Theoretical Computer Science 524 (2014) 41–47

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Recurrence in infinite partial words ☆

F. Blanchet-Sadri [a,*], Bob Chen [b], Sinziana Munteanu [c]

[a] Department of Computer Science, University of North Carolina, PO Box 26170, Greensboro, NC 27402-6170, USA
[b] Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Dept 0112, LaJolla, CA 92093-0112, USA
[c] School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891, USA

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form 25 March 2013
Accepted 20 December 2013
Communicated by B. Durand

Keywords:
Automata and formal languages
Combinatorics on words
Partial words
Recurrent words
Recurrence quotient
Recurrence function

ABSTRACT

The *recurrence function* $R_w(n)$ of an infinite word w was introduced by Morse and Hedlund in relation to symbolic dynamics. It is the size of the smallest window such that, wherever its position on w, all length n subwords of w will appear at least once inside that window. The *recurrence quotient* $\rho(w)$ of w, defined as $\limsup \frac{R_w(n)}{n}$, is useful for studying the growth rate of $R_w(n)$. It is known that if w is periodic, then $\rho(w) = 1$, while if w is not, then $\rho(w) \geqslant 3$. A long standing conjecture from Rauzy states that the latter can be improved to $\rho(w) \geqslant \frac{5+\sqrt{5}}{2} \sim 3.618$, this bound being true for each Sturmian word and being reached by the Fibonacci word. In this paper, we study in particular the spectrum of values taken by the recurrence quotients of infinite partial words, which are sequences that may have some undefined positions. In this case, we determine exactly the spectrum of values, which turns out to be 1, every real number greater than or equal to 2, and ∞. More precisely, if an infinite partial word w is "ultimately factor periodic", then $\rho(w) = 1$, while if w is not, then $\rho(w) \geqslant 2$, and we give constructions of infinite partial words achieving each value.

1. Introduction

A topic of interest on infinite words is the one of *recurrence*. An infinite *recurrent* word is one in which all of the finite subwords appear infinitely often, where finite subwords are finite contiguous blocks of letters. Many concepts dealing with recurrence were introduced by Morse and Hedlund in [8] in relation to symbolic dynamics. More recently, in [5], Cassaigne presented some results that establish connections between recurrence and the *subword complexity* $p_w(n)$ of an infinite word w, which is the number of distinct subwords of length n in w. He also described, under some conditions, a method for computing the *recurrence function* $R_w(n)$ of an infinite word w, which is the minimum length such that every contiguous block of letters in w of this length contains every length n subword of w.

The *recurrence quotient* of an infinite word w is defined to be $\rho(w) = \limsup \frac{R_w(n)}{n}$. Cassaigne studied in [4] the spectrum of possible recurrence quotients for Sturmian words (those with subword complexity $n+1$ [7]). It is a compact subset of $[0,\infty]$ with empty interior, it has cardinality of the continuum, and its smallest accumulation point is approximately 4.58565. He discussed in [5] the spectrum of values $\mathcal{S} \subset \mathbb{R} \cup \{\infty\}$ taken by ρ for arbitrary words, about which much less is known. Periodic words, those of the form x^ω, have a quotient ρ of 1, and he proved, using graph representations, that $\rho \geqslant 3$ otherwise.

☆ This material is based upon work supported by the National Science Foundation under Grant No. DMS-1060775. We thank the referees of preliminary versions of this paper for their very valuable comments and suggestions.

* Corresponding author.
E-mail address: blanchet@uncg.edu (F. Blanchet-Sadri).

0304-3975/$ – see front matter
http://dx.doi.org/10.1016/j.tcs.2013.12.016

Theorem 1. *(See [5].) Lower bounds for the recurrence quotients achievable by infinite words w are:*

- *If w is periodic, then $\rho(w) = 1$;*
- *If w is nonperiodic, then $\rho(w) \geqslant 3$.*

Note that this bound of 3 improved an earlier bound of 2 that can be easily deduced from Hedlund and Morse's inequalities that $R_w(n) \geqslant p_w(n) + n - 1$ and $p_w(n) \geqslant n + 1$ for nonperiodic recurrent words [8] (as was mentioned in [5], if w is a recurrent infinite word that is not periodic, then $R_w(n) \geqslant 2n$). However, the bound of 3 is not tight and Rauzy conjectured that the minimum value for nonperiodic words is $\frac{5+\sqrt{5}}{2} \sim 3.618$, which is achieved by the well-known Fibonacci word

$$0100101001001010010100100101001001\ldots$$

defined by $F_{n+2} = F_{n+1}F_n$, where $F_0 = 0$ and $F_1 = 01$ [10]. Very little else is known about the topological structure of $\mathcal{S}$.

How do these results related to recurrence in infinite words translate to the framework of infinite partial words? Partial words are sequences over a finite alphabet that may contain wildcard symbols, called holes, which match or are compatible with all letters; partial words without holes are said to be full words (or simply words). Combinatorics on partial words was initiated by Berstel and Boasson in the context of gene comparison [1] and has been developing since (see for instance [2]). In [3], Blanchet-Sadri et al. introduced recurrent infinite partial words. Their results culminate in showing that the various completions of an infinite partial word w, i.e., infinite full words built by filling in the holes of w with letters from the alphabet, can achieve subword complexities equal or close to that of w if and only if w is recurrent or ultimately recurrent (ultimate recurrence of w means recurrence of some suffix of w).

As mentioned earlier, the exact spectrum $\mathcal{S}$ of achievable recurrence quotients for infinite full words is not known, although it is known to be included in $\{1\} \cup [3, \infty]$ (see Theorem 1). In this paper, we consider the recurrence quotients of infinite partial words. Our main result, Theorem 2, states that the exact spectrum $\mathcal{S}_\diamond$ of achievable recurrence quotients for infinite partial words is $\{1\} \cup [2, \infty]$.

Theorem 2. *The spectrum of recurrence quotients achievable by partial words is $\mathcal{S}_\diamond = \{1\} \cup [2, \infty]$. More precisely, the following hold for infinite partial words w:*

- *If w is ultimately factor periodic, then $\rho(w) = 1$;*
- *If w is not ultimately factor periodic, then $\rho(w) \in [2, \infty]$.*

Note that to obtain different spectra in the context of partial words, we need to consider *factor periodicity* and not just *periodicity* as in the context of full words (see Section 2 for the definition of these terms). Moreover, we give constructions of infinite partial words achieving each value in $\mathcal{S}_\diamond$. We also provide some results showing how the distribution of holes in a recurrent infinite partial word w implies the nonultimate periodicity of w, strenghtening some results in [3].

The contents of our paper is as follows: In Section 2, we introduce our notations and terminology on partial words. In Section 3, we study the spectrum of values for recurrence quotients of infinite partial words. In Section 3.1 we prove that only the values in $\{1\} \cup [2, \infty]$ are possible recurrence quotients, and in Section 3.2 we give explicit constructions achieving each value. In Section 4, we give some properties of recurrent partial words. We present some relations between recurrence and periodicity as well as a relation between uniform recurrence and subword complexity. Finally in Section 5, we conclude with some remarks.

2. Preliminaries

For more information on basics of partial words, we refer the reader to [2]. Unless explicitly stated, A is a finite alphabet that contains at least two distinct letters 0 and 1. We denote the set of all words over A by A^*, which under the concatenation operation forms a free monoid whose identity is the empty word ε.

A *finite partial word* of length n over A is a function $w : \{0, \ldots, n-1\} \to A \cup \{\diamond\}$, where $\diamond \notin A$. The union set $A \cup \{\diamond\}$ is denoted by $A_\diamond$ and the length of w by $|w|$. A *right infinite partial word* or *infinite partial word* over A is a function $w : \mathbb{N} \to A_\diamond$. In both the finite and infinite cases, the symbol at position i in w is denoted by w_i. If $w_i \in A$, then i is defined in w, and if $w_i = \diamond$, then i is a hole in w. If w has no holes, then w is a *full word*. Two finite partial words u and v of same length are *compatible*, denoted $u \uparrow v$, if $u_i = v_i$ whenever $u_i, v_i \in A$. Equivalently, u and v are compatible if there exists a full word w which is a completion of both u and v, that is, we can fill the holes of u and v and obtain w in either case.

Let $w = w_0w_1w_2\ldots$ be an infinite partial word over A. We say that the finite partial word u is a *factor* of w if u is a block of consecutive symbols of w, that is, $w = xuy$ for some partial words x, y. We say that the finite full word v is a *subword* of w if v is compatible with some factor of w. We write $\mathrm{Sub}_w(n)$ to denote the set of length n subwords of w, and $\mathrm{Sub}(w)$ to denote the set of all subwords of w. The *subword complexity* of w is the function defined by $p_w(n) = |\mathrm{Sub}_w(n)|$. For example, if $w = 0010\underline{00\diamond 01}0\diamond 1\ldots$ and $A = \{0, 1\}$, then 00001 and 00101 are the subwords compatible with the underlined factor of w. We can check that $\mathrm{Sub}_w(2) = \{00, 01, 10, 11\}$ and $p_w(2) = 4$.