ON THE SUBSETS OF RANK TWO IN A FREE MONOID:
A FAST DECISION ALGORITHM

(EXTENDED ABSTRACT)

Jean NERAUD
LITP, Institut Blaise Pascal \ LIR
Université de Rouen,
Place Emile Blondel, F-76134 Mont-Saint-Aignan (FRANCE)

Abstract

Given a finite subset $X$ of a free monoid $A^*$, we define the rank of $X$ as $r(X) = \min\{|Y| : X \subseteq Y\}$. The problem we study here, is to decide whether $r(X) \leq 2$ or not. We propose an $O(n \ln m)$ algorithm, where $n$ stands for the sum of the lengths of the words in $X$, and $m$ stands for the length of the longest word.

INTRODUCTION

In the context of the free monoids theory, two notions of rank are actually proposed. First, given a set of words $X$ of the free monoid $A^*$, its rank can be defined as the cardinality $r_1(X)$ of the basis of the free hull of $X$, i.e. the smallest free submonoid which contains $X$. The Defect theorem (cf eg [Lo 83]) says that if $X$ is not a code then its cardinality $|X|$ satisfies $r_1(X) \leq |X| - 1$. In other words, $X$ is a code iff $r_1(X) = |X|$.

In another way, given a finite subset $X \subseteq A^*$, we can define the rank of $X$ as the smallest cardinality $r(X)$ of a finite set $Y$ satisfying $X \subseteq Y^*$, i.e. such that all the words in $X$ can be written as the concatenation product of words in $Y$. Clearly, we have the inequality $r_1(X) \leq r(X)$, moreover, this second notion corresponds to the concept of degree, introduced in [HK 86]. This topic meets that of elementariness: a finite subset $X$ is elementary (or independent)
iff \( r(X) = |X| \). Historically the notion of elementariness applied to morphisms, and its introduction constituted the major step in the delicate proof of the decidability of the famous DOL sequence equivalence problem.

As said in [HK 86], considering the second notion of rank is often more important in a combinatorial point of view. Moreover, although these two concepts seem very close, they leads to different topological properties: for instance, given an elementary set, being maximal is not equivalent to being a maximal code (cf N 88 2]). In [N 90], properties of our rank are established, with regard to some operations on sets.

Within the context of decision problems, an important difference concerns the properties of the integers \( r_1(X) \) and \( r(X) \). Indeed, it is well known that, given a finite set of words \( X \), deciding whether \( X \) is a code can be achieved by applying the classical Sardinas and Patterson algorithm ([SP 50]). This algorithm has been implanted so that it allows to process the set \( X \) in time complexity \( O(n \cdot \ln |X|) \) (cf [R 82] or [AG 84]), where \( n \) stands for the sum of the lengths of the words in \( X \), and \( |X| \) stands for the cardinality of \( X \). Moreover, in [S 76] the computation of \( r_1(X) \) is given in time \( O(|X| \cdot n^2) \).

On the contrary, deciding whether \( X \) is elementary, or deciding whether \( r(X) \) is smaller than a given integer \( k \), are NP-hard problems (cf [N 88 i]). In this way, it is of interest to examine restrictions of the last problem. For instance, consider a set \( X \) whose elements are words with length \( p \), where \( p \) is a given positive integer. If \( p = 2 \) then computing \( r(X) \) can be done in time \( O(|X|^2) \), by determining the family of the connected components of a direct graph associated with \( X \). Howethr, for \( p \geq 3 \), deciding whether \( r(X) = k \) remains NP-complete. As another example, given a finite set of words \( X \), deciding whether \( r(X) = 1 \) can be done by applying the Knuth Morris and Pratt algorithm ([KMP 77]) (indeed we have \( r(X) = 1 \) iff \( r_1(X) = 1 \)).

In this paper, we are interested in the following restriction: given a finite set of words \( X \), the problem is to decide whether \( r(X) = 2 \) or not. The naive algorithm consists in examining all the two elements subsets \( \{x, y\} \) of factors of the words in \( X \), thus we shall solve our problem in time \( O(n^5) \). A refinement in \( O(n^3) \) can be done by considering restrictive properties of prefixity on \( \{x, y\} \). Here, we establish the following: