A Note on the Commutative Closure of Star-Free Languages*

Anca Muscholl Holger Petersen
Institut für Informatik der Universität Stuttgart
Breitwiesenstraße 20–22, D-70565 Stuttgart
e-mail: {muscholl,petersen}@informatik.uni-stuttgart.de

Abstract

We show that the commutative closure of a star-free language is either star-free or not regular anymore. Actually, this property is shown to hold exactly for the closure with respect to a partial commutation corresponding to a transitive dependence relation.

Moreover, the question whether the closure of a star-free language remains star-free is decidable precisely for transitive partial commutation relations.

It is well-known that the commutative closure of regular languages need not be regular. A typical example is the star-free(!) language $(ab)^*$, the commutative closure of which consists of all words over $\{a, b\}$ with equal number of a and b. Moreover, as soon as the closure of languages under partial commutation is considered the question whether the closure of a given regular language remains regular or not is in general undecidable [9]. On the other hand, the question whether the closure of the star of a closed regular language remains regular or not is still an open question for partial commutations. Recently, some progress has been achieved towards a solution for this problem [6, 8] (for a survey and further references see also [2, Chapter 6]).

The aim of this paper is to clarify another aspect concerning the still mysterious behaviour of recognizable languages with respect to closure operations, namely the relationship between star-freeness and closure under commutation. We show that surprisingly there are no star-free languages, where the commutative closure is regular, but not star-free. In fact, we can show a more general result concerning the closure of star-free languages under partial commutations.

*The work of the authors was supported by the ESPRIT Basic Research Action WG 6317: Algebraic and Syntactic Methods in Computer Science (ASMICS 2).

Definition Let Σ be a finite alphabet and $I \subseteq \Sigma \times \Sigma$ be a symmetric relation. The closure of a language $L \subseteq \Sigma^*$ with respect to the partial commutation relation I is denoted by $[L]_I$. It is defined as the least language of Σ^* containing L, which satisfies $uabv \in [L]_I \Leftrightarrow ubav \in [L]_I$ for all $u, v \in \Sigma^*$, $(a, b) \in I$.

Note that if $I = \Sigma \times \Sigma$, then $[L]_I$ is the commutative closure, whereas in the other extreme, $[L]_\emptyset = L$.

Recall that the syntactic congruence $\sim_L$ of a language $L \subseteq \Sigma^*$ is given by $u \sim_L v$ if and only if $xuy \in L \Leftrightarrow xvy \in L$, for all $x, y \in \Sigma^*$. A language $L \subseteq \Sigma^*$ is *aperiodic*, if both the syntactic congruence $\sim_L$ is of finite index and the syntactic monoid of L, $\Sigma^*/\sim_L$, satisfies the equation $x^n = x^{n+1}$ for some integer $n \geq 0$. Equivalently, aperiodic languages are precisely the star-free resp. first-order definable languages (see [10, 5, 7]).

Assume that $D = (\Sigma \times \Sigma) \setminus I$ is transitive and let $\equiv_I \subseteq \Sigma^* \times \Sigma^*$ be the congruence generated by the set $\{(ab, ba) \mid (a, b) \in I\}$. Then the quotient monoid $\mathbb{M}(\Sigma, I) := \Sigma^*/\equiv_I$ is a direct product of free monoids, i.e., $\mathbb{M}(\Sigma, I) = \prod_{i=1}^k \Sigma_i^*$ for the partition $\Sigma = \dot{\bigcup}_{i=1}^k \Sigma_i$ of the alphabet corresponding to the connected components of (Σ, D). By abuse of notation we will identify in this case the closure $[L]_I \subseteq \Sigma^*$ of a language $L \subseteq \Sigma^*$ with the subset $L/\equiv_I$ of $\mathbb{M}(\Sigma, I)$. Hence in this case we will consider $[L]_I$ as a subset of $\prod_{i=1}^k \Sigma_i^*$. By $\pi_i : \Sigma^* \to \Sigma_i^*$ for $1 \leq i \leq k$ we denote the canonical projections.

Theorem 1 *Let Σ be a finite alphabet, $I \subseteq \Sigma \times \Sigma$ a partial commutation and $D = (\Sigma \times \Sigma) \setminus I$ the complementary relation. Then the following assertions hold:*

If D is transitive, then the closure $[L]_I$ of a star-free language $L \subseteq \Sigma^*$ is either star-free or not regular.

Conversely, if D is not transitive, then there exist star-free languages $L \subseteq \Sigma^*$ such that $[L]_I$ is regular, but not star-free.

Proof: Let us first assume that $D \subseteq \Sigma \times \Sigma$ is transitive and the closure $[L]_I$ of the star-free language L is regular. We have to show the existence of some integer $n \geq 0$ such that $xv^ny \in [L]_I \Rightarrow xv^{n+\delta}y \in [L]_I$, for all $x, v, y \in \Sigma^*$, $\delta \in \{-1, +1\}$.

By Mezei's theorem [1] we can express $[L]_I$ as a finite union $[L]_I = \bigcup_{\text{fin}} \prod_{i=1}^k L_i$, where every $L_i \subseteq \Sigma_i^*$ is recognizable. Moreover, different $L_i, L_i' \subseteq \Sigma_i^*$ in this union are either disjoint or equal. Let S be the set of all languages L_i occurring in the above representation of $[L]_I$. Let $m > 0$ denote the integer $\max_{L' \in S} (\min_{w \in L'} |w|)$. Further an integer $p > 0$ is chosen such that $u^p \sim_L u^{p+\delta}$ for all u and $\delta \in \{-1, +1\}$ (recall, L is aperiodic). We let $n = [(k-1)m+1]p$ and show that this value suffices in order to obtain the desired property for $[L]_I$.

Consider $xv^ny \in [L]_I$, $\delta \in \{-1, +1\}$, with $xv^ny = (w_1, \ldots, w_k)$ and let $xv^{n+\delta}y = (w_1', \ldots, w_k')$. Suppose, we have already shown that $(w_1', \ldots, w_{i-1}', w_i, \ldots, w_k) \in [L]_I$ for $i \geq 1$. Therefore we have for some $\prod_{i=1}^k L_i$ in the above representation of $[L]_I$: $(w'_1, \ldots, w'_{i-1}, w_i, \ldots, w_k) \in \prod_{i=1}^k L_i$. Let $z_i = w_i$ and consider $z_j \in \Sigma_j^*$, $1 \le j \le k$, $j \ne i$, of minimal length satisfying $z_j \in L_j$. Note that we also have $(z_1, \ldots, z_k) \in [L]_I$. Hence, there exists some $z \in L$ with $\pi_i(z) = z_i$, for every $1 \le i \le k$.

Let $q = \sum_{j \ne i} |z_j| \le (k-1)m$, then we can factorize z such that

$$z = u_0 x_1 u_1 \cdots x_q u_q\,,$$

with $u_l \in \Sigma_i^*$ and $x_l \in \Sigma \setminus \Sigma_i$ for all l. Due to the choice of n, there remain at least $[(k-1)m+1](p-1)+1$ occurences of v_i being factors of at most $(k-1)m+1$ words u_l, $0 \le l \le q$. Applying the pigeon-hole principle, we obtain that at least one u_l can be decomposed as $u_l = x' v_i^p y'$, for suitable $x', y' \in \Sigma_i^*$. This yields inevitably the word factor v_i^p in z. Hence, $z = x'' v_i^p y''$ and we have $x'' v_i^{p+\delta} y'' \in L$. Thus, $(z_1, \ldots, z_{i-1}, z'_i, z_{i+1}, \ldots, z_k) \in [L]_I$, where $z'_i = w'_i$. It follows that $z_j \in L'_j$, $j \ne i$, and $z'_i \in L'_i$ hold for some $\prod_{i=1}^k L'_i$ in the finite union representing $[L]_I$. For $j \ne i$ we have from $L_j \cap L'_j \ne \emptyset$ also $L'_j = L_j$. This is due to the property of the representation given by Mezei's theorem noticed above. Hence, we obtain $(w'_1, \ldots, w'_{i-1}, w'_i, w_{i+1}, \ldots, w_k) \in \prod_{i=1}^k L'_i$. In particular, it follows $(w'_1, \ldots, w'_{i-1}, w'_i, w_{i+1}, \ldots, w_k) \in [L]_I$.

For the converse, consider a monoid $\mathbb{M}(\Sigma, I)$ where $D = (\Sigma \times \Sigma) \setminus I$ is non-transitive. Hence, there are distinct letters $a, b, c \in \Sigma$ with $\{(a,c), (b,c)\} \subseteq D$, but $(a,b) \in I$. Now, it is easy to check that the language $(abcbac)^*$ is aperiodic (star-free):

$$(abcbac)^* = (abc\Sigma^* \cap \Sigma^* bac) \setminus (\Sigma^* abc(\Sigma^* \setminus bac\Sigma^*) \cup \Sigma^* bac(\Sigma^* \setminus abc\Sigma^*)).$$

However, the closure $K := [(abcbac)^*] = [(ab+ba)c(ab+ba)c]^*$ is obviously not aperiodic since for every $n > 0$ we have $(abc)^n \not\sim_K (abc)^{n+1}$. ∎

Remark Note that this result cannot be extended to languages over Σ^ω. As soon as we have a pair $(a,b) \in I$ we can exhibit the star-free language $(aab)^* b^\omega$, where the closure $\{w \in \{a,b\}^\omega \mid w$ contains an even number of a's and an infinite number of b's $\}$ is regular, but not star-free anymore.

It is natural to look for a decision procedure for the question, whether the closure of a given star-free language remains star-free. We can show that this problem is in general undecidable following [4, 9]. We follow directly the proof given in [9] for the undecidability of the question whether the closure of a regular language is still regular. We included the proof below in order to examine the languages involved in the construction and show that they are star-free.

Let B and C be disjoint alphabets. An instance of Post's correspondence problem (PCP) will be encoded by two homomorphisms $g, h : B^* \to C^*$. A solution for this instance is a word $w \in B^+$ such that $g(w) = h(w)$.

Consider now the language W_g (and analogously, W_h) which has been used in the reduction given in [9]:

$$W_g = \{(w\, g(w), c^{|g(w)|}) \mid w \in B^+\}.$$

Let our commutation relation be $I = \{(x,c),(c,x) \mid x \in B \cup C\}$. We define below a star-free language L_g such that $[L_g]_I = \overline{W}_g$ (analogously for L_h).

The following technical remark yields a concise star-free description of L_g (and L_h).

Remark Let Σ be an alphabet, $X \subseteq \Sigma$ a set of symbols, $c \notin X$, and $M \subset X\{c\}^*$ a finite set. Then M^* is a star-free language.

The last remark is justified by the fact that M is a very pure code and therefore M^+ is locally testable (i.e. it can be characterized by its suffixes, prefixes, and factors up to some given length) [7, p.120].

Now, L_g is formed as a union of sets described in the following. The first set takes care of the case that a symbol from C precedes a symbol from B, which is easily seen to be star-free. In the following sets we consider four ways in which the number of c's can differ from the correct one. (There could be strictly less/more c's than in the g-images of all symbols from B, respectively strictly less/more c's than symbols from C. It is not difficult to form the corresponding sets using the remark above.) Finally, the last set describes—under the assumption that the number of c's agrees with both numbers mentioned above—those words that fail to be encodings of solutions because some symbol x is not properly mapped to its image $g(x)$. This set is defined by the following expression (again using the remark):

$$B^* \left(\bigcup_{x \in B} x c^{|g(x)|} \{ y c^{|g(y)|} \mid y \in B\}^* \{zc \mid z \in C\}^* \{w \mid w \neq g(x), |w| = |g(x)|\} \right) C^*$$

Since star-free languages are by definition closed with respect to union we have shown that L_g (and L_h) are star-free languages.

Theorem 2 *Let Σ be a finite alphabet and $I \subseteq \Sigma \times \Sigma$ a partial commutation.*

It is decidable whether the closure $[L]_I$ of a star-free language L is star-free if and only if I is transitive.

Proof: With the notations introduced above note that the language $[L_g \cup L_h]_I = \overline{W}_g \cup \overline{W}_h$ is equal to $(B \cup C)^* \times c^*$ (and hence star-free) if the PCP encoded by g, h has no solution. On the other hand, if the PCP instance has a solution consider some $w \in B^+$ such that $g(w) = h(w)$. Then

$$((B \cup C)^* \times c^* \setminus [L_g \cup L_h]_I) \cap w^* g(w)^* \times c^* = \{(w^n g(w)^n, c^{n|g(w)|}) \mid n \geq 1\}$$

is not recognizable, hence $[L_g \cup L_h]_I$ is not recognizable (and thus not star-free).

Finally, we note that B and C can be encoded as usual by two letters. Moreover, if $I \subseteq \Sigma \times \Sigma$ is not transitive, then there exist different letters a, b, c in Σ such that $(a,c) \in I$, $(b,c) \in I$, but $(a,b) \in D$. This yields the first part of the claim, i.e., the undecidability of the question for I non-transitive.

For the second part we note that the closure $[L]_I$ of a regular language $L \subseteq \Sigma^*$ can be identified with a rational subset of the monoid $\mathbb{M}(\Sigma, I)$ (i.e., a subset generated from $\emptyset$ and the singleton sets using the operations concatenation, union and Kleene star). If I is transitive, then $\mathbb{M}(\Sigma, I)$ is a free product of free commutative monoids. Moreover, the question whether a rational subset of a free product of free commutative monoids is recognizable (i.e., saturated by a congruence of finite index) is effectively decidable [9]. More precisely, given a rational expression for $[L]_I$ a finite automaton can be computed, which recognizes exactly $[L]_I$ whenever $[L]_I$ is recognizable, which in turn is decidable. If the answer is positive, we can compute the syntactic monoid and decide whether $[L]_I$ is star-free or not. ∎

Remark Concerning the different assumptions in Thms. 1, 2 note that both I and $D = (\Sigma \times \Sigma) \setminus I$ are transitive if and only if the quotient monoid $\mathbb{M}(\Sigma, I)$ is either free or free commutative.

Acknowledgement: We would like to thank Volker Diekert for bringing this problem to our attention. Thanks are also due to the anonymous referee for useful comments which helped improving the presentation.

References

[1] J. Berstel. *Transductions and context-free languages.* Teubner Studienbücher, Stuttgart, 1979.

[2] V. Diekert and G. Rozenberg, editors. *The Book of Traces.* World Scientific, Singapore, 1995.

[3] G. Guaiana, A. Restivo, and S. Salemi. Star-free trace languages. *Theoretical Computer Science*, 97:301–311, 1992. A preliminary version was presented at STACS'91, Lecture Notes in Computer Science 480 (1991).

[4] L. P. Lisovik. The identity problem for regular events over the direct product of free and cyclic semigroups (in Ukrainian). *Dok. Akad. Nauk. Ukraiinskoj RSR*, ser. A(6):410–413, 1979.

[5] R. McNaughton and S. Papert. *Counter-free Automata.* MIT Press, Cambridge, MA, 1971.

[6] Y. Métivier and G. Richomme. New results on the star problem in trace monoids. *Information and Computation*, 119:240–251, 1995.

[7] J.-E. Pin. *Varieties of formal languages.* North Oxford Academic, London, 1986.

[8] G. Richomme. Some trace monoids where both the Star Problem and the Finite Power Property Problem are decidable. In I. Privara et al., editors, *Proceedings of the 19th Symposium on Mathematical Foundations of Computer Science (MFCS'94), Kosice (Slovakia) 1994*, number 841 in Lecture Notes in Computer Science, pages 577–586, Berlin-Heidelberg-New York, 1994. Springer.

[9] J. Sakarovitch. The "last" decision problem for rational trace languages. Report LITP 91.77, Inst. Blaise Pascal, Univ. Paris 6, 1991. Abstract presented at the 1st International Symposium of Latin American Theoretical Informatics (LATIN'92), Lecture Notes in Computer Science 583 (1992) 460–473.

[10] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.