Journal of Combinatorial Theory, Series A 136 (2015) 96–125

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Generalized trapezoidal words ☆

Amy Glen [a], Florence Levé [b]

[a] *School of Engineering & Information Technology, Murdoch University, 90 South Street, Murdoch, WA 6150, Australia*

[b] *Laboratoire MIS, Université de Picardie Jules Verne, 33 rue Saint Leu, 80039 Amiens, France*

A R T I C L E   I N F O

Article history:
Received 3 August 2014
Available online 3 July 2015

Keywords:
Word complexity
Trapezoidal word
Sturmian word
Palindrome
Rich word

A B S T R A C T

The *factor complexity function* $C_w(n)$ of a finite or infinite word w counts the number of distinct *factors* of w of length n for each $n \geq 0$. A finite word w of *length* $|w|$ is said to be *trapezoidal* if the graph of its factor complexity $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$) is that of a regular trapezoid (or possibly an isosceles triangle); that is, $C_w(n)$ increases by 1 with each n on some interval of length r, then $C_w(n)$ is constant on some interval of length s, and finally $C_w(n)$ decreases by 1 with each n on an interval of the same length r. Necessarily $C_w(1) = 2$ (since there is one factor of length 0, namely the *empty word*), so any trapezoidal word is on a binary alphabet. Trapezoidal words were first introduced by de Luca (1999) when studying the behaviour of the factor complexity of *finite Sturmian words*, i.e., factors of infinite "cutting sequences", obtained by coding the sequence of cuts in an integer lattice over the positive quadrant of $\mathbb{R}^2$ made by a line of irrational slope. Every finite Sturmian word is trapezoidal, but not conversely. However, both families of words (trapezoidal and Sturmian) are special classes of so-called *rich words* (also known as *full words*) – a wider family of finite and infinite words characterized by containing the maximal number of palindromes – studied in depth by the first author and others in 2009.

☆ *Note:* Preliminary work on this new class of words was presented in talks by the first author at the *35th Australasian Conference on Combinatorial Mathematics & Combinatorial Computing* (35ACCMCC) in December 2011 and at a workshop on *Outstanding Challenges in Combinatorics on Words* at the Banff International Research Station (BIRS) in February 2012.

E-mail addresses: amy.glen@gmail.com (A. Glen), florence.leve@u-picardie.fr (F. Levé).

http://dx.doi.org/10.1016/j.jcta.2015.06.008

In this paper, we introduce a natural generalization of trapezoidal words over an arbitrary finite alphabet $\mathcal{A}$, called *generalized trapezoidal words* (or *GT-words* for short). In particular, we study combinatorial and structural properties of this new class of words, and we show that, unlike the binary case, not all GT-words are rich in palindromes when $|\mathcal{A}| \geq 3$, but we can describe all those that are rich.

1. Introduction

Given a finite or infinite word w, let $C_w(n)$ (resp. $P_w(n)$) denote the *factor complexity function* (resp. the *palindromic complexity function*) of w, which associates with each integer $n \geq 0$ the number of distinct factors (resp. palindromic factors) of w of length n. The well-known *infinite Sturmian words* are characterized by both their factor complexity and palindromic complexity. In 1940 Morse and Hedlund [11] established that an infinite word $\boldsymbol{w}$ is Sturmian if and only if $C_{\boldsymbol{w}}(n) = n + 1$ for each $n \geq 0$. Almost half a century later, in 1999, Droubay and Pirillo [7] showed that an infinite word $\boldsymbol{w}$ is Sturmian if and only if $P_{\boldsymbol{w}}(n) = 1$ whenever n is even, and $P_{\boldsymbol{w}}(n) = 2$ whenever n is odd. In the same year, de Luca [4] studied the factor complexity function of finite words and showed, in particular, that if w is a *finite Sturmian word* (meaning a factor of an infinite Sturmian word), then the graph of $C_w(n)$ as a function of n (for $0 \leq n \leq |w|$, where $|w|$ denotes the *length* of w) is that of a regular trapezoid (or possibly an isosceles triangle). That is, $C_w(n)$ increases by 1 with each n on some interval of length r, then $C_w(n)$ is constant on some interval of length s, and finally $C_w(n)$ decreases by 1 with each n on an interval of the same size r. Such a word is said to be *trapezoidal*. Since $C_w(1) = 2$, any trapezoidal word is necessarily on a binary alphabet.

In this paper, we study combinatorial and structural properties of the following natural generalization of trapezoidal words over an arbitrary finite alphabet.

Definition 1. A finite word w with *alphabet* $Alph(w) := \mathcal{A}$, $|\mathcal{A}| \geq 2$, is said to be a ***generalized trapezoidal word*** (or ***GT-word*** for short) if there exist positive integers m, M with $m \leq M$ such that the factor complexity function $C_w(n)$ of w increases by 1 for each n in the interval $[1, m]$, is constant for each n in the interval $[m, M]$, and decreases by 1 for each n in the interval $[M, |w|]$. That is, w is a GT-word if there exist positive integers m, M with $m \leq M$ such that C_w satisfies the following:

$$\begin{aligned}
&C_w(0) = 1, && \\
&C_w(i) = |\mathcal{A}| + i - 1 && \text{for } 1 \leq i \leq m, \\
&C_w(i+1) = C_w(i) && \text{for } m \leq i \leq M - 1, \\
&C_w(i+1) = C_w(i) - 1 && \text{for } M \leq i \leq |w|.
\end{aligned}$$