A Note on SLRE

Peter Habermehl and Richard Mayr

LIAFA - Université Denis Diderot - Case 7014 - 2, place Jussieu,
F-75251 Paris Cedex 05. France. E-mail: {haberm,mayr}@liafa.jussieu.fr

Abstract. We point out common misconceptions about SLRE and bounded languages and correct some errors in various sources.

Definition 1. *A SLRE is a regular language that can be written as a finite union of expressions of the form $w_1^* w_1' w_2^* w_2' \ldots w_{n-1}' w_n^*$ where the w_i, w_i' are words [FIS00].*

Definition 2. *A language is* bounded *iff it is a* **subset** *of a language of the form $w_1^* w_2^* \ldots w_m^*$, where $w_1, \ldots, w_m$ are words [Gin66, HRS76].*

Note that bounded languages need not be equivalent to a language of the form $w_1^* w_2^* \ldots w_m^*$. Bounded languages need not even be regular. In [FIS00, Sut00] the definition of 'bounded' is wrong, while the real bounded languages are called 'plat' (or 'flat') in [FIS00, Sut00]. Note also that all SLRE are bounded, since $w_1^* v_1 w_2^* v_2 \ldots v_{n-1} w_n^* \subseteq w_1^* v_1^* w_2^* v_2^* \ldots v_{n-1}^* w_n^*$ and the class of bounded languages is closed under union.

Definition 3. *A language L has polynomial density iff there exists a polynomial P s.t. for every $n \in \mathbb{N}$ L contains at most $P(n)$ words of length n [Yu97].*

Note that, e.g., $(a+b)^*$ does not have polynomial density. However, all languages over a unary alphabet have polynomial (even constant) density.

Theorem 4. *[HRS76]*
Given a regular expression (or finite automaton) R it is decidable in polynomial time if its language is bounded. If yes, then one expression $w_1^* w_2^* \ldots w_m^*$ with $R \subseteq w_1^* w_2^* \ldots w_m^*$ has polynomial size and can be effectively constructed in polynomial time.

A similar theorem holds for context-free languages, but the complexity result does not carry over.

Theorem 5. *[Gin66]*
Given a context-free grammar G it is decidable if its language is bounded. If yes, then one expression $w_1^* w_2^* \ldots w_m^*$ with $G \subseteq w_1^* w_2^* \ldots w_m^*$ can be effectively constructed.

An analysis of the proof of Theorem 5 in [Gin66] shows that the algorithm requires exponential time. One cannot do better, because there are bounded context-free grammars G for which the smallest expression $w_1^* w_2^* \ldots w_m^*$ with $G \subseteq w_1^* w_2^* \ldots w_m^*$ has exponential size. For example, the grammar with initial symbol A_0 and the rules $A_i \to A_{i+1} A_{i+1}$ for $0 \leq i \leq n-1$ and $A_n \to a$ and $A_n \to b$ generates the language of all words of length 2^n. This language is finite and thus bounded, but the smallest expression $w_1^* w_2^* \ldots w_m^*$ that contains it has exponential size in n.

Theorem 6. *Every bounded language has polynomial density.*

Proof. *For every bounded language L there exists $w_1^* w_2^* \dots w_m^*$ s.t. $L \subseteq w_1^* w_2^* \dots w_m^*$ by Def. 2. Then L contains at most n^m words of length n, because for every i there are at most n possibilities for the position of the border between w_i and w_{i+1}. Thus we define the polynomial $P(n) := n^m$.* □

Note that the converse of Theorem 6 is not true. There are languages of polynomial density that are not bounded. For example, consider the language $\{z_n \mid n \in \mathbb{N}\}$ where z_n is the word of the first n digits of π in decimal coding. This language even has constant density, but it is still not bounded.
However, it is known that all context-free languages of polynomial density are bounded [IR86].

Theorem 7. *Let R be a regular language. Then we have*

$$R \text{ bounded} \iff R \text{ has polynomial density} \iff R \text{ is SLRE}$$

Proof. *If R is bounded then it has polynomial density by Theorem 6. If R has polynomial density and is regular then it is SLRE by [Yu97]. If R is SLRE then it is trivially bounded, because $w_1^* v_1 w_2^* v_2 \dots v_{n-1} w_n^* \subseteq w_1^* v_1^* w_2^* v_2^* \dots v_{n-1}^* w_n^*$ and the class of bounded languages is closed under union.* □

Note that the condition that R is regular is essential. For example the context-free language $a^n b^n$ is bounded (since $a^n b^n \subseteq a^* b^*$) and thus has polynomial density, but is not regular (and of course not SLRE).

Theorem 8. *[HRS76]*
Let R_1 be a regular expression that describes a bounded language and R_2 a general regular expression. The question if $R_1 = R_2$ is co-$\mathcal{NP}$-complete.

Proof. *(sketch) Use Theorem 4 to construct the expression $w_1^* w_2^* \dots w_m^*$ for R_1 in polynomial time. Also check if R_2 is bounded. If no, then $R_1 \neq R_2$. If yes, then construct in polynomial time $v_1^* v_2^* \dots v_l^*$ with $R_2 \subseteq v_1^* v_2^* \dots v_l^*$. Guess numbers $k_1, \dots, k_m$ or $k_1, \dots, k_l$ that describe a word $w_1^{k_1} w_2^{k_2} \dots w_m^{k_m}$ or $v_1^{k_1} v_2^{k_2} \dots v_l^{k_l}$ that distinguishes R_1 and R_2 (i.e., occurs in one, but not in the other). These numbers are at most exponential and can be stored in polynomial space. It can be checked in polynomial time (by using some special matrix construction) if these words are in R_1 and R_2. So the question if $R_1 \neq R_2$ is in $\mathcal{NP}$and thus the question if $R_1 = R_2$ is in co-$\mathcal{NP}$. The co-$\mathcal{NP}$-hardness follows from Theorem 9.* □

Note that there is an error in [Yu97]. There it is claimed that the problem above is $\mathcal{NP}$-complete instead of co-$\mathcal{NP}$-complete. This error has been copied into [FIS00, Sut00].

Theorem 9. *[SM73]*
Given a SLRE E and a single letter 'a'. The question if $E = a^*$ is co-$\mathcal{NP}$-complete.

Proof. *(sketch) Inclusion in co-$\mathcal{NP}$ follows from Theorem 8. The co-$\mathcal{NP}$-lower bound is implicit in a result in [SM73]. In [SM73] it is proved that $INEQ(\{0\}, \{\cup, \cdot, *\})$ is $\mathcal{NP}$-complete, i.e., the inequality of regular languages over a one-letter alphabet with are constructed with the operations of union, concatenation and Kleene-star is $\mathcal{NP}$-complete. By examining their proof it is easy to see that they in fact prove a stronger result. They only use one SLRE and the language a^*. The proof is done by a reduction of 3-SAT to $E \neq a^*$ that uses Gödel-coding to store variable assignments in natural numbers.* □

Note that it follows immediately from Theorem 9 that the inclusion problem $a^* \subseteq E$ for a SLRE E is also co-$\mathcal{NP}$-complete.

Theorem 10. *Given two SLREs E_1, E_2, a bounded regular language B, a general regular language R and a single letter 'a'. The following questions are all co-$\mathcal{NP}$-complete.*

1. $B = R$?
2. $E_1 = R$?
3. $E_1 = B$?
4. $E_1 = E_2$?
5. $R = a^*$?
6. $B = a^*$?
7. $E_1 = a^*$?
8. $E_1 \supseteq a^*$?

Proof. *Directly from Theorem 8 and Theorem 9.* □

Note in particular that equality and inclusion of two SLREs is **not** in $\mathcal{NP} \cap$ co-$\mathcal{NP}$ as claimed in [FIS00, Sut00] (unless $\mathcal{NP}$ = co-$\mathcal{NP}$), but co-$\mathcal{NP}$-complete.
Note also that for a general regular language R the question $R = a^*$ is co-$\mathcal{NP}$-complete, while the question $R = \Sigma^*$ (for a non-unary alphabet Σ) is $PSPACE$-complete [SM73].

References

[FIS00] A. Finkel, S. Purushothaman Iyer, and G. Sutre. Well-abstracted transition systems: Application to fifo automata. In *Proc. 11th Int. Conf. Concurrency Theory (CONCUR'2000)*, pages 566–580. Springer Verlag, 2000.

[Gin66] S. Ginsburg. *The Mathematical Theory of Context-Free Languages*. McGraw-Hill, 1966.

[HRS76] H. B. Hunt III, D. J. Rosenkrantz, and T. G. Szymanski. On the equivalence, containment, and covering problems for the regular and context-free languages. *Journal of Computer and System Sciences*, 12(2):222–268, 1976.

[IR86] O. Ibarra and B. Ravikumar. On sparseness, ambiguity and other decision problems for acceptors and transducers. In *Proceedings of the 3rd Symposium on Theoretical Aspects of Computer Science*, volume 210 of *LNCS*, pages 171–179. Springer Verlag, 1986.

[SM73] L. Stockmeyer and A. Meyer. Word problems requiring exponential time. In *Proceedings of the Fifth ACM Symposium on the Theory of Computing (STOC-73)*, pages 1–9, 1973.

[Sut00] G. Sutre. *Abstraction et accélération de systèmes infinis*. PhD thesis, ENS de Cachan, 2000.

[Yu97] S. Yu. Regular languages. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages, Vol. 1*, pages 41–110. Springer Verlag, 1997.