One-Nonterminal Conjunctive Grammars over a Unary Alphabet

Artur Jeż[1,⋆] and Alexander Okhotin[2,3,⋆⋆]

[1] Institute of Computer Science, University of Wrocław, Poland
aje@ii.uni.wroc.pl
[2] Academy of Finland
[3] Department of Mathematics, University of Turku, Finland
alexander.okhotin@utu.fi

Abstract. Conjunctive grammars over an alphabet $\Sigma = \{a\}$ are studied, with the focus on the special case with a unique nonterminal symbol. Such a grammar is equivalent to an equation $X = \varphi(X)$ over sets of natural numbers, using union, intersection and addition. It is shown that every grammar with multiple nonterminals can be encoded into a grammar with a single nonterminal, with a slight modification of the language. Based on this construction, the compressed membership problem for one-nonterminal conjunctive grammars over $\{a\}$ is proved to be EXPTIME-complete, while the equivalence, finiteness and emptiness problems for these grammars are shown to be undecidable.

1 Introduction

Conjunctive grammars are an extension of the context-free grammars with an explicit intersection operation, introduced by Okhotin [9]. These grammars are characterized by language equations of the form

$$\begin{cases} X_1 = \varphi_1(X_1, \ldots, X_n) \\ \quad\vdots \\ X_n = \varphi_n(X_1, \ldots, X_n) \end{cases} \tag{*}$$

in which the right-hand sides φ_i may contain union, intersection and concatenation of languages, as well as singleton constants; in context-free grammars, intersection is not allowed. Despite their higher expressive power compared to context-free grammars, conjunctive grammars still possess efficient parsing algorithms and can be parsed in $DTIME(n^3) \cap DSPACE(n)$ [10], which makes them potentially useful for practical use.

Consider the special case of a *unary alphabet* $\Sigma = \{a\}$. All unary context-free languages are known to be regular. The question of whether conjunctive grammars can generate any non-regular unary languages has been an open problem

⋆ Supported by MNiSW grants N206 024 31/3826 2006–2009 and N N206 259035 2008–2010.

⋆⋆ Supported by the Academy of Finland under grant 118540.

A. Frid et al. (Eds.): CSR 2009, LNCS 5675, pp. 191–202, 2009.

for some years [10], until recently solved by Jeż [2], who constructed a conjunctive grammar for the language $\{a^{4^n} | n \geqslant 0\}$. This result was followed by a general theorem due to the authors [3], which asserts representability of a large class of unary languages by conjunctive grammars.

As strings over a one-letter alphabet can be regarded as natural numbers, language equations (*) corresponding to unary conjunctive grammars accordingly become *equations over sets of numbers*, in which concatenation of languages is represented by element-wise addition of sets defined as $S + T = \{m + n \mid m \in S, n \in T\}$. Such equations can be regarded as a generalization of *integer expressions*, introduced by Stockmeyer and Meyer [13] and further studied by McKenzie and Wagner [8]. In particular, the membership problem for integer expressions with the operations of union and addition is NP-complete [13], and allowing intersection makes it PSPACE-complete [8]. For equations over sets of numbers of the form (*) and using the same operations, the membership problem is NP-complete without intersection [1] and EXPTIME-complete with intersection [4]. The latter result was established by constructing a conjunctive grammar that generates unary notation of all numbers in a certain EXPTIME-complete set. The results on unary conjunctive grammars have further led to realising the computational completeness of more general equations over sets of natural numbers [5].

This paper investigates a special case of unary conjunctive grammars containing a unique nonterminal symbol. This subfamily has recently been studied by Okhotin and Rondogiannis [11], who constructed a one-nonterminal conjunctive grammar generating a non-periodic language, as well as presented two classes of conjunctive languages that are not representable by any grammars with a single nonterminal. The goal of this paper is to generalize the example of a nonregular language given by Okhotin and Rondogiannis [11] to a general representability theorem, and to explore its implications on decidability and complexity.

The main result of the paper, established in Section 3, is that for every unary conjunctive grammar the languages generated by all of its nonterminal symbols can be encoded together in a single unary language generated by a one-nonterminal conjunctive grammar. This construction is used in Section 4 to show that the EXPTIME-completeness result for unary conjunctive grammars [4] holds already for one-nonterminal grammars. In Section 5, the new theorem is used to demonstrate that equivalence, finiteness and co-finiteness problems are already undecidable for one-nonterminal unary conjunctive grammars. At the same time, equality to a constant language is shown to be decidable for a large class of constants, in contrast to the multiple-nonterminal case where it is undecidable for every fixed conjunctive constant [3].

2 Conjunctive Grammars and Systems of Equations

Conjunctive grammars generalize context-free grammars by allowing an explicit conjunction operation in the rules.

Definition 1 (Okhotin [9]). *A conjunctive grammar is a quadruple $G = (\Sigma, N, P, S)$, in which Σ and N are disjoint finite non-empty sets of terminal and nonterminal symbols respectively; P is a finite set of grammar rules, each of the form*

$$A \to \alpha_1 \& \ldots \& \alpha_n \quad (\textit{with } A \in N,\ n \geqslant 1 \textit{ and } \alpha_1, \ldots, \alpha_n \in (\Sigma \cup N)^*),$$

while $S \in N$ is a nonterminal designated as the start symbol.

The semantics of conjunctive grammars may be defined either by term rewriting [9], or, equivalently, by a system of language equations. According to the definition by language equations, conjunction is interpreted as intersection of languages as follows.

Definition 2 ([10]). *Let $G = (\Sigma, N, P, S)$ be a conjunctive grammar. The associated system of language equations is the following system in variables N:*

$$A = \bigcup_{A \to \alpha_1 \& \ldots \& \alpha_n \in P} \ \bigcap_{i=1}^{n} \alpha_i \quad (\textit{for all } A \in N)$$

Let $(\ldots, L_A, \ldots)$ be its least solution and denote $L_G(A) := L_A$ for each $A \in N$. Define $L(G) := L_G(S)$.

If the right-hand side of the equation is taken as an operator on vectors of languages, then the least fixed point of this operator is the least solution of the system with respect to componentwise inclusion.

An equivalent definition of conjunctive grammars is given via *term rewriting*, which generalizes the string rewriting used by Chomsky to define context-free grammars.

Definition 3 ([9]). *Given a conjunctive grammar G, consider terms over concatenation and conjunction with symbols from $\Sigma \cup N$ as atomic terms. The relation $\Longrightarrow$ of immediate derivability on the set of terms is defined as follows:*

- *Using a rule $A \to \alpha_1 \& \ldots \& \alpha_n$, a subterm $A \in N$ of any term $\varphi(A)$ can be rewritten as $\varphi(A) \Longrightarrow \varphi(\alpha_1 \& \ldots \& \alpha_n)$.*
- *A conjunction of several identical strings can be rewritten by one such string: $\varphi(w \& \ldots \& w) \Longrightarrow \varphi(w)$, for every $w \in \Sigma^*$.*

The language generated by a term φ is $L_G(\varphi) = \{w \mid w \in \Sigma^*,\ \varphi \Longrightarrow^* w\}$. The language generated by the grammar is $L(G) = L_G(S) = \{w | w \in \Sigma^*,\ S \Longrightarrow^* w\}$.

Consider the following grammar generating the language $\{a^{4^n} \mid n \geqslant 0\}$, which was the first example of a unary conjunctive grammar representing a non-regular language.

Example 1 (Jeż [2]). The conjunctive grammar

$$\begin{aligned}
A_1 &\to A_1 A_3 \& A_2 A_2 \mid a \\
A_2 &\to A_1 A_1 \& A_2 A_6 \mid aa \\
A_3 &\to A_1 A_2 \& A_6 A_6 \mid aaa \\
A_6 &\to A_1 A_2 \& A_3 A_3
\end{aligned}$$

with the start symbol A_1 generates the language $L(G) = \{a^{4^n} \mid n \geqslant 0\}$. In particular, $L_G(A_i) = \{a^{i \cdot 4^n} \mid n \geqslant 0\}$ for $i = 1, 2, 3, 6$.

The construction in Example 1 essentially uses all four nonterminal symbols, and there seems to be no apparent way to replicate it using a single nonterminal. However, this was achieved in the following example:

Example 2 (Okhotin, Rondogiannis [11]). The conjunctive grammar

$$S \to a^{22}SS\&a^{11}SS \mid a^9SS\&aSS \mid a^7SS\&a^{12}SS \mid a^{13}SS\&a^{14}SS \mid a^{56} \mid a^{113} \mid a^{181}$$

generates the language $\{a^{4^n-8} \mid n \geqslant 3\} \cup \{a^{2 \cdot 4^n-15} \mid n \geqslant 3\} \cup \{a^{3 \cdot 4^n-11} \mid n \geqslant 3\}$ $\cup \{a^{6 \cdot 4^n-9} \mid n \geqslant 3\}$.

This grammar is actually derived from Example 1, and the language it generates encodes the languages of all four nonterminals in Example 1. Each of the four components in the generated language represents one of the nonterminals in Example 1 with a certain *offset* (8, 15, 11 and 9).

Note that the set from Example 2 is exponentially growing. At the same time, it has been proved that if a set grows faster than exponentially (for example, $\{n! \mid n \geqslant 1\}$), then it is not representable by univariate equations:

Proposition 1 (Okhotin, Rondogiannis [11]). *Let $S = \{n_1, n_2, \ldots, n_i, \ldots\}$ with $0 \leqslant n_1 < n_2 < \ldots < n_i < \ldots$ be an infinite set of numbers, for which $\liminf_{i \to \infty} \frac{n_i}{n_{i+1}} = 0$. Then S is not the least solution of any equation $X = \varphi(X)$.*

On the other hand, it is known that unary conjunctive grammars can generate a set that grows faster than any given recursive set:

Proposition 2 (Jeż, Okhotin [3]). *For every recursively enumerable set of natural numbers S there exists a conjunctive grammar G over an alphabet $\{a\}$, such that the set $\widehat{S} = \{n \mid a^n \in L(G)\}$ grows faster than S, in the sense that the n-th smallest number of $\widehat{S}$ is greater than the n-th smallest number of S, for all $n \geqslant 1$.*

Thus one-nonterminal conjunctive grammars are weaker in power than arbitrary unary conjunctive grammars. However, even though one-nonterminal conjunctive grammars cannot generate *all* unary conjunctive languages, it will now be demonstrated that they can represent a certain encoding of every conjunctive language.

3 One-Nonterminal Conjunctive Grammars

The goal is to simulate an arbitrary conjunctive grammar over $\{a\}$ by a conjunctive grammar with a single nonterminal symbol. The construction formalizes and elaborates the intuitive idea of Example 2, making it provably work for any grammar.

The first step towards the construction is a small refinement of the known normal form for conjunctive grammars. It is known that every conjunctive language

$L \subseteq \Sigma^+$ over any alphabet Σ can be generated by a conjunctive grammar in the *binary normal form* [9,10], with all rules of the form $A \to B_1C_1\& \ldots \& B_nC_n$ with $n \geqslant 1$ or $A \to a$. The following stronger form is required by the below construction.

Lemma 1. *For every conjunctive grammar $G = (\Sigma, N, P, S)$ with $\varepsilon \notin L(G)$ there exists a conjunctive grammar $G' = (\Sigma, N', P', S')$ generating the same language, in which every rule is of the form $A \to a$ with $a \in \Sigma$, or*

$$A \to B_1C_1\& \ldots \& B_nC_n \quad (\textit{with } n \geqslant 2),$$

in which the sets $\{B_1, C_1\}$, ..., $\{B_n, C_n\}$ are pairwise disjoint.

A grammar in the binary normal form can be easily converted this stronger form by making multiple copies of each nonterminal.

The next theorem is the core result of this paper.

Theorem 1. *For every unary conjunctive grammar $G = (\{a\}, \{A_1, \ldots, A_m\}, P, A_1)$ of the form given in Lemma 1 there exist numbers $0 < d_1 < \ldots < d_m < p$ depending only on m and a conjunctive grammar $G' = (\{a\}, \{B\}, P', B)$ generating the language $L(G') = \{a^{np-d_i} \mid 1 \leqslant i \leqslant m,\ a^n \in L_G(A_i)\}$.*

Accordingly, the corresponding equation $X = \varphi(X)$ over sets of natural numbers has a unique solution $S = \bigcup_{i=1}^m S_i$, where $S_i = \{np - d_i \mid a^n \in L_G(A_i)\}$.

Let $p = 4^{m+2}$ and let $d_i = \frac{p}{4} + 4^i$ for every nonterminal A_i. For every number $t \in \{0, \ldots, p\}$, the set $\{np - t \mid n \geqslant 0\}$ is called *track number* t. The goal of the construction is to represent each set S_i in the track d_i. The rest of the tracks should be empty.

For every rule $A_i \to A_{j_1}A_{k_1}\& \ldots \& A_{j_\ell}A_{k_\ell}$ in G, the new grammar G' contains the rule

$$B \to a^{d_{j_1}+d_{k_1}-d_i}BB\& \ldots \& a^{d_{j_\ell}+d_{k_\ell}-d_i}BB, \tag{1}$$

and for every rule $A_i \to a$ in G, let G' have a rule

$$B \to a^{p-d_i}.$$

The proof of correctness of the construction will be done in terms of equations over sets of numbers. The task is to prove that the unique solution of the equation corresponding to G' is $S = \bigcup_i S_i$, where $S_i = \{np - d_i \mid a^n \in L_G(A_i)\}$.

Each time X appears in the right-hand side of the equation, it is used in the context of an expression $\psi(X) = X + X + (d_i + d_j - d_k)$. The proof of the theorem is based upon the following property of these expressions.

Lemma 2. *Let $i, j, k, \ell \in \{1, \ldots, m\}$ with $\{i, j\} \cap \{k, \ell\} = \varnothing$. Then*

$$(S + S + d_i + d_j) \cap (S + S + d_k + d_\ell) = (S_i + S_j + d_i + d_j) \cap (S_k + S_\ell + d_k + d_\ell).$$

Proof. As addition is distributive over union and union is distributive over intersection,

$$(S + S + d_i + d_j) \cap (S + S + d_k + d_\ell) = \\ = \bigcup_{i',j',k',\ell'} (S_{i'} + S_{j'} + d_i + d_j) \cap (S_{k'} + S_{\ell'} + d_k + d_\ell)$$

It is sufficient to prove that if $\{i', j'\} \neq \{i, j\}$ or $\{k', \ell'\} \neq \{k, \ell\}$, then the intersection is empty. Consider any such intersection

$$(S_{i'} + S_{j'} + d_i + d_j) \cap (S_{k'} + S_{\ell'} + d_k + d_\ell) = \\ \big(\{np \mid a^n \in L(A_{i'})\} - d_{i'} + \{np \mid a^n \in L(A_{j'})\} - d_{j'} + d_i + d_j\big) \cap \\ \big(\{np \mid a^n \in L(A_{k'})\} - d_{k'} + \{np \mid a^n \in L(A_{\ell'})\} - d_{\ell'} + d_k + d_\ell\big),$$

and suppose it contains any number, which must consequently be equal to $d_i + d_j - d_{i'} - d_{j'}$ modulo p and to $d_k + d_\ell - d_{k'} - d_{\ell'}$ modulo p. As each d_t satisfies $\frac{p}{4} < d_t \leqslant \frac{p}{2}$, both offsets are strictly between $-\frac{p}{2}$ and $\frac{p}{2}$, and therefore they must be equal to each other:

$$d_i + d_j - d_{i'} - d_{j'} = d_k + d_\ell - d_{k'} - d_{\ell'}.$$

Equivalently, $d_i + d_j + d_{k'} + d_{\ell'} = d_k + d_\ell + d_{i'} + d_{j'}$, and since each d_t is defined as $\frac{p}{4} + 4^t$, this holds if and only if

$$4^i + 4^j + 4^{k'} + 4^{\ell'} = 4^k + 4^\ell + 4^{i'} + 4^{j'}.$$

Consider the largest of these eight numbers, let its value be d. Without loss of generality, assume that it is on the left-hand side. Then the left-hand side is greater than d. On the other hand, if no number on the right-hand side is d, then the sum is at most $4 \cdot \frac{d}{4} = d$. Thus at least one number on the right-hand side must be equal to d as well. Removing those two numbers and giving the same argument for the sum of 3, 2 and 1 summands yields that

$$\{d_i, d_j, d_{k'}, d_{\ell'}\} = \{d_k, d_\ell, d_{i'}, d_{j'}\}.$$

Then, by the assumption that $\{i, j\} \cap \{k, \ell\} = \varnothing$,

$$\{d_i, d_j\} = \{d_{i'}, d_{j'}\} \quad \text{and} \quad \{d_{k'}, d_{\ell'}\} = \{d_k, d_\ell\},$$

and since the addition is commutative,

$$i = i', \quad j = j', \quad k = k' \quad \text{and} \quad \ell = \ell'.$$

This completes the proof of the lemma. □

With this property established, it can be verified that every rule for every A_i in G is correctly simulated by the corresponding rule of G', and that the data from different tracks is never mixed.

Consider the equation $X = \varphi(X)$ over sets of numbers corresponding to G'. Every "long" rule $A \to \mathscr{A}$ in G, where $\mathscr{A} = A_{j_1}A_{k_1}\& \ldots \& A_{j_\ell}A_{k_\ell}$, is represented in the new grammar by a rule (1), which contributes the following subexpression to φ

$$\varphi_{i,\mathscr{A}}(S) = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S + S = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t}$$

Altogether, the equation $X = \varphi(X)$ takes the following form:

$$X = \bigcup_{A_i \to \mathscr{A} \in P} \varphi_{i,\mathscr{A}}(X) \cup \bigcup_{A_i \to a \in P} \{p - d_i\}$$

Now the task is to prove that the unique solution of this equation is $S = \bigcup_i S_i$, where $S_i = \{np - d_i \mid a^n \in L_G(A_i)\}$.

Consider each "long" rule $A_i \to \mathscr{A}$ with $\mathscr{A} = A_{j_1}A_{k_1}\& \ldots \& A_{j_t}A_{k_t}$. Then

$$\varphi_{i,\mathscr{A}}(S) = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S + S = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t}$$

by Lemma 2, and it is easy to calculate that

$$\bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t} = \{np - d_i \mid a^n \in L_G(\mathscr{A})\}.$$

Similarly, for a "short" rule $A_i \to a$, $\{p - d_i\} = \{np - d_i \mid a^n \in L_G(\{a\})\}$, and altogether,

$$\varphi(S) = \bigcup_i \bigcup_{A_i \to \mathscr{B} \in P} \{np - d_i \mid a^n \in L_G(\mathscr{B})\} = \bigcup_i \{np - d_i \mid a^n \in L_G(A_i)\} = S,$$

where the inner equality is due to the fact that $(\ldots, L_G(A_i), \ldots)$ is the solution of the system of language equations associated to G. This completes the proof of Theorem 1.

For an example of this transformation, consider the four-nonterminal grammar from Example 1. It satisfies the condition in Lemma 1, but it is not precisely in the binary normal form, as it contains rules $A_2 \to aa$ and $A_3 \to aaa$. However, these rules do not affect the general construction, and one can extend the transformation of Theorem 1 to this grammar as follows:

Example 3. For the grammar in Example 1, the constants are $m = 4$, $p = 4^{m+2} = 4^6 = 4096$, $d_1 = 1028$, $d_2 = 1040$, $d_3 = 1088$ and $d_4 = 1280$, and the transformation yields the grammar

$$\begin{aligned} S \to a^{1088}SS\&a^{1052}SS \mid a^{1016}SS\&a^{1280}SS \mid a^{980}SS\&a^{1472}SS \mid \\ a^{788}SS\&a^{896}SS \mid a^{3068} \mid a^{7152} \mid a^{11200} \end{aligned}$$

generating the language

$$\{a^{4^n-1028} | n \geqslant 6\} \cup \{a^{2\cdot 4^n-1040} | n \geqslant 6\} \cup \{a^{3\cdot 4^n-1088} | n \geqslant 6\} \cup \{a^{6\cdot 4^n-1280} | n \geqslant 6\}.$$

4 Complexity of the Membership Problem

The *(general) membership problem* for a family of grammars is stated as follows: "Given a string w and a grammar G, determine whether $w \in L(G)$". The membership problem is P-complete both for context-free and conjunctive grammars.

A variant of this problem, which received considerable attention in the recent years, is the *compressed membership problem* [12], where the string w is given as a context-free grammar G_w generating $\{w\}$. This problem is PSPACE-complete for context-free grammars [7,12] and EXPTIME-complete for conjunctive grammars [4]. For unary languages, the compressed representation of a^n is its binary notation of length $\Theta(\log n)$. In this case the problem is NP-complete for context-free grammars [1,12], but still EXPTIME-complete for conjunctive grammars [4].

Proposition 3 (Jeż, Okhotin [4]). *There exists an EXPTIME-complete set of numbers $S \subseteq \mathbb{N}$, such that the language $L = \{a^n \mid n \in S\}$ is generated by a conjunctive grammar. The compressed membership problem for conjunctive grammars over a unary alphabet is EXPTIME-complete.*

To show that both results still hold for one-nonterminal unary conjunctive grammars, it is sufficient to take the grammar generating L and to transform it according to Theorem 1.

Theorem 2. *There exists an EXPTIME-complete set of numbers $S \subseteq \mathbb{N}$, such that the language $L = \{a^n \mid n \in S\}$ is generated by a one-nonterminal conjunctive grammar. The compressed membership problem for one-nonterminal unary conjunctive grammars is EXPTIME-complete.*

For unary context-free grammars, the compressed membership problem has different complexity depending on the number of nonterminals. For multiple nonterminals it is NP-complete [12]. However, in the one-nonterminal case an efficient algorithm for solving this problem can be obtained using the following property:

Lemma 3. *Let $G = (\{a\}, \{S\}, P, S)$ be a one-nonterminal context-free grammar with m rules and with the longest right-hand side of a rule of length k. Then $L(G)$ is periodic starting from $\widehat{n} \leqslant 2mk^3(2k+1)$ with a period at most $2k^2$.*

First it is shown that from any single pair of rules $S \to S^{k_i}a^{\ell_i}$ and $S \to a^{\ell_j}$ one can deduce that $p = (k_i - 1)\ell_j + \ell_i$ is a period of $L(G)$, which gives an upper bound on the least period. As the language $L(G)$ is over a unary alphabet, a derivation is determined by the number of occurrences of each rule, and it is shown that a derivation exists if and only if those numbers satisfy a certain numerical condition. Using this representation, the derivation of every sufficiently long string can be reduced to obtain a derivation of a shorter string, which has the same length modulo p. Thus the periodicity starts from a short string.

Theorem 3. *The compressed membership problem for one-nonterminal unary CFGs is in NLOGSPACE.*

Table 1. Complexity of general membership problems

	uncompressed	compressed	fully compressed
Context-free			
general case	P-complete	PSPACE-complete [12]	n/a
$\Sigma = \{a\}$, any N	P-complete	NP-complete [1]	NP-complete [1]
$\Sigma = \{a\}$, $N = \{S\}$	**in NLOGSPACE**	**in NLOGSPACE**	**NP-complete**
Conjunctive			
general case	P-complete	EXPTIME-complete [4]	n/a
$\Sigma = \{a\}$, any N	P-complete	EXPTIME-complete [4]	EXPTIME-compl. [4]
$\Sigma = \{a\}$, $N = \{S\}$	in P	**EXPTIME-complete**	EXPTIME-compl. [4]

By Lemma 3, the period of $L(G)$ is small. Then the input string can be replaced with a shorter string of a logarithmic length, and a nondeterministic algorithm can guess its derivation, storing the compressed sentential form in logarithmic space.

Note that the grammars are supplied to the above algorithm uncompressed, that is, rules of the form $S \to a^\ell S^k$ are stored using $\ell + k$ symbols. Consider another variant of the problem, the *fully compressed membership problem*, in which the grammar is compressed as well, and a rule $S \to a^\ell S^k$ is given by binary notations of ℓ and k. This problem is hard already for one nonterminal:

Theorem 4. *The fully compressed membership problem for one-nonterminal unary CFGs is NP-complete.*

As in the NP-hardness result for integer expressions by Stockmeyer and Meyer [13, Thm. 5.1], the proof is by encoding an instance of the *knapsack problem* in a grammar, with the sizes of the objects compressed. The knapsack problem is stated as follows: "Given integers $b_1, \ldots, b_n$ and z in binary notation, determine whether there exist $c_1, \ldots, c_n \in \{0, 1\}$ with $\sum_{i=1}^n b_i c_i = z$". Given an instance of this problem, assume that $z > \max_i b_i$, and let m be the least power of two with $m \geqslant \max\{z, 2^n\} + 2$. Then the grammar G with the following rules is constructed: $S \to S^n$, $S \to a^{m^2+2^{i-1}}$ and $S \to a^{m^2+mb_i+2^{i-1}}$ for $1 \leqslant i \leqslant n$. It can be proved that the string $w = a^{nm^2+zm+(2^n-1)}$ is in $L(G)$ if and only if the numbers $c_1, \ldots, c_n \in \{0, 1\}$ with $\Sigma_{i=1}^n b_i c_i = z$ do exist.

The complexity of different cases is summarized in Table 1.

5 Decision Problems

Having established the complexity of the membership problem, let us now consider some other basic properties of one-nonterminal unary conjunctive grammars. In the case of multiple nonterminals, most properties are undecidable:

Proposition 4 (Jeż, Okhotin [3]). *For every fixed unary conjunctive language $L_0 \subseteq a^*$, the problem of whether a given conjunctive grammar over $\{a\}$ generates the language L_0 is Π_1-complete.*

In contrast, in the case of one-nonterminal grammars, the equality to any fixed ultimately periodic set is clearly decidable: it is sufficient to substitute it into the equation and check whether it is turned into an equality. This approach extends to a larger class of fixed languages:

Theorem 5. *There exists an algorithm, which, given an equation $X = \varphi(X)$ using union, intersection, addition and ultimately periodic constants, and a finite automaton M over an alphabet $\Sigma_k = \{0, 1, \ldots, k-1\}$, determines whether $\{n \mid$ the k-ary notation of n is in $L(M)\}$ is the least solution of the equation.*

In particular, the theorem applies to such languages as in Examples 1–3, which are recognized by finite automata in base-4 notation.

The algorithm works by substituting the set of numbers defined by M into the equation. The value of each subexpression is computed in the form of a finite automaton over Σ_k representing base-k notation. For Boolean operations this is clearly possible, while the addition of sets can be done symbolically on finite automata for their base-k representation according to the following lemma:

Lemma 4. *Let L_1 and L_2 be regular languages over $\Sigma_k = \{0, 1, \ldots, k-1\}$, with $L_1 \cap 0\Sigma_k^* = L_2 \cap 0\Sigma_k^* = \varnothing$. Then the language $\{$the k-ary notation of $n_1 + n_2 \mid$ the k-ary notation of n_i is in $L_i\}$ over Σ_k is (effectively) regular.*

This shows that equality to a fixed language is decidable for one-nonterminal conjunctive grammars for a fairly large class of constants. At the same time, the more general problem of equivalence of two given grammars is undecidable.

Theorem 6. *The equivalence problem for one-nonterminal unary conjunctive grammars is Π_1-complete.*

The equivalence problem for unary conjunctive grammars with multiple nonterminals is Π_1-complete and the proof is by reduction from this problem. Two grammars G_1 and G_2 are combined into G, such that $L(G_1) = L_G(A_1)$ and $L(G_2) = L_G(A_2)$. Let G' be G with A_1 and A_2 exchanged. Once the construction of Theorem 1 is applied to G and G', the two resulting one-nonterminal grammars are equivalent if and only if the original grammars generate the same language.

Theorem 7. *The co-finiteness problem for one-nonterminal unary conjunctive grammars is Σ_1-complete.*

The problem is in Σ_1 by Theorem 5: an algorithm solving this problem can nondeterministically guess an NFA N recognizing a co-finite language and test whether the grammar generates $L(N)$. The hardness is proved by reduction from the non-emptiness problem for unary conjunctive grammars. The transformation of Theorem 1 is applied to a given grammar, resulting in a one-nonterminal grammar G', and then the following p extra rules are added to G':

$$S \to a^{d_1+i}X \& a^{d_2+i}X \quad (0 \leqslant i \leqslant p-1)$$

Table 2. Decision problems for grammars over $\{a\}$

	equiv. to reg. L_0	equivalence	finiteness	co-finiteness
Context-free				
general case	undecidable	undecidable	decidable	undecidable
$\Sigma = \{a\}$, any N	decidable	decidable	decidable	decidable
$\Sigma = \{a\}$, $N = \{S\}$	decidable	decidable	decidable	decidable
Conjunctive				
general case	undecidable	undecidable	undecidable	undecidable
$\Sigma = \{a\}$, any N	Π_1-complete [3]	Π_1-complete [3]	undecidable [3]	$\Sigma_1 \leqslant \cdot \leqslant \Sigma_2$
$\Sigma = \{a\}$, $N = \{S\}$	**decidable**	Π_1**-complete**	Σ_1**-complete**	Σ_1**-complete**

If any string appears on track d_1, these rules will generate all longer strings, thus "spamming" the language to make it co-finite.

Theorem 8. *The finiteness problem for one-nonterminal unary conjunctive grammars is Σ_1-complete.*

The problem is in Σ_1 for the same reason as in Theorem 7. The Σ_1-hardness is proved by reduction from the problem of whether $L(G) \neq a^+$ for a given unary conjunctive grammar G. Once G is transformed to a one-nonterminal grammar G' (as in the previous proof), an additional conjunct is added to each rule as follows: for every rule

$$S \to a^{\ell_1} SS \& \ldots \& a^{\ell_k} SS$$

created in Theorem 7, which generates a subset of $\{a^{np-d_i} \mid n \geqslant 1\}$ for some number i, the final grammar has the rule

$$S \to a^{\ell_1} SS \& \ldots \& a^{\ell_k} SS \& a^{p+d_1-d_i} S.$$

This additional conjunct acts as a filter: if any string in track d_1 is missing, then no strings of any greater length can be generated, making the generated language finite.

The level of undecidability of both finiteness and co-finiteness problems for conjunctive grammars with multiple nonterminals remains open, see the summary in Table 2.

References

1. Huynh, D.T.: Commutative grammars: the complexity of uniform word problems. Information and Control 57(1), 21–39 (1983)
2. Jeż, A.: Conjunctive grammars can generate non-regular unary languages. International Journal of Foundations of Computer Science 19(3), 597–615 (2008)
3. Jeż, A., Okhotin, A.: Conjunctive grammars over a unary alphabet: undecidability and unbounded growth. Theory of Computing Systems (to appear)
4. Jeż, A., Okhotin, A.: Complexity of equations over sets of natural numbers. In: 25th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2008), Bordeaux, France, February 21-23, pp. 373–383 (2008)

5. Jeż, A., Okhotin, A.: On the computational completeness of equations over sets of natural numbers. In: 35th International Colloquium on Automata, Languages and Programming (ICALP 2008), Reykjavik, Iceland, July 7-11, pp. 63–74 (2008)
6. Kunc, M.: What do we know about language equations? In: Harju, T., Karhumäki, J., Lepistö, A. (eds.) DLT 2007. LNCS, vol. 4588, pp. 23–27. Springer, Heidelberg (2007)
7. Lohrey, M.: Word problems and membership problems on compressed words. SIAM Journal on Computing 35(5), 1210–1240 (2006)
8. McKenzie, P., Wagner, K.W.: The complexity of membership problems for circuits over sets of natural numbers. Computational Complexity 16, 211–244 (2007)
9. Okhotin, A.: Conjunctive grammars. Journal of Automata, Languages and Combinatorics 6(4), 519–535 (2001)
10. Okhotin, A.: Nine open problems for conjunctive and Boolean grammars. Bulletin of the EATCS 91, 96–119 (2007)
11. Okhotin, A., Rondogiannis, P.: On the expressive power of univariate equations over sets of natural numbers. In: IFIP Intl. Conf. on Theoretical Computer Science (TCS 2008), Milan, Italy, September 8-10. IFIP, vol. 273, pp. 215–227 (2008)
12. Plandowski, W., Rytter, W.: Complexity of language recognition problems for compressed words. In: Karhumäki, J., Maurer, H.A., Păun, G., Rozenberg, G. (eds.) Jewels are Forever, pp. 262–272. Springer, Heidelberg (1999)
13. Stockmeyer, L.J., Meyer, A.R.: Word problems requiring exponential time. In: STOC 1973, pp. 1–9 (1973)