closed formula where $\bar{Q}$ is a string of quantifiers with n alternations, the outermost a $\exists(\forall)$, $\bar{Q}\; x\equiv y$ is true in ${}^{\tau}/{\equiv_\Gamma}\}$

and if we let Σ_p^n and Π_p^n represent the n^th^ Σ and Π levels of the polynomial time hierarchy,[7] we have by the preceding results

i) $S_0 = V_0$ is $\le^m_{\log}$-complete for $\Sigma_p^0 = \Pi_p^0 = P$;

ii) S_1 is $\le^m_{\log}$-complete for $\Sigma_p^1 = NP$;

iii) V_1 is $\le^m_{\log}$-complete for $\Pi_p^1 = \text{co-}NP$.

It is conjectured that, like other results in this area,[6,7] S_n is $\le^m_{\log}$-complete for Σ_p^n and V_n is $\le^m_{\log}$-complete for Π_p^n, for all n; and $\cup_n S_n \cup V_n$ is complete for PSPACE.

5. Isomorphism of finitely presented algebras

In this section we wish to show that the problem of isomorphism of finitely presented algebras (ISOM) is polynomially equivalent to the problem of graph isomorphism.

As before, the reduction from the graph problem (the more specific) to the algebra problem (the more general) is trivial. To go the other way, we show that every finitely presented algebra has a "reduced" presentation, which is unique in a certain well defined sense. In view of the relationship between finitely presented algebras and regular tree languages noted in §1, this result corresponds roughly to the minimization of states in a finite tree automaton.

In proving ISOM $\equiv_p^m$ graph isomorphism, we use the reduction sequence

isomorphism of undirected graphs without multiple edges or loops

$\le_{\log}^m$ ISOM

$\le_p^m$ isomorphism of labeled directed graphs

$\le_{\log}^m$ isomorphism of directed graphs

$\le_{\log}^m$ isomorphism of directed graphs without multiple edges or loops

$\le_{\log}^m$ isomorphism of undirected graphs without multiple edges or loops.

The $\leq^m_{\log}$ reductions in the above sequence are easy exercises and are left to the reader.

Definition

ISOM = $\{\langle\Gamma,\Delta\rangle \mid \Gamma$ and Δ present isomorphic algebras$\}$. ■

We will assume that the number of operator symbols of each arity in O_Γ and O_Δ is the same. The interested reader may verify that there is a polynomial time algorithm to check whether two operator symbols in any O specify the same operation, hence the assumption is without loss of generality.

To prove the reduction ISOM $\leq^m_p$ graph isomorphism, we will show that every finitely presented algebra $\mathcal{A}$ has a "reduced" presentation Γ, which can be found in polynomial time, such that r_Γ is unique (up to isomorphism). But r_Γ is uniquely represented by the characteristic graph χ_Γ introduced in §3, as shown by the following lemma:

Lemma 24

r_Γ and r_Δ are isomorphic (as subsets of algebras) iff χ_Γ and χ_Δ are isomorphic (as graphs).

Proof

Follows directly from Lemma 10. ■

Definition

A presentation is reduced provided

i) if $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ is an axiom, then one of $x_1 \not\equiv y_1, \ldots, x_m \not\equiv y_m$.

ii) no axiom of the form $a \equiv x$ occurs, where $a \in G$ and a does not occur in x. ■

Lemma 25

There is a polynomial time algorithm which for input Γ gives an equivalent reduced Γ^*.

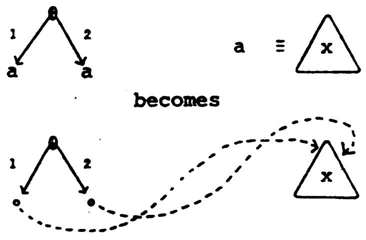
Proof

Given R_Γ, find all congruent pairs, using the word problem algorithm of Theorem 1. Repeat the following two steps until no more changes occur:

a) If $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ is an axiom and $x_1 \equiv y_1, \ldots, x_m \equiv y_m$ follow, replace $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ in Γ with new axioms $x_1 \equiv y_1, \ldots, x_m \equiv y_m$.

b) If $a \equiv x$ is an axiom, $a \in G$, and a does not occur in x, replace all occurrences of a in other terms of R_Γ with pointers to x, and eliminate the axiom $a \equiv x$.

E.g.



We claim first that this algorithm is polynomially time bounded. Note that step b) occurs at most n times, since each time, a generator symbol disappears. For each occurrence of b), step a) occurs at most n^2 times, provided whenever $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ and $\theta z_1 \ldots z_k \equiv \theta w_1 \ldots w_k$ are axioms, and $x_i \equiv y_i$, $z_j \equiv w_j$ for $1 \le i \le m$, $1 \le j \le k$, and $\theta z_1 \ldots z_k < \theta x_1 \ldots x_m$, step a) is applied to the axiom $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ first. We must also insure that every time b) occurs, a valid presentation results, i.e. the graph remains acyclic. This follows from the requirement in b) that a not occur in x in the axiom $a \equiv x$.

Hence the algorithm halts in polynomial time with a reduced presentation Γ^*, so it remains to show that Γ^* and Γ are equivalent, i.e. present the same algebra. If a) is applied, we have axiom $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m \in \Gamma$ and $x_1 \equiv_\Gamma y_1, \ldots, x_m \equiv_\Gamma y_m$. Let $\Gamma' = \Gamma - \{\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m\}$. Since $\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m$ follows from $\Gamma' \cup \{x_1 \equiv y_1, \ldots, x_m \equiv y_m\}$, we have that under the assumptions Γ', $\{x_1 \equiv y_1, \ldots, x_m \equiv y_m\}$ and $\{\theta x_1 \ldots x_m \equiv \theta y_1 \ldots y_m\}$ are equivalent. If b) is applied, let $a \equiv x$ be the axiom removed. Let $\tau' = \{$terms over O & G-$\{a\}\}$, and let $f : \tau \to \tau'$ be given by $f(y) = y[a \setminus x]$. An application of b) replaces axioms $\Gamma = \{x_1 \equiv y_1, \ldots, x_k \equiv y_k\}$ with $\Gamma' = \{f(x_1) \equiv f(y_1), \ldots, f(x_k) \equiv f(y_k)\}$, inducing congruence $\equiv' = \equiv_{\Gamma'}$, so we need to show that $\tau/_\equiv$ and $\tau'/_{\equiv'}$ are isomorphic. But it is easily verified that for $z, y \in \tau'$, $z \equiv' y$ iff $z \equiv y$, and each $y \in \tau - \tau'$ is congruent via $\equiv$ to $y[a \setminus x] \in \tau'$, thus there exists an h such that the diagram

$$\begin{tikzcd}
\tau \arrow[r, "f"] \arrow[d, "{[\ ]_\Gamma}"'] & \tau' \arrow[d, "{[\ ]_{\Gamma'}}"] \\
\tau/\equiv \arrow[r, "h"] & \tau'/\equiv'
\end{tikzcd}$$

commutes, and h is an isomorphism. ∎

In the following, let Γ and Δ be finite presentations of algebras $\mathscr{A}$ and $\mathscr{B}$, respectively. The symbols G, O, τ, etc. will have their usual meaning, except we will attach subscripts Γ and Δ to denote the presentation with which they are associated.

Lemma 26

Suppose $\mathscr{A}$ and $\mathscr{B}$ are isomorphic via h, and suppose Γ is reduced. Then there is a function $f\colon \tau_\Gamma \to \tau_\Delta$ such that

i) the diagram

$$\begin{tikzcd}
\tau_\Gamma \arrow[r, "f"] \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[d, "{[\ ]_\Delta}"] \\
\tau_\Gamma/\equiv_\Gamma \arrow[r, "h"] & \tau_\Delta/\equiv_\Delta
\end{tikzcd}$$

commutes, and

ii) $f[R_\Gamma] \subseteq R_\Delta$.

Proof

We have assumed previously that the number of operators of each arity in O_Γ and O_Δ are the same. Since $\mathscr{A}$ and $\mathscr{B}$ are isomorphic, there is a 1-1 arity preserving correspondence between O_Γ and O_Δ. Hence for notational convenience and without loss of generality, we will assume $O_\Gamma = O_\Delta$ and the correspondence is given by identity.

We first define an $f_1\colon \tau_\Gamma \to \tau_\Delta$ satisfying i) only. For $a \in G_\Gamma$, let $f_1(a) = y$ where y is any term in τ_Δ such that $[y]_\Delta = h([a]_\Gamma)$. f_1 extends uniquely to a homomorphism $\tau_\Gamma \to \tau_\Delta$, by taking $f_1(\theta x_1 \ldots x_m) = \theta f_1(x_1) \ldots f_1(x_m)$. Then f_1 satisfies i), since for $a \in G$, $[f_1(a)]_\Delta = h([a]_\Gamma)$ by definition, and proceeding by structural induction,

$$
\begin{aligned}
[f_1(\theta x_1 \ldots x_m)]_\Delta &= \theta [f_1(x_1)]_\Delta \ \ldots\ [f_1(x_m)]_\Delta \\
&= \theta h([x_1]_\Gamma) \ \ldots\ h([x_m]_\Gamma) \\
&= h([\theta x_1 \ldots x_m]_\Gamma),
\end{aligned}
$$

since $[\ ]_\Delta \circ f_1$ and $h \circ [\ ]_\Gamma$ are homomorphisms.

Let us take $f(x) = f_1(x)$ for $x \notin R_\Gamma$. We need only to find, for every $x \in R_\Gamma$, a $y \in R_\Delta$ such that $y \equiv_\Delta f_1(x)$; then we can take $f(x) = y$, and ii) will hold, but f will satisfy i), since $\forall x\ f(x) \equiv_\Delta f_1(x) \to \forall x\ [f(x)]_\Delta = [f_1(x)]_\Delta = h([x]_\Gamma)$.

Case 1: x appears in an axiom $x \equiv y \in \Gamma$, and neither x nor y are generator symbols.

Observe that since f_1 satisfies i), $x \equiv_\Gamma y \leftrightarrow f_1(x) \equiv_\Delta f_1(y)$. Let $f_1(x) \overset{*}{\to} f_1(y)$ be a proof of $f_1(x) \equiv_\Delta f_1(y)$. If the entire tree is ever replaced, then we have $f_1(x) \overset{*}{\to} z \to w \overset{*}{\to} f_1(y)$ where $z \equiv w$ is an axiom of Δ. Then $z \in R_\Delta$ hence we can take $f(x) = z$, and we are done. Otherwise, since neither x nor y are generator symbols, by construction of f_1 we have

$f_1(x) = \theta f_1(x_1)\ldots f_1(x_m)$, $f_1(y) = \theta f_1(y_1)\ldots f_1(y_m)$, and $f_1(x_1) \equiv_\Delta f_1(y_1)$,..., $f_1(x_m) \equiv_\Delta f_1(y_m)$, where $x = \theta x_1 \ldots x_m$ and $y = \theta y_1 \ldots y_m$. But then $x_1 \equiv_\Gamma y_1$,..., $x_m \equiv_\Gamma y_m$, contradicting the fact that Γ was reduced.

<u>Case 2</u>: $a \in G_\Gamma$ and a occurs in an axiom $a \equiv x$, where $x \notin G_\Gamma$.

Since Γ is reduced, x must contain occurrences of a. Then $f_1(a) \equiv_\Delta f_1(x)$ and $f_1(a) \precneqq f_1(x)$. Let y be a $\prec$-minimal element of $[f_1(a)]_\Delta$, and let $w = f_1(x)[f_1(a) \setminus y]$. Then $w \overset{*}{\rightarrow} y$ and $y \precneqq w$. In a proof $w \overset{*}{\rightarrow} y$, either an ancestor of some occurrence of y in w is the root of a transformation, in which case $y \equiv_\Delta u \in R_\Delta$ as above; or not, in which case y is congruent to a proper subterm of itself, contradicting the assumption of $\prec$-minimality.

<u>Case 3</u>: x is a proper subterm of a term y appearing in an axiom $y \equiv z$.

By cases 1 and 2, we have $f(y) \in R_\Delta$. But then $f_1(y) \equiv_\Delta f(y)$ and $f_1(x) \prec f_1(y)$. In a proof $f_1(y) \overset{*}{\rightarrow} f(y)$, either an ancestor of $f_1(x)$ is the root of a transformation or not. If so, $f_1(x)$ is congruent to a subterm of an axiom of Δ, if not, $f_1(x)$ is congruent to a subterm of $f(y)$. But in either case, $\exists w \in R_\Delta$ $f_1(x) \equiv_\Delta w$, hence we can take $f(x) = w$.

Case 4: $a \in G_\Gamma$ and a occurs only in the axiom $a \equiv a$. Then $f_1(a)$ must be in G_Δ, otherwise $\exists x_1 \ldots x_n \in \tau_\Gamma$ such that $f_1(a) \equiv_\Delta \bullet f_1(x_1) \ldots f_1(x_n)$, since h is an isomorphism; but then $a \equiv_\Gamma \bullet x_1 \ldots x_n$, which is impossible if a occurs only in $a \equiv a$. Thus take $f(a) = f_1(a) \in R_\Delta$. ■

Theorem 27

Let $\mathscr{A}, \mathscr{B}$ be isomorphic via h, and let Γ and Δ be reduced. Then r_Γ and r_Δ are isomorphic via h.

Proof

Using the lemma, form f and g such that

$$\begin{tikzcd}
\tau_\Gamma \arrow[r, "f"] \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[d, "{[\ ]_\Delta}"] \\
\mathscr{A} \arrow[r, "h"] & \mathscr{B}
\end{tikzcd}$$

and

$$\begin{tikzcd}
\tau_\Gamma \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[l, "g"'] \arrow[d, "{[\ ]_\Delta}"] \\
\mathscr{A} & \mathscr{B} \arrow[l, "h^{-1}"']
\end{tikzcd}$$

commute, and $f[R_\Gamma] \subseteq R_\Delta$, $g[R_\Delta] \subseteq R_\Gamma$.

Now $x \in R_\Gamma \rightarrow h([x]_\Gamma) = [f(x)]_\Delta \in r_\Delta$, since $f(x) \in R_\Delta$; hence $h[r_\Gamma] \subseteq r_\Delta$. Similarly, using g, $h^{-1}[r_\Delta] \subseteq r_\Gamma$. But then $h[r_\Gamma] = r_\Delta$, since h is an isomorphism. ■

Theorem 28

Let r_Γ and r_Δ be isomorphic via h. Then h extends to an isomorphism between $\mathscr{A}$ and $\mathscr{B}$.

Proof

Define f: $\tau_\Gamma \to \tau_\Delta$ by taking f(a) = y where $y \in R_\Delta$ such that $[y]_\Delta = h([a]_\Gamma)$ for $a \in G_\Gamma$, and extend f to the unique homomorphism $\tau_\Gamma \to \tau_\Delta$, as in the previous proof.

Claim

$$\begin{tikzcd}
R_\Gamma \arrow[r, "f \restriction R\Gamma"] \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[d, "{[\ ]_\Delta}"] \\
r_\Gamma \arrow[r, "h"] & r_\Delta
\end{tikzcd}$$

commutes.

Proof of claim

For $a \in G_\Gamma$, $h([a]_\Gamma) = [f(a)]_\Delta$ by definition, and for $\theta x_1 \ldots x_m \in R_\Gamma$, we have $x_1, \ldots, x_m \in R_\Gamma$, hence by structural induction,

$$\begin{aligned}
h([\theta x_1 \ldots x_m]_\Gamma) &= \theta h([x_1]_\Gamma) \ \ldots\ h([x_m]_\Gamma) \\
&= \theta [f(x_1)]_\Delta \ldots [f(x_m)]_\Delta \\
&= [f(\theta x_1 \ldots x_m)]_\Delta,
\end{aligned}$$

and the claim is verified.

We wish to extend h to $\hat{h}$ on domain $\mathscr{A}$ by taking $\hat{h}([x]_\Gamma) = [f(x)]_\Delta$, but first we must show that $[\ ]_\Delta \circ f$ is well-defined on $\equiv_\Gamma$-congruence classes, i.e. $x \equiv_\Gamma y \to [f(x)]_\Delta = [f(y)]_\Delta$, so that $\hat{h}$ will be well-defined.

For $x, y \in \tau_\Gamma$, take $x \simeq y$ iff $[f(x)]_\Delta = [f(y)]_\Delta$. Since $[\ ]_\Delta \circ f$ is a homomorphism, $\simeq$ is a congruence relation on τ_Γ. By the above claim, we have for $x, y \in R_\Gamma$

$$\begin{aligned} x \equiv_\Gamma y &\text{ iff } [x]_\Gamma = [y]_\Gamma \\ &\text{ iff } h([x]_\Gamma) = h([y]_\Gamma) \\ &\text{ iff } [f(x)]_\Delta = [f(y)]_\Delta \\ &\text{ iff } x \simeq y. \end{aligned}$$

Since $\equiv_\Gamma$ is defined to be the smallest (most general) congruence satisfying the axioms of Γ, and $\simeq$ satisfies the axioms of Γ, it follows that $\equiv_\Gamma$ is a refinement of $\simeq$. Hence $\forall x,y \in \tau_\Gamma$, $x \equiv_\Gamma y \rightarrow x \simeq y \rightarrow [f(x)]_\Delta = [f(y)]_\Delta$, as was to be shown. Now we have that

$$\begin{tikzcd} \tau_\Gamma \arrow[r, "f"] \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[d, "{[\ ]_\Delta}"] \\ \mathscr{A} \arrow[r, "\hat{h}"] & \mathscr{B} \end{tikzcd}$$

commutes, and $\hat{h} \upharpoonright r_\Gamma = h$. We can also form g such that

$$\begin{tikzcd} \tau_\Gamma \arrow[d, "{[\ ]_\Gamma}"'] & \tau_\Delta \arrow[l, "g"'] \arrow[d, "{[\ ]_\Delta}"] \\ \mathscr{A} & \mathscr{B} \arrow[l, "\widehat{h^{-1}}"'] \end{tikzcd}$$

commutes, and $\widehat{h^{-1}} \upharpoonright r_\Delta = h^{-1}$. But now $\forall x \in \tau_\Gamma$, $g \circ f(x) \equiv_\Gamma x$, by structural induction: certainly for $a \in G_\Gamma$,

$$\begin{aligned} [g \circ f(a)]_\Gamma &= \widehat{h^{-1}}([f(a)]_\Delta) \\ &= \widehat{h^{-1}}(h([a]_\Gamma)) \\ &= h^{-1}(h([a]_\Gamma)) \\ &= [a]_\Gamma, \text{ and then} \end{aligned}$$

$$\begin{aligned}[g\circ f(\theta x_1 \ldots x_m)]_\Gamma &= \theta[g\circ f(x_1)]_\Gamma \ldots [g\circ f(x_m)]_\Gamma \\ &= \theta[x_1]_\Gamma \ldots [x_m]_\Gamma \\ &= [\theta x_1 \ldots x_m]_\Gamma .\end{aligned}$$

Similarly $\forall x \in \tau_\Delta$, $f\circ g(x) \equiv_\Delta x$. But this says that $\hat{h}$ and $\widehat{h^{-1}}$ are inverses, since $\forall x \in \tau_\Gamma$

$$\begin{aligned}\widehat{h^{-1}}\ (\hat{h}([x]_\Gamma)) &= \widehat{h^{-1}}\ (\ [f(x)]_\Delta) \\ &= [g\circ f(x)]_\Gamma \\ &= [x]_\Gamma\end{aligned}$$

and similarly $\forall y \in \tau_\Delta$, $\hat{h}\ (\widehat{h^{-1}}([y]_\Delta)) = [y]_\Delta$. Thus $\mathscr{A}$ and $\mathscr{B}$ are isomorphic via $\hat{h}$. ■

Corollary 29

ISOM $\leq_p^m$ isomorphism of labeled directed graphs.

Proof

Given an instance of ISOM $\langle\Gamma,\Delta\rangle$, reduce Γ and Δ to get Γ^* and Δ^*, and then form the graphs $\chi_\Gamma{}^*$ and $\chi_\Delta{}^*$. By Lemmas 10 and 25, this can be done in polynomial time. Then by Theorems 27 and 28 and Lemma 24,

$$\mathscr{A} \cong \mathscr{B} \quad \text{iff} \quad r_\Gamma{}^* \cong r_\Delta{}^* \quad \text{iff} \quad \chi_\Gamma{}^* \cong \chi_\Delta{}^* . \qquad ■$$

Again, it is rather remarkable that isomorphism of possibly infinite structures should reduce to that of finite ones, unlike previous results in this area.[15,16]

The $\equiv_p^m$ equivalence of graph isomorphism and ISOM should be of great interest to those who believe graph isomorphism is NP complete. It is clear that graph isomorphism is in NP, but the form of the problem is so restricted (i.e., two graphs with the same numbers of vertices of each in- and out-degree) that standard reduction techniques fail to show even that it is hard for P. However, there are no such restrictions on the form of instances of ISOM, and it is quite trivial to show ISOM is P-hard:

Theorem 30

ISOM is $\leq^{m}_{\log}$-hard for P.

Proof

Use Γ' of Theorem 5 and a presentation of the trivial algebra. ■

Thus it would surely be easier to show that ISOM, the more general of the two problems, is NP-hard.

6. Open questions

The following is a list of interesting and relevant open problems:

1) Prove the conjecture at the end of §4, thus generalizing the results of [7]. The techniques of §4 do not immediately generalize, nevertheless it is likely that this problem will be easier than 2) or 3).

2) Show graph isomorphism is $\leq_p^m$-complete for NP, by showing ISOM is $\leq_p^m$-hard for NP.

3) In lieu of 2), improve the $\leq_p^m$ reduction from ISOM to graph isomorphism to a $\leq_{\log}^m$ reduction, thereby showing graph isomorphism is $\leq_{\log}^m$-hard for P. The $\leq_p^m$ reduction given involves an algorithm for a problem complete for P, thus it will not be altered easily to a $\leq_{\log}^m$ reduction (unless LOGSPACE = P).

Acknowledgment

I would like to thank Erik Schmidt for turning me on to finite tree automata.

References

1) Lipton, R.J. and Y. Zalcstein, Word Problems Solvable in Logspace, Technical Report #48, Department of Computer Science, SUNY at Stony Brook, 1975.

2) Cardoza, E., R. Lipton, and A.R. Meyer, "Exponential Space Complete Problems for Petri Nets and Commutative Semigroups: Preliminary Report," Proc. 8th ACM Symposium on Theory of Computing, May 1976, 50-54.

3) Ladner, R.E., "The Circuit Value Problem is Logspace Complete for P," SIGACT News 7:1, January 1975.

4) Jones, N.D., and W.T. Laaser, "Complete Problems for Deterministic Polynomial Time," Proc. 6th ACM Symposium on Theory of Computing, April 1974, 40-46.

5) Cook, S.A., "The Complexity of Theorem Proving Procedures," Proc. 3rd ACM Symposium on Theory of Computing, 1971, 151-158.

6) Kozen, D., "On Parallelism in Turing Machines," Proc. 17th IEEE Symposium on Foundations of Computer Science, October 1976, 89-97.

7) Meyer, A.R., and L.J. Stockmeyer, "Word Problems Requiring Exponential Time," Proc. 5th ACM Symposium on Theory of Computing, April 1973, 1-9.

8) Ladner, R.E., "Polynomial Time Reducibility," Proc. 5th ACM Symposium on Theory of Computing, April 1973, 122-129.

9) Hopcroft, J.E., and J.D. Ullman, Formal Languages and Their Relation to Automata, Addison-Wesley, Reading, MA, 1969, p. 29.

10) Thatcher, J.W., "Tree Automata: An Informal Survey," Currents in the Theory of Computing, ed. Aho, Prentice-Hall, Englewood Cliffs, NJ, 1973.

11) Engelfriet, J., "Tree Automata and Tree Grammars," Department of Computer Science Report DAIMI FN-10, University of Aarhus, Denmark, April 1975.

12) Thatcher, J.W., and J.B. Wright, "Generalized Finite Automata Theory with an Application to a Decision Problem of Second Order Logic," Math. Syst. Th. 2, 1968.

13) Doner, J., "Decidability of the Weak Second Order Theory of Two Successors," Notices Am. Math. Soc. 12, 1965.

14) Rabin, M.O., "Decidability of Second Order Theories and Automata on Infinite Trees," Trans. Am. Math. Soc. 141, 1969, 1-35.

15) Booth, K.S., "Problems Polynomially Equivalent to Graph Isomorphism," Carnegie-Mellon Symposium on New Directions and Recent Results in Algorithms and Complexity, April 1976.

16) Aho, A.V., J.E. Hopcroft, and J.D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1975, p. 402.