



If a machine did it, it is probably transcendental (even p -adically)

Laura Capuano¹ · Sara Checcoli² · Marzio Mula³ · Lea Terracini⁴

Received: 7 April 2025 / Accepted: 27 June 2026
© The Author(s) 2026

Abstract

Continued fraction expansions provide a well-established bridge between algebraic properties of numbers and combinatorics on words. In this article, we investigate the algebraicity of p -adic numbers whose continued fractions arise from certain classes of words which generalize the classical automatic, periodic and palindromic words. Our main result shows that, under mild conditions on the p -adic continued fraction expansion, such numbers are either algebraic of degree at most 2 or transcendental. This result provides an analogue of results of Bugeaud and Adamczewski–Bugeaud in the real setting and extends previous works that were limited to specific choices of p -adic floor functions and less general classes of words.

Keywords p -adic numbers · Continued fractions · Transcendence · Combinatorics on words

Mathematics Subject Classification 11D88 · 11J70 · 11J81 · 68R15

1 Introduction

Continued fractions can be viewed as a way to map numbers in a given set to words, i.e. possibly infinite sequences of elements from a specified alphabet $\mathcal{A}$.

There are several ways to construct continued fraction algorithms depending on the context. In this article, we will focus on the case where the numbers belong to a field K , which is either $\mathbb{R}$ or $\mathbb{Q}_p$, and where the alphabet $\mathcal{A}$ is a subset of either $\mathbb{Z}$ or $\mathbb{Z}[1/p]$. We denote by $\mathbb{N}$ the set of strictly positive integers.

✉ Laura Capuano
laura.capuano@uniroma3.it

Sara Checcoli
sara.checcoli@univ-grenoble-alpes.fr

Marzio Mula
marzio.mula@unibw.de

Lea Terracini
lea.terracini@unito.it

¹ Dipartimento di Matematica e Fisica, Univ. degli Studi Roma Tre, Rome, Italy

² Univ. Grenoble Alpes, CNRS, IF, 38000 Grenoble, France

³ Research Institute CODE, Univ. of the Bundeswehr Munich, Munich, Germany

⁴ Dipartimento di Informatica, Univ. di Torino, Turin, Italy

Most classical constructions are based on the choice of a function, called *floor function*, $s : K \rightarrow K$, which determines the choice of the alphabet $\mathcal{A} = \text{Im}(s)$.

In the archimedean case, we have $K = \mathbb{R}$, $\mathcal{A} \subset \mathbb{Z}$ and s is the classical floor function. However, in the p -adic setting where $K = \mathbb{Q}_p$ and $\mathcal{A} \subset \mathbb{Z}[1/p]$, there are infinitely many possible choices for the floor function s and they all share analogous properties to the classical one (see Sect. 3.1).

In all cases, given a floor function s and $\alpha \in K$, the corresponding word $a_0 a_1 a_2 \dots$ is obtained via the algorithm:

$$\gamma_0 = \alpha, \quad a_n = s(\gamma_n), \quad \gamma_{n+1} = \frac{1}{\gamma_n - a_n} \quad \text{if } \gamma_n \neq a_n, \quad (1)$$

which stops when $\gamma_n = a_n$. The a_n 's are called the *partial quotients* of α . From these, one can define the two important sequences of *continuants* $(A_n)_{n \geq -1}$ and $(B_n)_{n \geq -1}$, given by the following recurrences:

$$\begin{aligned} A_{-1} &= 1, & A_0 &= a_0, & A_n &= a_n A_{n-1} + A_{n-2}, \\ B_{-1} &= 0, & B_0 &= 1, & B_n &= a_n B_{n-1} + B_{n-2}. \end{aligned} \quad (2)$$

The quotient A_n/B_n is called the n -th *convergent* of the continued fraction expansion (with respect to s) of α . Indeed, it is a classical result that the sequence $(A_n/B_n)_n$ converges p -adically to α [18, Theorem 2].

If a number α corresponds, via a continued fraction algorithm with floor function s , to the word $a_0 a_1 a_2 \dots$, we write

$$\alpha = [a_0, a_1, a_2, \dots]_s.$$

When $K = \mathbb{R}$ and s is the classical floor function, we simply write $\alpha = [a_0, a_1, a_2, \dots]$.

Continued fractions share several features with the more common base- b representations of numbers [16]: they can be computed algorithmically, they are well-suited for (some) arithmetic operations, and, surprisingly, they carry over some properties from the realm of word-combinatoric to the realm of numbers. In this article we will investigate the latter feature in the case of p -adic continued fractions.

Before doing so, let us briefly consider the classical archimedean case, from which our approach is inspired. It is well known [10, Ch. 4] that archimedean continued fractions provide a bijection between real numbers in $[0, 1)$ and words on $\mathbb{N}$ (excluding finite words ending with 1). Not only does this bijection restrict to a bijection between rationals and finite words, but, as already shown by Lagrange, it also restricts to a bijection between algebraic numbers of degree 2 and eventually periodic words. This motivated the following:

Question 1.1 Are there families of *special* words that correspond to families of algebraic numbers of degree > 2 ?

Notably, the answer to this question is expected to be negative. The reasoning behind this expectation can be seen as a sort of statement in *unlikely intersections*: any set of words exhibiting some form of structure constitutes a small subset within the set of all words. Consequently, words from such a set arise from the continued fraction expansion of a small subset of the real numbers. Such a small set is unlikely to intersect the set of algebraic numbers, which is itself small, unless the words we took were *very special* i.e. finite or periodic.

Historically, two families of special words have mainly been considered as natural generalizations of periodic words: words on finite subsets of $\mathbb{N}$ and words containing increasingly large repeated or mirrored prefixes. Although these are the primary candidates for being new

special families with algebraic continued fractions, all available partial results suggest they fall short of this goal, revealing instead a substantial gap between periodic words and any form of generalization.

Indeed, the former family, i.e. words on a finite alphabet, was already studied by Borel, who proved that it is in bijection with a measure-zero set of real numbers. As mentioned in [38, §4], it seems that Khintchine was the first to ask whether this set contains any algebraic number of degree > 2 . While this question remains open in general, in the meantime several specific cases have been studied and suggest that the answer should be negative. Such known partial results include families of quasi-periodic words [26, Ch. VII; 13], and some words of linear complexity i.e. words containing only $O(n)$ distinct subwords of length n , such as Thue–Morse words [2, 31] or Sturmian words [7, 30].

In a series of later works [1, 4–6, 19] the authors considered generalisations of quasi-periodic and quasi-palindromic words, namely the $\spadesuit$ words and $\clubsuit$ words.

Before giving their definitions, we fix some notation: for a finite word ω , we denote its length by $|\omega|$ and we denote by $\widetilde{\omega}$ the word obtained by reversing the order of the letters of ω . According to [19, beginning of Sections 3 and 5], we have the following:

Definition 1.2 Let $c > 0$ be a real number. A word ω has *property $\spadesuit$* (respectively *property $\clubsuit$*) *with constant c* if it is not ultimately periodic and if there exists three sequences $(U_n)_n, (V_n)_n, (W_n)_n$ of finite words such that:

- (i) for every $n \geq 1$, $W_n U_n V_n U_n$ (respectively, $W_n U_n V_n \widetilde{U}_n$) is a prefix of ω ;
- (ii) for every $n \geq 1$, $\max(|V_n|/|U_n|, |W_n|/|U_n|) \leq c$;
- (iii) the sequence $(|U_n|)_n$ is increasing.

Notably, $\spadesuit$ words include all linear complexity words [19, Theorem. 1.1]. In particular, thanks to a theorem by Cobham [20], they include all automatic words, which are words on a finite alphabet that can be generated by a finite state machine. However, both classes of $\spadesuit$ words and $\clubsuit$ words include also words on infinite alphabets (see Appendix A for more details and examples).

In [1, 5, 6] the authors showed that, if $\omega = a_1 a_2 \dots$ has property $\spadesuit$ or $\clubsuit$, and assuming the extra condition that the limit superior of the sequence $(|W_n|/|U_n|)_n$ is *small enough*, then the real number $[0, a_1, a_2, \dots]$ is transcendental. This was generalised by Bugeaud [19, Theorem 1.3] to all $\spadesuit$ and $\clubsuit$ words as it follows:

Theorem 1.3 (Bugeaud) *Let $\alpha = [0, a_1, a_2, \dots]$, with $a_i \in \mathbb{Z}_{>0}$, and let A_n/B_n be the n th convergent of its continued fraction expansion. Suppose that the word $a_1 a_2 a_3 \dots$ satisfies condition $\spadesuit$ or $\clubsuit$, and that there exist a real $C > 1$ such that $|B_n| \leq C^n$. Then, α is either quadratic or transcendental.*

In this article, we consider the analogue of Theorem 1.3 in the p -adic setting, where p is any odd prime. A first naive attempt could be to simply replace $\mathbb{N}$ with $\mathbb{Z}[1/p]$ and $\mathbb{R}$ with $\mathbb{Q}_p$. However, this translation is far from straightforward for various reasons.

A structural key difficulty in the p -adic case lies in the absence of a *canonical* floor function $s : \mathbb{Q}_p \rightarrow \mathbb{Q}$. As we already noticed, there are infinitely many possible choices for s , and each leads to a distinct construction of p -adic continued fractions. In the last decades, many constructions of p -adic floor functions have been proposed, such as the classical ones by Ruban [34] and Browkin [18]. However, none of them gives approximants as good as in the real case and all of them lack complete analogues of fundamental results from the real setting, such as Lagrange’s theorem. More precisely, it is either not true or unknown, in general, that all rational numbers arise from finite words (except for very specific choices of s), nor that all algebraic numbers of degree 2 arise from periodic words.

These differences make Question 1.1 even more challenging in this p -adic setting. The only known results in this context provide negative answers to Question 1.1 only for very specific choices of floor functions and very special sets of words (see for instance [23, 29]).

In this article, we provide, under some mild conditions, a negative answer to Question 1.1 for $\spadesuit$ words and $\clubsuit$ words for any floor function, generalising most of the previous known results. More precisely, our main result is the following p -adic version of Theorem 1.3:

Theorem 1.4 *Let p be an odd prime, let $\alpha \in \mathbb{Q}_p$ and let s be a p -adic floor function. Suppose that $\alpha = [0, a_1, a_2, \dots]_s$ and let A_n/B_n be the n -th convergent of this continued fraction expansion. Suppose that the word $a_1 a_2 a_3 \dots$ satisfies either condition $\spadesuit$ or condition $\clubsuit$ with constant $c > 0$ and that for every $n \geq 1$:*

- (i) *the sequences $(|A_n|_\infty^{1/n})_n, (|B_n|_\infty^{1/n})_n, (|B_n|_p^{1/n})_n$ are bounded;*
- (ii) *$|a_n|_p \geq p^k$ where k is an explicit constant depending on p, c and the bound for the sequence $\left(\max\left(|A_n|_\infty^{1/n}, |B_n|_\infty^{1/n}\right)\right)_n$.*

Then, α is either rational, quadratic, or transcendental.

A more precise version of Theorem 1.4 with explicit constants can be found in Sect. 4 (more precisely, see Theorem 4.1 for $\spadesuit$ words and Theorem 4.3 for $\clubsuit$ words). As a first corollary, we have the following result (proved in Sect. 4.2.1) concerning the classical Ruban's and Browkin's continued fractions (see Sect. 3.2.1 for more details):

Corollary 1.5 *Let p be an odd prime and let s be a p -adic floor function. Let $\alpha = [0, a_1, a_2, \dots]_s$ and suppose that the sequence $(|B_n|_p^{1/n})_n$ is bounded, where B_n is defined in (2). Moreover, suppose that one of the following holds:*

- (a) *the word $a_1 a_2 a_3 \dots$ starts with arbitrarily long repetitions and either*
 - *s is Ruban's floor function, α is irrational and $|a_n|_p \geq p^4$ for every n , or*
 - *s is Browkin's floor function and $|a_n|_p \geq p^3$ for every n .*
- (b) *the word $a_1 a_2 a_3 \dots$ starts with arbitrarily long palindromes and either*
 - *s is Ruban's floor function, α is irrational and $|a_n|_p \geq p^2$ for every n , or*
 - *s is Browkin's floor function.*

Then, α is either quadratic or transcendental.

In another direction, we also have the following corollary (proved in Sect. 4.2.2) on continued fractions arising from automatic sequences:

Corollary 1.6 *Let n, m be non zero distinct integers. Let $\omega = a_1 a_2 \dots$ be an automatic word on the alphabet $\mathcal{A} = \{\frac{n}{p}, \frac{m}{p}\}$ where p is an odd prime. Then, for every p large enough, there is a p -adic floor function s such that $\mathcal{A} \subset \text{Im}(s)$ and the number $[0, a_1, a_2, \dots]_s$ is either rational, quadratic or transcendental.*

Our proof strategy for Theorem 1.4 is an adaptation of that of Bugeaud [19, Theorem 1.3]. There, the main tool used is Schmidt's subspace theorem, which is a generalization of fundamental results by Liouville, Thue, and Roth, on the approximation of algebraic numbers by means of sequences of rational numbers (we refer to [14, Ch. 6–7] for an overview). The basic idea is that, if a number is approximated *too well* by infinitely many rational numbers, then it must be transcendental. This already gives an intuition, in the real case, of why words containing arbitrarily large repeated prefixes fail to yield an algebraic number, given that

infinitely many of its convergents are *good approximations* of some periodic words (i.e. algebraic numbers of degree 2). In our non-archimedean world, this general philosophy still applies, but one needs some extra work and a p -adic version of the subspace theorem proved by Schlikewei [36, 37] (see Sect. 5 for details).

More specifically, the main idea underlying the proof of Theorem 1.4 is to leverage the $\spadesuit$ (resp. $\clubsuit$) structure of α to approximate it by a sequence of quadratic (resp. rational) numbers α_n having a periodic (resp. palindromic) continued fraction expansion of the form $[W_n \overline{U_n V_n}]_s$ (resp. $[W_n U_n V_n \widetilde{U_n}]_s$). Then α and α_n have the same $|W_n| + 2|U_n| + |V_n|$ -th convergent, so that the quality of approximation is very good relatively to the arithmetic complexity (more precisely, the Weil height, cf. Definition 5.1) of α_n which only depends on the $|W_n| + |U_n| + |V_n|$ -th continuants. Thus, assuming α algebraic, one can construct suitable linear forms and a sequence of vectors depending on α_n that make the p -adic Subspace Theorem (Theorem 5.2) applicable. Then one gets a linear relation between some continuant of the continued fraction for α . Finally, a brute force approach case by case, allows to transform it into a linear relation between 1, α , α^2 .

The structure of our article is as follows. In Sect. 2, we recall a result of Bugeaud linking word complexity and condition $\spadesuit$. In Sect. 4, we present explicit versions of Theorem 1.4 (namely, Theorem 4.1 for $\spadesuit$ words and Theorem 4.3 for $\clubsuit$ words) along with applications, including proofs of Corollary 1.5 and Corollary 1.6. Section 5 provides a statement of the p -adic Subspace Theorem, a key ingredient in our proofs. In Sect. 6 we prove Theorem 4.1 and in Sect. 7 we prove Theorem 4.3. Finally, in Appendix A, we examine $\spadesuit$ and $\clubsuit$ words in detail, exploring their connections with well-known classes of words, such as those arising from automatic and Sturmian sequences, and providing concrete examples.

2 Some results on word complexity

The goal of this short section is to recall a result of Bugeaud [19] relating word complexity and condition $\spadesuit$.

Recall that a *word* ω is a (possibly infinite) string of symbols from a non-empty set $\mathcal{A}$, called *alphabet*. For a finite word ω , we recall that we denote its length by $|\omega|$ and we denote by $\widetilde{\omega}$ the word obtained by reversing the order of the letters of ω . We also denote by $\overline{\omega}$ the infinite purely periodic word of period ω .

The *complexity function* of a word $\omega = a_1 a_2 a_3 \dots$ is the function $p_\omega : \mathbb{N} \rightarrow \mathbb{N}$ defined as

$$p_\omega(n) = \#\{a_{\ell+1} \dots a_{\ell+n} \mid \ell \geq 0\},$$

that is, $p_\omega(n)$ is the number of distinct blocks of length n appearing in the writing of ω .

For example, the periodic word $\omega = \overline{011}$ has $p_\omega(1) = 2$, $p_\omega(n) = 3$ for all $n \geq 2$. More generally, a classical result in [27, Thm. 7.3] shows that either ω is ultimately periodic, in which case the function p_ω is bounded, or $p_\omega(n) \geq n + 1$ for all $n \geq 1$.

In view of this result and linking to our study of continued fraction, we have that, if the alphabet is $\mathbb{N}$, the word complexity of an infinite word $\omega = a_1 a_2 a_3 \dots$ gives information on the nature of the real number $\alpha = [0, a_1, a_2, \dots]$.

Indeed, if $p_\omega(n) \leq n$ for some $n \geq 1$, then ω is ultimately periodic, hence, by Lagrange theorem, α is algebraic of degree 2. On the other hand, Bugeaud [19, Theorem 1.1] proves that if ω is not ultimately periodic and α is algebraic, then $p_\omega(n)$ cannot grow *too slowly* i.e. one must have

$$\lim_{n \rightarrow \infty} \frac{p_\omega(n)}{n} = \infty. \quad (3)$$

A first step in his proof is to show that *low complexity* words satisfy condition ♠. More precisely (see [19, Proof of Theorem 1.1]):

Theorem 2.1 *Let $(a_i)_{i \geq 1}$ be a sequence of strictly positive integers which is not ultimately periodic and consider the word $\omega = a_1 a_2 a_3 \dots$. Suppose that there exists an infinite set $\mathcal{N} \subset \mathbb{N}$ and a positive real $C > 0$ such that $p_\omega(n) \leq nC$ for all $n \in \mathcal{N}$. Then ω satisfies condition ♠ with constant $c = 3C + 1$.*

To the best of the authors' knowledge, there is no direct analogue of Theorem 2.1 for words satisfying condition ♣. In particular, it remains unclear whether this set contains specific complexity classes.

3 Toolbox for the p -adic case

In this section we recall some basic properties of the p -adic continued fraction algorithms that will be used later. In what follows, we will fix a prime $p > 2$ and we will denote by $|\cdot|_p$ and $|\cdot|_\infty$ the p -adic and the Euclidean norm respectively and by v_p the usual p -adic valuation.

3.1 p -adic floor functions

There are several ways to construct continued fraction algorithms over $\mathbb{Q}_p$, and several authors as Mahler [25], Schneider [35], Ruban [34] and Browkin [17, 18] proposed different algorithms in the attempt of achieving the same properties which hold in the real case. All these constructions are based on the choice of a so called *p -adic floor function* $s : \mathbb{Q}_p \rightarrow \mathbb{Z}[1/p]$. We give the precise definition (which can also be found in [18, Section 1]).

Definition 3.1 A *p -adic floor function* is any function $s : \mathbb{Q}_p \rightarrow \mathbb{Z}[1/p]$ such that

- (i) $|\alpha - s(\alpha)|_p < 1$ for every $\alpha \in \mathbb{Q}_p$;
- (ii) $s(0) = 0$;
- (iii) if $|\alpha - \beta|_p < 1$, then $s(\alpha) = s(\beta)$.

Given $\alpha \in \mathbb{Q}_p$ and a floor function s , the continued fraction expansion $[a_0, a_1, \dots]_s$ of α is obtained via the algorithm:

$$\begin{cases} \gamma_0 &= \alpha, \\ a_n &= s(\gamma_n), \\ \gamma_{n+1} &= \frac{1}{\gamma_n - a_n} \text{ if } \gamma_n - a_n \neq 0. \end{cases}$$

and it stops if $\gamma_n = a_n$. Notice that, by definition, $|a_i|_p > 1$ for every $i \geq 1$.

It is proved in [18, Section 2, Theorem 2] that, for any choice of s , the map sending α to the partial quotients of its continued fraction expansion is a bijection between $\mathbb{Q}_p$ and the (possibly infinite) sequences $a_0, a_1, a_2, \dots$ with $a_i \in \text{Im}(s)$ and $|a_i|_p > 1$ for $i \geq 1$. Since $a_0 = 0$ if and only if $|\alpha|_p < 1$ (this follows directly from Definition 3.1), such map restricts to a bijection between $p\mathbb{Z}_p$ and the (possibly infinite) sequences $0, a_1, a_2, \dots$ with $a_i \in \text{Im}(s)$ and $|a_i|_p > 1$ for $i \geq 1$.

3.2 Some useful estimates on the continuants

Given a floor function s and a continued fraction

$$[a_0, a_1, a_2, \dots]_s$$

we define, analogously to the real case, the classic sequences $(A_n)_{n \geq -1}$, $(B_n)_{n \geq -1}$ of *continuants* given by the recurrences (2). We have

$$\frac{A_n}{B_n} = [a_0, \dots, a_n]_s. \quad (4)$$

Notice that A_n, B_n only depend on $a_0, \dots, a_n$.

Using matrices, we can write

$$\mathcal{A}_n = \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathcal{B}_n = \begin{pmatrix} A_n & A_{n-1} \\ B_n & B_{n-1} \end{pmatrix},$$

for $n \in \mathbb{N}$. Moreover, if we define

$$\mathcal{B}_{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we have

$$\mathcal{B}_n = \mathcal{B}_{n-1} \mathcal{A}_n = \mathcal{A}_0 \mathcal{A}_1 \dots \mathcal{A}_n. \quad (5)$$

This easily implies by induction the useful relation

$$A_n B_{n-1} - B_n A_{n-1} = (-1)^{n+1} \quad \text{for every } n \geq 0. \quad (6)$$

Moreover, for every $k \geq 1$ we have

$$\alpha = [a_0, \dots, a_{k-1}, \gamma]_s = \frac{\gamma A_{k-1} + A_{k-2}}{\gamma B_{k-1} + B_{k-2}}. \quad (7)$$

We will also use the following properties.

Lemma 3.2 *Let s be a p -adic floor function, $\alpha = [0, a_1, a_2, \dots]_s$ and let $(A_n)_n, (B_n)_n$ be the sequences of continuants of α . Then:*

- (i) *for every $n \geq 2$, $|A_n|_p = \prod_{i=2}^n |a_i|_p$, $|B_n|_p = \prod_{i=1}^n |a_i|_p$.*
- (ii) *for every $n \geq 0$, $|A_n|_p < |A_{n+1}|_p$, $|B_n|_p < |B_{n+1}|_p$, $|A_n|_p \leq |B_n|_p$.*

Suppose moreover that there exists a real number $M > 0$ such that, for every i , $|a_i|_\infty \leq M$. Then

$$\max(|A_n|_\infty, |B_n|_\infty) \leq (M+1)^n. \quad (8)$$

Proof Notice that

$$|A_0|_p = 0, |A_1|_p = 1, |A_2|_p = |a_2|_p, |A_3|_p = \max(|a_3|_p |a_2|_p, 1) = |a_3|_p |a_2|_p$$

and

$$|B_0|_p = 1, |B_1|_p = |a_1|_p, |B_2|_p = |a_1|_p |a_2|_p, |B_3|_p = |a_1|_p |a_2|_p |a_3|_p.$$

Notice that, as $|a_i|_p > 1$, one can show by induction that, for every $n \geq 2$, $|A_{n-2}|_p < |a_n A_{n-1}|_p$ and similarly $|B_{n-2}|_p < |a_n B_{n-1}|_p$. This, together with the ultrametric inequality, proves (i), which easily implies (ii). Finally, (8) also easily follows by induction. $\square$

The proof of the following useful lemma is obtained combining [18, Theorem 1] and [18, Theorem 2]:

Lemma 3.3 *Let s be a p -adic floor function, $\alpha = [a_0, a_1, a_2, \dots]_s$ and let $(A_n)_n, (B_n)_n$ be the sequences of continuants of α . Then for $n \geq 0$*

$$|B_n \alpha - A_n|_p = \prod_{j=1}^{n+1} \frac{1}{|a_j|_p} = \frac{1}{|B_{n+1}|_p} \leq \frac{1}{p^{n+1}}.$$

In particular

$$\left| \alpha - \frac{A_n}{B_n} \right|_p = \frac{1}{|B_n B_{n+1}|_p} \leq \frac{1}{|B_n|_p^2}. \quad (9)$$

Notice that, by (ii) of Lemma 3.2, for every $n \geq 0$, $|B_n|_p < |B_{n+1}|_p$, we have that the sequence $|\alpha - \frac{A_n}{B_n}|_p \rightarrow 0$ as n tends to ∞ ; then, $\frac{A_n}{B_n}$ converges p -adically to α .

The last lemma shows that, if the n th convergents of α and β are the same, then the two numbers are p -adically close. We have then the following result.

Lemma 3.4 *Let s be a p -adic floor function, $\alpha = [a_0, a_1, \dots, a_n, a_{n+1} \dots]_s$. Suppose that $\beta = [a_0, a_1, \dots, a_n, b_{n+1}, \dots]_s$. Then α and β have the same first $n + 1$ continuants $A_0, \dots, A_n, B_0, \dots, B_n$ and*

$$|\alpha - \beta|_p < \frac{1}{|B_n|_p^2}.$$

Proof By (4) and (9) we have

$$\left| \alpha - \frac{A_n}{B_n} \right|_p < \frac{1}{|B_n|_p^2} \quad \text{and} \quad \left| \beta - \frac{A_n}{B_n} \right|_p < \frac{1}{|B_n|_p^2}.$$

By adding and subtracting $\frac{A_n}{B_n}$ and applying the ultrametric inequality we obtain the conclusion. $\square$

3.2.1 Estimates for Ruban's and Browkin's floor functions

The two main definitions of p -adic continued fraction algorithms are due to Ruban [34] and Browkin [18]: in these cases the p -adic floor functions are given by

$$s(\alpha) = \sum_{n=k}^0 x_n p^n \in \mathbb{Q} \quad \text{where} \quad \alpha = \sum_{n=k}^{+\infty} x_n p^n \in \mathbb{Q}_p,$$

where $k \leq 0$ is an integer, the x_n are representatives modulo p in the interval $[0, p - 1]$ for Ruban's and in the interval $[-(p - 1)/2, (p - 1)/2]$ for Browkin's definition. Notice that, for these two floor functions, the partial quotients are bounded in absolute value by a constant M , which is p in Ruban's case and $p/2$ in Browkin's. By (8), this implies that $\max(|A_n|_\infty, |B_n|_\infty) \leq (M + 1)^n$.

3.3 Periodic and palindromic continued fractions

The following result is well known (see e.g. [33, Th. III.1, p. 40; 32, Remark 6.1]). Since it is widely used in our proofs, we reproduce here the proof for clarity in our case where $a_0 = 0$. We recall that, given a finite word ω , we denote by $\overline{\omega}$ the infinite purely periodic word obtained repeating ω .

Lemma 3.5 *Let s be a p -adic floor function and let*

$$\alpha = [0, a_1, \dots, a_w, \overline{a_{w+1}, \dots, a_\ell}]_s.$$

Let $(A_n)_{n \geq -1}, (B_n)_{n \geq -1}$ be the sequences of continuants of α . Then α is a root of the polynomial

$$P(X) = aX^2 - bX + c \quad (10)$$

where

$$\begin{aligned} a &= B_{w-1}B_\ell - B_wB_{\ell-1} \\ b &= B_{w-1}A_\ell - B_wA_{\ell-1} + A_{w-1}B_\ell - A_wB_{\ell-1} \\ c &= A_{w-1}A_\ell - A_wA_{\ell-1}. \end{aligned} \quad (11)$$

Proof Let $\beta = [\overline{a_{w+1}, \dots, a_\ell}]_s$. Then

$$\alpha = \frac{A_w\beta + A_{w-1}}{B_w\beta + B_{w-1}}$$

but also

$$\alpha = \frac{A_\ell\beta + A_{\ell-1}}{B_\ell\beta + B_{\ell-1}}.$$

Therefore α is a fixed point of the linear fractional transformation defined by the matrix

$$\begin{pmatrix} A_\ell & A_{\ell-1} \\ B_\ell & B_{\ell-1} \end{pmatrix} \begin{pmatrix} A_w & A_{w-1} \\ B_w & B_{w-1} \end{pmatrix}^{-1} = \pm \begin{pmatrix} A_\ell B_{w-1} - A_{\ell-1} B_w & -A_\ell A_{w-1} + A_{\ell-1} A_w \\ B_\ell B_{w-1} - B_{\ell-1} B_w & -B_\ell A_{w-1} + B_{\ell-1} A_w \end{pmatrix}.$$

The claim follows from (11). $\square$

The following result establishes a relation between the continuants of a continued fraction and its palindrome.

Lemma 3.6 *Let s be a p -adic floor function, let $\alpha = [a_0, a_1, \dots, a_n]_s$ and let $(A_n)_n, (B_n)_n$ be the sequences of continuants of α . If $|a_0|_p > 1$, then*

$$\frac{B_n}{B_{n-1}} = [a_n, \dots, a_0]_s.$$

If $a_0 = 0$, we have

$$\frac{B_{n-1}}{B_n} = [0, a_n, \dots, a_1]_s.$$

Proof This can be easily checked by induction using the definition of B_n from (2). $\square$

4 Our main results and some applications

In this section we give an explicit version of our main result Theorem 1.4. We state separately the part dealing with property $\spadesuit$ (see Theorem 4.1) and the part dealing with property $\clubsuit$ (see Theorem 4.3).

4.1 An explicit version of Theorem 1.4

Theorem 4.1 *Let p be an odd prime and let s be a p -adic floor function. Let $\alpha = [0, a_1, a_2, \dots]_s$, and let A_n, B_n be its n -th continuants.*

Suppose that the word $a_1 a_2 a_3 \dots$ satisfies condition ♠ with constant c and that there exist reals $C_\infty, C_p > 1$ such that, for every $n \geq 1$:

- (i) $\max(|A_n|_\infty, |B_n|_\infty) \leq C_\infty^n$;
- (ii) $|B_n|_p \leq C_p^n$;
- (iii) $|a_n|_p \geq p^k$ where $k = \left\lfloor \max \left(3 \frac{\log C_\infty}{\log p}, 2(3c + 1) \frac{\log C_\infty}{\log p} \right) \right\rfloor + 1$.

Then α is either rational, quadratic, or transcendental.

Remark 4.2 Unlike the archimedean case, it can actually occur that α is rational in our statement.

For example, let s be the floor function defined by Ruban (see Sect. 3.2.1). Then, as remarked in [22, §3], one has

$$-p = \left[0, \frac{p-1}{p} + (p-1) \right]_s.$$

So, the associated word satisfies condition ♠ with constant $c = 0$ and the conditions of Theorem 4.1 are satisfied (with $C_\infty = 2$, $C_p = p$ and $k = 1$).

Theorem 4.3 *Let p be an odd prime. Let s be a p -adic floor function. Let $\alpha = [0, a_1, a_2, \dots]_s$, and let A_n, B_n be its n -th continuants.*

Suppose that the word $a_1 a_2 a_3 \dots$ satisfies condition ♣ with constant c and that there exist reals $C_\infty, C_p > 1$ such that, for every $n \geq 1$:

- (i) $\max(|A_n|_\infty, |B_n|_\infty) \leq C_\infty^n$;
- (ii) $|B_n|_p \leq C_p^n$;
- (iii) for every n , $|a_n|_p \geq p^k$ where

$$k = \begin{cases} \left\lfloor \frac{\log C_\infty}{\log p} \right\rfloor + 1 & \text{if } c = 0, \\ \left\lfloor (4 + 6c) \frac{\log C_\infty}{\log p} \right\rfloor + 1 & \text{if } c \neq 0. \end{cases}$$

Then α is either rational, quadratic, or transcendental.

Remark 4.4 Notice that:

- For both Theorems 4.1 and 4.3, condition (ii) implies the same bound also on $|A_n|_p$, since $|A_n|_p \leq |B_n|_p$ by Lemma 3.2. However, this does not generally hold for the Archimedean absolute value, so we need to impose both bounds in condition (i).
- Moreover, if there exists $T > 0$ such that $|a_i|_\infty \leq T$ for every i , as it happens in most definitions of p -adic continued fractions, condition (i) from Theorems 4.1 and 4.3 always holds true. Indeed, by Lemma 3.2, one can take $C_\infty = T + 1$.

4.2 Some corollaries of our results

4.2.1 Ruban's and Browkin's floor functions: proof of Corollary 1.5

Since the word $a_1 a_2 a_3 \dots$ starts with arbitrarily long repetitions (resp. palindromes), it satisfies condition ♠ (resp. ♣) with $c = 0$. It is then enough to check that the hypotheses of

Theorem 4.1 (resp. Theorem 4.3) are satisfied. Note that infinite continued fraction expansions are always irrational in Browkin's case [18, Thm. 3], so the irrationality of α needs to be assumed only in Ruban's case.

As observed in Sect. 3.2.1, hypothesis (i) of Theorem 4.1 (resp. Theorem 4.3) is automatically satisfied by setting $C_\infty = p + 1$ when s is Ruban's floor function, and $C_\infty = p/2 + 1$ when s is Browkin's floor function, applying Lemma 3.2. Hypothesis (ii) of Theorem 4.1 (resp. Theorem 4.3) holds by assumption. Finally, hypothesis (iii) of Theorem 4.1 (resp. Theorem 4.3) can be easily checked, case by case, by setting $c = 0$ and using the respective values of C_∞ . This concludes the proof of Corollary 1.5.

4.2.2 Words on a finite alphabet

For words on a finite alphabet, hypotheses (i) and (ii) of Theorems 4.1 and 4.3 are always fulfilled. Therefore, our results can be specialized as follows.

Corollary 4.5 *Let p be an odd prime. Let s be a p -adic floor function, and $\alpha = [0, a_1, a_2, \dots]_s$. Suppose that the word $\omega = a_1 a_2 a_3 \dots$ consists of letters from a finite alphabet and set $T = \max(|a_n|_\infty)_n$. Moreover, suppose that ω satisfies one of the following:*

(1) *condition ♠ with constant c and*

$$(T + 1)^{\max(3, 6c+2)} < p, \quad (12)$$

(2) *condition ♣ with constant c and*

$$\begin{cases} T < p - 1 & \text{if } c = 0, \\ (T + 1)^{4+6c} < p & \text{otherwise.} \end{cases} \quad (13)$$

Then α is either rational, quadratic, or transcendental.

Proof It is enough to check that the conditions (i), (ii), and (iii) in Theorem 4.1 (resp. Theorem 4.3) are satisfied. By Remark 4.4, (i) is satisfied taking $C_\infty = T + 1$. From part (i) of Lemma 3.2, one sees that (ii) holds for $C_p = \max(|a_i|_p)_i$. Finally, (12) (resp. (13)) implies that $k = 1$ in (iii), which is then automatically satisfied since $a_i \in \text{Im}(s) \setminus \{0\}$. $\square$

The following result specializes Corollary 4.5 to low complexity words.

Corollary 4.6 *Let p be an odd prime. Let $a, b \in \mathbb{Z}[1/p]$ such that:*

- (i) $|a - b|_p \geq 1$;
- (ii) $\min(|a|_p, |b|_p) \geq p$.

Then, there exists a p -adic floor function s such that $\{a, b\} \subset \text{Im}(s)$.

Let s be such a floor function. Suppose moreover that $\omega = a_1 a_2 a_3 \dots$ is a word on the alphabet $\{a, b\}$ whose complexity function satisfies $p_\omega(n) \leq nC$ for all $n \geq 1$ where $C \geq 0$ is a real number such that:

- (iii) $(\max(|a|_\infty, |b|_\infty) + 1)^{18C+8} < p$.

Then $[0, a_1, a_2, \dots]_s$ is either rational, quadratic, or transcendental.

Proof Notice that, in general, the construction of a p -adic floor function s amounts to choosing a representative in $\mathbb{Z}[1/p]$ from any coset of $\mathbb{Q}_p$ modulo $p\mathbb{Z}_p$. Then the image of s on the given class is the chosen element. Conditions (i) and (ii) ensure that the cosets $a + p\mathbb{Z}_p$ and

$b + p\mathbb{Z}_p$ are distinct and non zero. Hence one can choose any floor function s mapping such cosets to a and b respectively (notice that there are infinitely many possible choices).

We fix such an s . If ω is ultimately periodic, then, by classical results on p -adic continued fractions (see Lemma 3.5), the number $[0, a_1, a_2, \dots]_s$ is either rational or quadratic. Otherwise, by Theorem 2.1, ω satisfies condition ♠ with constant $c = 3C + 1$. Therefore the hypotheses of Corollary 4.5—in particular (i), which in this case is equivalent to (iii)—are satisfied, and the conclusion follows immediately. $\square$

Finally, Corollary 1.6 is a special case of the above result.

Proof of Corollary 1.6 For p large enough, $n, m, n - m$ are coprime to p . This is equivalent to saying that n/p and m/p satisfy hypotheses (i) and (ii) of Corollary 4.6. Since also (iii) must be satisfied up to taking a larger p , the conclusion follows directly from Corollary 4.6. $\square$

Remark 4.7 Notice that condition (iii) of Corollary 4.6 can be made explicit, or even improved, for some families of words (see Sect. A.3 of Appendix. A).

5 A key tool: the p -adic subspace theorem

The proof of our main results requires a slightly more general version of the subspace theorem than the one by Schmidt used in [19], in order to include p -adic places. We recall the classical definition of Weil height (see for instance [14, §1.5.6]). For our purposes, we restrict our attention to the case of an affine point in $\mathbb{Z}[1/p]^N$.

Definition 5.1 Let $\mathbf{z}$ be a vector $(z_1, z_2, \dots, z_N) \in \mathbb{Z}[1/p]^N$. The (multiplicative) Weil height of $\mathbf{z}$ is

$$H(\mathbf{z}) = \max(1, \max_i(|z_i|_\infty)) \cdot \max(1, \max_i(|z_i|_p)).$$

We now state the generalisation of Schmidt's subspace theorem, due to Schlickewei [36, 37], which can be obtained, for instance, as easy consequence of [15, Theorem 2.3']. We refer to [15] for a comprehensive overview on the subject.

Theorem 5.2 Let $N \geq 1$ be an integer, let S be the set containing the archimedean and the p -adic place of $\mathbb{Q}$. For each $v \in S$, let $\{L_{1,v}, \dots, L_{N,v}\}$ be a set of linearly independent linear forms in the variables $X_1, \dots, X_N$ with algebraic coefficients in $\mathbb{Q}_v$. Let $(\mathbf{z}_n)_{n \geq 1}$ be a sequence of vectors in $\mathbb{Z}[1/p]^N$ and set $\mathbf{z}_n = (z_{n,1}, z_{n,2}, \dots, z_{n,N})$. Assume that there exist strictly positive reals C_1, C_2, A, B with $A < 1 \leq B$ and an increasing sequence of positive integers $(u_n)_{n \geq 1}$ such that

$$\prod_{v \in S} \prod_{i=1}^N |L_{i,v}(\mathbf{z}_n)|_v \leq C_1 \cdot A^{u_n}$$

$$H(\mathbf{z}_n) \leq C_2 \cdot B^{u_n}.$$

Then there exist $x_1, \dots, x_N \in \mathbb{Z}$, not all zero, and an infinite subset $\mathcal{N} \subseteq \mathbb{N}$ such that, for every $j \in \mathcal{N}$, one has

$$x_1 z_{j,1} + x_2 z_{j,2} + \dots + x_N z_{j,N} = 0.$$

Proof Up to normalizing the coefficients of the linear forms and increasing the constant B , we can assume that

$$\prod_{v \in S} \prod_{i=1}^N |L_{i,v}(\mathbf{z}_n)|_v \leq A^{u_n} \text{ and } H(\mathbf{z}_n) \leq B^{u_n}.$$

Let $\epsilon > 0$ be such that $A < 1/B^\epsilon$. Then,

$$\prod_{v \in S} \prod_{i=1}^N |L_{i,v}(\mathbf{z}_n)|_v \leq A^{u_n} < \frac{1}{B^{u_n \epsilon}} \leq \frac{1}{H(\mathbf{z}_n)^\epsilon}.$$

Then the result follows from [15, Theorem 2.3'], by noticing that $H(\mathbf{z}_n)$ is larger or equal than the height of $\mathbf{z}_n$ as defined in [15, p.5, formula (3)]. $\square$

In what follows, given two real sequences $(t_n)_{n \geq 1}$ and $(s_n)_{n \geq 1}$, we will use the notation $t_n \ll s_n$ to indicate that there exists a positive constant C such that $t_n \leq C s_n$ for every $n \in \mathbb{N}$.

6 Proof of Theorem 4.1

Our proof strategy closely follows the one of [19, Thm. 3.1].

Before proving the result we fix some notation. By hypothesis the sequence $(a_n)_{n \geq 1}$ satisfies condition $\spadesuit$ with constant c . Hence, there exist three sequences of finite words $(W_n)_n, (U_n)_n, (V_n)_n$ satisfying conditions (i), (ii) and (iii) of Definition 1.2.

Setting $w_n = |W_n|$, $u_n = |U_n|$, and $v_n = |V_n|$, we have that, for every $n \geq 0$, $u_{n+1} \geq u_n$ and

$$\max(v_n/u_n, w_n/u_n) \leq c. \quad (14)$$

We let also

$$\ell_n = w_n + u_n + v_n.$$

We assume by contradiction that α is algebraic of degree at least 3 over $\mathbb{Q}$.

6.1 A preliminary lemma

The proof of Theorem 4.1 is based on several applications of the p -adic Subspace Theorem (see Theorem 5.2), the first one of which is used to prove the following lemma.

Lemma 6.1 Let A_n and B_n be as in Theorem 4.1 and ℓ_n, w_n as above. Then, there exist a non zero quadruple $(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4$ and an infinite set $\mathcal{N}_1 \subseteq \mathbb{N}$ such that

$$\begin{aligned} & x_1(B_{w_n-1}B_{\ell_n} - B_{w_n}B_{\ell_n-1}) + x_2(B_{w_n-1}A_{\ell_n} - B_{w_n}A_{\ell_n-1}) + \\ & + x_3(A_{w_n-1}B_{\ell_n} - A_{w_n}B_{\ell_n-1}) + x_4(A_{w_n-1}A_{\ell_n} - A_{w_n}A_{\ell_n-1}) = 0 \end{aligned} \quad (15)$$

for any $n \in \mathcal{N}_1$.

Proof To simplify the notation, we set

$$\mathbf{z}_n = (z_{n,1}, z_{n,2}, z_{n,3}, z_{n,4}),$$

where

$$\begin{aligned} z_{n,1} &= B_{w_n-1} B_{\ell_n} - B_{w_n} B_{\ell_n-1} \\ z_{n,2} &= B_{w_n-1} A_{\ell_n} - B_{w_n} A_{\ell_n-1} \\ z_{n,3} &= A_{w_n-1} B_{\ell_n} - A_{w_n} B_{\ell_n-1} \\ z_{n,4} &= A_{w_n-1} A_{\ell_n} - A_{w_n} A_{\ell_n-1}. \end{aligned} \quad (16)$$

We now want to apply the Subspace Theorem 5.2 to the sequence $(\mathbf{z}_n)_n$ with carefully chosen linear forms.

To this aim, for $1 \leq i \leq 4$ we choose the linear forms

$$L_{i,\infty}(X_1, X_2, X_3, X_4) = X_i$$

and

$$\begin{aligned} L_{1,p}(X_1, X_2, X_3, X_4) &= \alpha^2 X_1 - \alpha(X_2 + X_3) + X_4, \\ L_{2,p}(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_2, \\ L_{3,p}(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_3, \\ L_{4,p}(X_1, X_2, X_3, X_4) &= X_1. \end{aligned}$$

We now want to estimate the quantities $|L_{i,\infty}(\mathbf{z}_n)|_\infty$ and $|L_{i,p}(\mathbf{z}_n)|_p$.

To this purpose, we set $\alpha_n = [W_n U_n V_n]_s$. By Lemma 3.5, α_n is a root of the following polynomial of degree 2

$$P_n(X) = z_{n,1} X^2 - (z_{n,2} + z_{n,3})X + z_{n,4}.$$

Moreover, by (9),

$$\left| \alpha - \frac{A_n}{B_n} \right|_p = \frac{1}{|B_n B_{n+1}|_p}$$

and

$$\left| \alpha_n - \frac{A_{\ell_n+u_n}}{B_{\ell_n+u_n}} \right|_p = \frac{1}{|B_{\ell_n+u_n} B_{\ell_n+u_n+1}|_p}.$$

Thus,

$$|\alpha - \alpha_n|_p \leq \frac{1}{|B_{\ell_n+u_n} B_{\ell_n+u_n+1}|_p}. \quad (17)$$

Notice that, by Lemma 3.2 and the fact that $|a_1|_p > 1$, for every $m \geq 2$, we have

$$|A_m|_p = \prod_{i=2}^m |a_i|_p < |A_{m+1}|_p, \quad |B_m|_p = \prod_{i=1}^m |a_i|_p < |B_{m+1}|_p, \quad |A_m|_p < |B_m|_p.$$

Moreover, for every $n \geq 1$, we have

$$\max_{1 \leq j \leq 4} |z_{n,j}|_p \leq |B_{w_n} B_{\ell_n}|_p. \quad (18)$$

By (9) and Lemma 3.2 (ii), we have

$$\begin{aligned} |L_{2,p}(\mathbf{z}_n)|_p &= |(B_{w_n-1} B_{\ell_n} - B_{w_n} B_{\ell_n-1})\alpha - (B_{w_n-1} A_{\ell_n} - B_{w_n} A_{\ell_n-1})|_p \\ &\leq \max(|B_{w_n-1}|_p |B_{\ell_n} \alpha - A_{\ell_n}|_p, |B_{w_n}|_p |B_{\ell_n-1} \alpha - A_{\ell_n-1}|_p) \\ &\leq |B_{w_n}|_p |B_{\ell_n}|_p^{-1} \end{aligned} \quad (19)$$

and

$$\begin{aligned} |L_{3,p}(\mathbf{z}_n)|_p &= |(B_{w_n-1}B_{\ell_n} - B_{w_n}B_{\ell_n-1})\alpha - (A_{w_n-1}B_{\ell_n} - A_{w_n}B_{\ell_n-1})|_p \\ &\leq \max(|B_{\ell_n}|_p|B_{w_n-1}\alpha - A_{w_n-1}|_p, |B_{\ell_n-1}|_p|B_{w_n}\alpha - A_{w_n}|_p) \\ &\leq |B_{\ell_n}|_p|B_{w_n}|_p^{-1}. \end{aligned} \quad (20)$$

Also, by (17), (18), (19) and (20), we have

$$\begin{aligned} |L_{1,p}(\mathbf{z}_n)|_p &= |P_n(\alpha)|_p = |P_n(\alpha) - P_n(\alpha_n)|_p = |z_{n,1}(\alpha^2 - \alpha_n^2) - (z_{n,2} + z_{n,3})(\alpha - \alpha_n)|_p \\ &= |(\alpha - \alpha_n)|_p \cdot |(z_{n,1}\alpha - z_{n,2}) + (z_{n,1}\alpha_n - z_{n,3})|_p = \\ &= |(\alpha - \alpha_n)|_p \cdot |L_{2,p}(\mathbf{z}_n) + L_{3,p}(\mathbf{z}_n) + z_{n,1}(\alpha_n - \alpha)|_p \\ &\leq |\alpha - \alpha_n|_p \max(|B_{w_n}|_p|B_{\ell_n}|_p^{-1}, |B_{w_n}|_p^{-1}|B_{\ell_n}|_p, |B_{w_n}|_p|B_{\ell_n}|_p|\alpha - \alpha_n|_p) \\ &\leq |B_{w_n}|_p^{-1}|B_{\ell_n}|_p|B_{\ell_n+u_n}|_p^{-2}. \end{aligned} \quad (21)$$

Moreover, by (18), we have

$$|L_{4,p}(\mathbf{z}_n)|_p = |z_{n,1}|_p \leq |B_{w_n}B_{\ell_n}|_p. \quad (22)$$

By the above formulas (19), (20), (21), (22), we have that

$$\prod_{j=1}^4 |L_{j,p}(\mathbf{z}_n)|_p \leq |B_{\ell_n}|_p^2 |B_{\ell_n+u_n}|_p^{-2}.$$

Since, by the Assumption (iii) of Theorem 4.1,

$$|B_{\ell_n+u_n}|_p = |B_{\ell_n}|_p \prod_{i=\ell_n+1}^{\ell_n+u_n} |a_i|_p \geq |B_{\ell_n}|_p \cdot p^{ku_n},$$

and, by (i),

$$\max_{1 \leq j \leq 4} |z_{n,j}|_\infty \leq 2C_\infty^{2w_n+u_n+v_n} \leq 2C_\infty^{(3c+1)u_n}, \quad (23)$$

we find

$$\prod_{j=1}^4 |L_{j,p}(\mathbf{z}_n)|_p \prod_{j=1}^4 |z_{n,j}|_\infty \leq 16 \left(\frac{C_\infty^{4(3c+1)}}{p^{2k}} \right)^{u_n}. \quad (24)$$

Using (18), (23), and assumption (ii) of Theorem 4.1, we also have

$$H(\mathbf{z}_n) \leq 2C_\infty^{(3c+1)u_n} |B_{w_n}|_p |B_{\ell_n}|_p \leq 2 \cdot ((C_\infty C_p)^{3c+1})^{u_n}. \quad (25)$$

Then, by Theorem 5.2, the conclusion follows provided that the quantity $\frac{C_\infty^{4(3c+1)}}{p^{2k}}$ in (24) is < 1 , and this is in fact the case by hypothesis (iii) of Theorem 4.1. $\square$

6.2 Splitting the proof in two cases

We are now ready to prove Theorem 4.1. By Lemma 6.1, we may consider only the indices in the subset $\mathcal{N}_1$ so that (15) is satisfied.

Here the proof is split in two cases, depending on whether the sequence $(w_n)_{n \in \mathcal{N}_1}$ has a constant subsequence or not.

In the first case, using the continuants of α , Lemma 6.1 and the Subspace theorem, we construct an irrational number β such that both $(1, \alpha, \beta)$ and $(1, \alpha, 1/\beta)$ are linearly dependent over $\mathbb{Z}$. Combining these linear relations, we obtain a contradiction.

In the second case, using Lemma 6.1 and suitable auxiliary words, we show that α is a root of a certain polynomial of degree at most 2, which must therefore be trivial. The vanishing of its coefficients gives relations among the continuants of α and these, together with the Subspace theorem, allow to conclude.

Case (1): $(w_n)_{n \in \mathcal{N}_1}$ has a constant subsequence

Let w be such that $w_n = w$ for infinitely many $n \in \mathcal{N}_1$. By extracting an infinite subset of $\mathcal{N}_1$ and replacing α by

$$[0, a_{w+1}, a_{w+2}, \dots]_s,$$

we may without loss of generality assume that $w_n = 0$ and hence $\ell_n = u_n + v_n$ for every $n \in \mathcal{N}_1$. Let now

$$\beta = \lim_{\substack{n \rightarrow \infty \\ n \in \mathcal{N}_1}} \frac{B_{\ell_n-1}}{B_{\ell_n}}.$$

We claim that:

- (a) β is irrational;
- (b) there is a non zero triple of integers (y_1, y_2, y_3) such that

$$y_1 + y_2\beta + y_3\alpha = 0; \quad (26)$$

- (c) there is a non zero triple of integers (z_1, z_2, z_3) such that

$$\frac{z_1}{\beta} + z_2 + z_3\alpha = 0. \quad (27)$$

Assuming these claims, from (26) and (27) we obtain

$$(z_3\alpha + z_2)(y_3\alpha + y_1) = y_2z_1.$$

Since from (a) α and β are both irrational, we get from (26) and (27) that $y_3z_3 \neq 0$. This shows that α is an algebraic number of degree at most two, which is a contradiction with our assumption that α is algebraic of degree at least 3. This concludes the proof in this case.

We now prove the above claims. Let $(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4$ and $\mathcal{N}_1 \subseteq \mathbb{N}$ be as in Lemma 6.1. Recalling that $B_{-1} = A_0 = 0$ and $B_0 = A_{-1} = 1$, we deduce from (15) that

$$x_1 B_{\ell_n-1} + x_2 A_{\ell_n-1} - x_3 B_{\ell_n} - x_4 A_{\ell_n} = 0 \quad (28)$$

for any $n \in \mathcal{N}_1$. Dividing (28) by B_{ℓ_n} and letting n tend to infinity in $\mathcal{N}_1$ we get

$$\beta = \frac{x_3 + x_4\alpha}{x_1 + x_2\alpha}. \quad (29)$$

Observe that $x_1 + x_2\alpha \neq 0$, since otherwise α would be rational.

We now want to prove (a) and (b) (notice that (a) is not automatically guaranteed by (29) since a priori one could have $x_3x_2 - x_4x_1 = 0$).

This is again obtained by an application of the Subspace Theorem 5.2 with $N = 3$ and the linear forms, for $1 \leq i \leq 3$

$$L'_{i,\infty}(X_1, X_2, X_3) = X_i$$

and

$$\begin{aligned} L'_{1,p}(X_1, X_2, X_3) &= \beta X_1 - X_2, \quad L'_{2,p}(X_1, X_2, X_3) \\ &= \alpha X_1 - X_3, \quad L'_{3,p}(X_1, X_2, X_3) = X_2, \end{aligned}$$

evaluated at the points $(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n})$.

Setting $Q_n = \frac{A_n}{B_n}$, we have, by (28),

$$\frac{B_{\ell_n-1}}{B_{\ell_n}} = \frac{x_3 + x_4 Q_{\ell_n}}{x_1 + x_2 Q_{\ell_n-1}},$$

and therefore, by (29),

$$\left| \beta - \frac{B_{\ell_n-1}}{B_{\ell_n}} \right|_p = \left| \frac{x_1 x_4 (\alpha - Q_{\ell_n}) - x_2 x_3 (\alpha - Q_{\ell_n-1}) + x_2 x_4 \alpha (Q_{\ell_n-1} - Q_{\ell_n})}{(x_1 + x_2 \alpha)(x_1 + x_2 Q_{\ell_n-1})} \right|_p.$$

We notice that, since Q_{ℓ_n-1} converges p -adically to α and $x_1 + x_2 \alpha \neq 0$ since α is irrational, we have that, for n sufficiently large, $|x_1 + x_2 Q_{\ell_n}|_p = |x_1 + x_2 \alpha|_p$. Hence, for $n \in \mathcal{N}_1$, we have

$$|B_{\ell_n} \beta - B_{\ell_n-1}|_p \leq \frac{C'}{|B_{\ell_n-1}|_p} \quad (30)$$

for some real $C' > 0$.

Furthermore, observe that, for any integer $n \in \mathcal{N}_1$, we have, by Lemma 3.3 and (30),

$$\max(|B_{\ell_n} \alpha - A_{\ell_n}|_p, |B_{\ell_n-1} \alpha - A_{\ell_n-1}|_p, |B_{\ell_n} \beta - B_{\ell_n-1}|_p) \ll \frac{1}{|B_{\ell_n-1}|_p}. \quad (31)$$

We obtain, by (31) and hypotheses (i) and (iii) of Theorem 4.1,

$$\prod_{1 \leq j \leq 3} |L'_{j,p}(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n})|_p |L'_{j,\infty}(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n})|_{\infty} \ll \frac{C_{\infty}^{3\ell_n}}{|B_{\ell_n-1}|_p} \ll \left(\frac{C_{\infty}^3}{p^k} \right)^{\ell_n} \quad (32)$$

and

$$H(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n}) \ll (C_{\infty} C_p)^{\ell_n}. \quad (33)$$

Notice that, up to passing to a subset of $\mathcal{N}_1$, we might assume that ℓ_n is increasing. By the definition of k in Theorem 4.1 we have $C_{\infty}^3/p^k < 1$ in (32). Therefore, by Theorem 5.2, there exist a non zero integer triple (y_1, y_2, y_3) and an infinite set of distinct positive integers $\mathcal{N}_2 \subseteq \mathcal{N}_1$ such that

$$y_1 B_{\ell_n} + y_2 B_{\ell_n-1} + y_3 A_{\ell_n} = 0 \quad (34)$$

for any $n \in \mathcal{N}_2$. Dividing (34) by B_{ℓ_n} and letting n tend to infinity along $\mathcal{N}_2$, we get that (26) holds, so that claim (b) is proven.

Equations (34) and (26) then prove claim (a): indeed, if β was rational, (26) would imply $y_3 = 0$ (recall that we are supposing α irrational). But then, by (34) we would have

$$B_{\ell_n-1}/B_{\ell_n} = -y_1/y_2,$$

for all $n \in \mathcal{N}_2$ and, for each such index, by Lemma 3.6 we would have

$$-\frac{y_2}{y_1} = \frac{B_{\ell_n-1}}{B_{\ell_n}} = [0, a_{\ell_n}, \dots, a_1]_s,$$

contradicting the uniqueness of p -adic continued fraction expansion of $-y_2/y_1$ attached to the floor function s .¹

To prove (c), we also apply the Subspace Theorem 5.2 with $N = 3$ and the linear forms

$$L''_{i,\infty}(X_1, X_2, X_3) = X_i$$

for $1 \leq i \leq 3$ and

$$\begin{aligned} L''_{1,p}(X_1, X_2, X_3) &= \beta X_1 - X_2, & L''_{2,p}(X_1, X_2, X_3) &= \alpha X_2 - X_3, \\ L''_{3,p}(X_1, X_2, X_3) &= X_2. \end{aligned}$$

Evaluating them on the triple $(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n-1})$ with $n \in \mathcal{N}_2$, we infer from (31) that

$$\prod_{1 \leq j \leq 3} |L''_{j,p}(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n-1})|_p |L''_{j,\infty}(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n-1})|_\infty \ll \left(\frac{C_\infty^3}{p^k} \right)^{\ell_n}.$$

As before, estimating $H(B_{\ell_n}, B_{\ell_n-1}, A_{\ell_n-1})$ as in (33), and using the definition of k in Theorem 4.1, we can apply Theorem 5.2 and deduce that there exist a non zero integer triple (z_1, z_2, z_3) and an infinite subset $\mathcal{N}_3 \subseteq \mathcal{N}_2$ such that

$$z_1 B_{\ell_n} + z_2 B_{\ell_n-1} + z_3 A_{\ell_n-1} = 0 \quad (35)$$

for any $n \in \mathcal{N}_3$. Dividing (35) by B_{ℓ_n-1} and letting n tend to infinity along $\mathcal{N}_3$, we get (27), concluding the proof of this case.

Case (2): $(w_n)_{n \in \mathcal{N}_1}$ has no constant subsequences

As detailed in [19, page 1013], up to suitably modifying the words U_n and W_n , we can assume that the last letter of $U_n V_n$ is different from that of W_n , that is $a_{w_n} \neq a_{\ell_n}$ for any $n \in \mathcal{N}_1$.

We now set

$$Q_n = \frac{B_{w_n-1} B_{\ell_n}}{B_{w_n} B_{\ell_n-1}} \quad (36)$$

and

$$R_m = \alpha - \frac{A_m}{B_m}. \quad (37)$$

Dividing (15) by $B_{w_n} B_{\ell_n-1}$, we obtain

$$\begin{aligned} x_1(Q_n - 1) + x_2(Q_n(\alpha - R_{\ell_n}) - (\alpha - R_{\ell_n-1})) + x_3(Q_n(\alpha - R_{w_n-1}) - (\alpha - R_{w_n})) + \\ + x_4(Q_n(\alpha - R_{w_n-1})(\alpha - R_{\ell_n}) - (\alpha - R_{w_n})(\alpha - R_{\ell_n-1})) = 0. \end{aligned}$$

Letting also

$$P(\alpha) = x_1 + (x_2 + x_3)\alpha + x_4\alpha^2,$$

this yields

$$\begin{aligned} (Q_n - 1)P(\alpha) = & x_2 Q_n R_{\ell_n} - x_2 R_{\ell_n-1} + x_3 Q_n R_{w_n-1} - x_3 R_{w_n} - x_4 Q_n R_{w_n-1} R_{\ell_n} + \\ & + x_4 R_{w_n} R_{\ell_n-1} + \alpha x_4 (Q_n R_{w_n-1} + Q_n R_{\ell_n} - R_{w_n} - R_{\ell_n-1}). \end{aligned} \quad (38)$$

We want to prove

$$P(\alpha) = 0. \quad (39)$$

¹ Notably, this is the first time in the proof where we use the fact that the word associated to α arises from a given floor function.

Suppose first that there exists $\epsilon > 0$ and an infinite subset $\mathcal{M} \subset \mathcal{N}_1$ such that $|Q_n - 1|_p \geq \epsilon$ for all $n \in \mathcal{M}$.

Suppose further that $|Q_n|_p \leq 1$ for infinitely many $n \in \mathcal{M}$. Then, for n tending to infinity along the elements of this subset, the right-hand side of (38) tends to 0 (indeed, by (37), R_m tends to 0 for m tending to infinity and Q_m is bounded), which implies (39). If instead $|Q_n|_p > 1$ for all but finitely many $n \in \mathcal{M}$, dividing both sides of formula (38) by Q_n and letting n tend to infinity, we obtain, as in the previous case, that (39) holds.

Suppose now that

$$\lim_{n \in \mathcal{N}_1} |Q_n - 1|_p = 0.$$

Then, for n sufficiently large, we have $|Q_n|_p = |1|_p = 1$.

Notice that, by Lemma 3.6, Q_n is the quotient of the two continued fractions $\mathbf{a}_{\ell_n} = [a_{\ell_n}, a_{\ell_n-1}, \dots, a_1]_s$ and $\mathbf{a}_{w_n} = [a_{w_n}, a_{w_n-1}, \dots, a_1]_s$, hence must have

$$|a_{\ell_n}|_p = |\mathbf{a}_{\ell_n}|_p = |\mathbf{a}_{w_n}|_p = |a_{w_n}|_p = p^{k_n},$$

for n sufficiently large, with $k_n > 0$.

Let us now assume, by contradiction, that $P(\alpha) \neq 0$. Then, since $|R_\ell|_p \ll 1/|B_\ell B_{\ell+1}|_p$ for $\ell \geq 1$, from (38) we deduce

$$|Q_n - 1|_p \ll \frac{1}{|B_{w_n-1} B_{w_n}|_p},$$

which we can rephrase as

$$\left| \frac{\mathbf{a}_{\ell_n}}{\mathbf{a}_{w_n}} - 1 \right|_p \ll \frac{1}{|a_{w_n}|_p \cdot \prod_{i=1}^{w_n-1} |a_i|_p^2}.$$

Now, multiplying both the numerator and denominator of the left-hand side by p^{k_n} and noticing that $|p^{k_n} \mathbf{a}_{w_n}|_p = 1$ and $|a_i|_p \geq p$ for every $i \leq 1$, we obtain, in particular, that

$$\left| p^{k_n} \mathbf{a}_{\ell_n} - p^{k_n} \mathbf{a}_{w_n} \right|_p \ll \frac{1}{p^{2(w_n-1)+k_n}}.$$

Therefore

$$|[a_{\ell_n}, a_{\ell_n-1}, \dots, a_1]_s - [a_{w_n}, a_{w_n-1}, \dots, a_1]_s|_p \ll \frac{1}{p^{2(w_n-1)}} \leq \frac{1}{p}.$$

This implies that $\mathbf{a}_{\ell_n} \equiv \mathbf{a}_{w_n} \pmod{p}$, so that $a_{\ell_n} \equiv a_{w_n} \pmod{p}$. Therefore, by property (iii) in the definition of floor function, they are equal, contradicting our initial assumption that $a_{w_n} \neq a_{\ell_n}$.

Therefore (39) is true; since we are assuming that α is neither rational nor quadratic, we must have $x_1 = x_2 + x_3 = x_4 = 0$. Then, equation (15) gives

$$B_{w_{n-1}} A_{\ell_n} - B_{w_n} A_{\ell_n-1} = A_{w_{n-1}} B_{\ell_n} - A_{w_n} B_{\ell_n-1}.$$

In particular, in (16) we have $z_{n,2} = z_{n,3}$. We can thus refine the argument in Lemma 6.1. Let $\alpha_n = [W_n, \overline{U_n V_n}]_s$. Then, by Lemma 3.5, α_n is a root of the polynomial

$$P_n(X) = z_{n,1} X^2 - 2z_{n,2} X + z_{n,4}.$$

Consider now the two sets of linearly independent linear forms

$$L'''_{1,p}(T_1, T_2, T_3) = \alpha^2 T_1 - 2\alpha T_2 + T_3,$$

$$\begin{aligned}
L_{2,p}'''(T_1, T_2, T_3) &= \alpha T_1 - T_2, \\
L_{3,p}'''(T_1, T_2, T_3) &= T_1, \\
L_{1,\infty}'''(T_1, T_2, T_3) &= T_1, \\
L_{2,\infty}'''(T_1, T_2, T_3) &= T_2, \\
L_{3,\infty}'''(T_1, T_2, T_3) &= T_3;
\end{aligned}$$

evaluating them on the triple $\mathbf{z}'_n = (z_{n,1}, z_{n,2}, z_{n,4})$ we find by (19), (21), (22), and (23) that

$$\begin{aligned}
\prod_{1 \leq j \leq 3} |L_{j,p}'''(\mathbf{z}'_n)|_p \prod_{1 \leq j \leq 3} |L_{j,\infty}'''(\mathbf{z}'_n)|_\infty &\ll |B_{w_n}|_p |B_{\ell_n}|_p |B_{\ell_n+u_n}|_p^{-2} C_\infty^{3(\ell_n+w_n)} \\
&\leq \frac{C_\infty^{3(3c+1)u_n}}{\prod_{j=\ell_n+1}^{\ell_n+u_n} |a_j|_p^2} \\
&\leq \left(\frac{C_\infty^{3(3c+1)}}{p^{2k}} \right)^{u_n}.
\end{aligned}$$

Also, by (25) we have

$$H(\mathbf{z}_n) \ll (C_\infty C_p)^{(3c+1)u_n}.$$

Again using the definition of k in Theorem 4.1 and applying Theorem 5.2 we obtain that there exists a non zero integer triple (t_1, t_2, t_3) and an infinite set $\mathcal{N}_4 \subseteq \mathcal{N}_1$ such that

$$t_1 z_{n,1} + t_2 z_{n,2} + t_3 z_{n,4} = 0 \quad (40)$$

for any $n \in \mathcal{N}_4$. We proceed exactly as above. Divide (40) by $B_{w_n} B_{\ell_n-1}$ and set Q_n as in (36). Then, we get

$$t_1(Q_n - 1) + t_2 \left(Q_n \frac{A_{\ell_n}}{B_{\ell_n}} - \frac{A_{\ell_n-1}}{B_{\ell_n-1}} \right) + t_3 \left(Q_n \frac{A_{w_n-1}}{B_{w_n-1}} \frac{A_{\ell_n}}{B_{\ell_n}} - \frac{A_{w_n}}{B_{w_n}} \frac{A_{\ell_n-1}}{B_{\ell_n-1}} \right) = 0,$$

for any $n \in \mathcal{N}_4$. By arguing as after (39) we derive that $t_1 + t_2\alpha + t_3\alpha^2 = 0$, a contradiction since α is irrational and not quadratic. This concludes the proof of Theorem 4.1. $\square$

7 Proof of Theorem 4.3

Suppose by contradiction that α were algebraic of degree at least 3 over $\mathbb{Q}$.

By hypothesis the sequence $(a_n)_n$ satisfies condition $\clubsuit$ with constant c . Hence, there exist three sequences of finite words $(W_n)_n$, $(U_n)_n$, $(V_n)_n$ satisfying conditions (i), (ii) and (iii) of Definition 1.2. Setting $w_n = |W_n|$, $u_n = |U_n|$, and $v_n = |V_n|$, for every $n \geq 0$,

$$\max(v_n/u_n, w_n/u_n) \leq c.$$

We let also

$$\ell_n = w_n + u_n + v_n.$$

Here the proof splits into three cases.

The first case occurs when the set $\{n \geq 1 \mid v_n = w_n = 0\}$ is infinite: here, using the fact that the words $a_1 \dots a_{2u_n}$ are palindromic for infinitely many n , an application of the Subspace Theorem 5.2 allows us to conclude.

In the second case, we consider the situation where the set $\{n \geq 1 \mid v_n = w_n = 0\}$ is finite, but the set $\{n \geq 1 \mid w_n = 0\}$ is still infinite: in this case, on one hand we show that the continued fraction expansion of α has ‘many’ partial quotients in common with that of certain non palindromic words constructed using certain continuants of α . On the other hand, an application of the Subspace theorem and the uniqueness of the continued fraction expansion for a given floor function, imply that the words we considered are actually palindromic, leading to a contradiction.

Finally, the third case occurs when $\{n \geq 1 \mid w_n = 0\}$ is finite. Here, we use the words U_n , V_n and W_n together with the recurrence formulas defining the continuants to build two sequences of numbers whose continued fraction expansions have ‘many’ partial quotients in common with that of α . A first application of the Subspace theorem gives a relation between some of these words and a second application of the Subspace theorem allows us to conclude.

Case (1): the set $\{n \geq 1 \mid w_n = v_n = 0\}$ is infinite

In this case (which occurs, in particular, when $c = 0$) we have that, up to passing to an infinite subset of positive integers n , we may assume that $\ell_n = u_n$ and $U_n \widetilde{U}_n$ is a prefix of α for every n . Since the word $a_1 \dots a_{2u_n} = U_n \widetilde{U}_n$ is palindromic, we have that the matrix

$$\begin{pmatrix} B_{2u_n} & B_{2u_n-1} \\ A_{2u_n} & A_{2u_n-1} \end{pmatrix} = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{2u_n} & 1 \\ 1 & 0 \end{pmatrix}$$

is symmetric, hence, for every n

$$A_{2u_n} = B_{2u_n-1}. \quad (41)$$

Notice that by (9) we have

$$\left| \alpha - \frac{A_{2u_n}}{B_{2u_n}} \right|_p \leq \frac{1}{|B_{2u_n}|_p^2} \quad (42)$$

and also, by (41),

$$\left| \alpha - \frac{A_{2u_n-1}}{A_{2u_n}} \right|_p = \left| \alpha - \frac{A_{2u_n-1}}{B_{2u_n-1}} \right|_p \leq \frac{1}{|B_{2u_n-1}|_p^2}. \quad (43)$$

Moreover

$$\begin{aligned} \left| \alpha^2 - \frac{A_{2u_n-1}}{B_{2u_n}} \right|_p &= \left| \alpha^2 - \frac{A_{2u_n-1}}{B_{2u_n-1}} \frac{A_{2u_n}}{B_{2u_n}} \right|_p \\ &= \left| \left(\alpha + \frac{A_{2u_n-1}}{B_{2u_n-1}} \right) \left(\alpha - \frac{A_{2u_n}}{B_{2u_n}} \right) + \alpha \left(\frac{A_{2u_n}}{B_{2u_n}} - \frac{A_{2u_n-1}}{B_{2u_n-1}} \right) \right|_p \\ &\leq \max \left(|\alpha|_p \left| \alpha - \frac{A_{2u_n}}{B_{2u_n}} \right|_p, \frac{|\alpha|_p}{|B_{2u_n} B_{2u_n-1}|_p} \right) \end{aligned}$$

and since $|\alpha|_p < 1$, we finally obtain

$$\left| \alpha^2 - \frac{A_{2u_n-1}}{B_{2u_n}} \right|_p < \max \left(\frac{1}{|B_{2u_n}|_p^2}, \frac{1}{|B_{2u_n} B_{2u_n-1}|_p} \right) \leq \frac{1}{|B_{2u_n} B_{2u_n-1}|_p}. \quad (44)$$

We apply Theorem 5.2 with $N = 3$ and the linear forms

$$L'_{1,p}(X_1, X_2, X_3) = X_1 \alpha - X_2,$$

$$L'_{2,p}(X_1, X_2, X_3) = X_1\alpha^2 - X_3,$$

$$L'_{3,p}(X_1, X_2, X_3) = X_2\alpha - X_3,$$

and $L'_{i,\infty}(X_1, X_2, X_3) = X_i$ for every $1 \leq i \leq 3$. Evaluating them at the points $\mathbf{z}'_n = (B_{2u_n}, A_{2u_n}, A_{2u_n-1})$, by (42), (43) and (44) we get

$$\prod_{i=1}^3 |L'_{i,p}(\mathbf{z}'_n)|_p |L'_{i,\infty}(\mathbf{z}'_n)|_\infty \leq \frac{C_\infty^{6u_n}}{|B_{2u_n-1}|_p} \ll \left(\frac{C_\infty}{p^k}\right)^{6u_n}.$$

Moreover

$$H(\mathbf{z}'_n) \leq (C_\infty C_p)^{2u_n}.$$

As by the definition of k in Theorem 4.3 we have $k > \frac{\log C_\infty}{\log p}$, Theorem 5.2 ensures there exist $x_1, x_2, x_3 \in \mathbb{Z}$ not all zero and an infinite subsequence of positive integers n such that

$$x_1 B_{2u_n} + x_2 A_{2u_n} + x_3 A_{2u_n-1} = 0 \quad (45)$$

for all n . Dividing (45) by B_{2u_n} we have

$$x_1 + x_2 \frac{A_{2u_n}}{B_{2u_n}} + x_3 \frac{A_{2u_n-1}}{B_{2u_n-1}} \frac{B_{2u_n-1}}{B_{2u_n}} = 0.$$

Since by (41) $A_{2u_n} = B_{2u_n-1}$, letting n go to infinity, we get

$$x_1 + x_2\alpha + x_3\alpha^2 = 0$$

contradicting the hypothesis that α is algebraic of degree at least 3.

Case (2): the set $\{n \geq 1 \mid w_n = v_n = 0\}$ is finite and the set $\{n \geq 1 \mid w_n = 0\}$ is infinite

Without loss of generality we might assume that, up to passing to a subsequence, for all n , $v_n \neq 0$, $w_n = 0$ and that, by enlarging the words U_n , the first and the last letters of the word V_n are always different (if not, it means that we are in Case (1)). So, in particular, the word $U_n V_n \widetilde{U}_n$ is never palindromic.

Notice first that, by Lemma 3.6,

$$\frac{B_{n-1}}{B_n} = [0, a_n, a_{n-1}, \dots, a_1]_s.$$

In particular

$$\frac{B_{\ell_n+u_n-1}}{B_{\ell_n+u_n}} = [0, U_n \widetilde{V}_n \widetilde{U}_n]_s.$$

On the other hand

$$\frac{A_{\ell_n+u_n}}{B_{\ell_n+u_n}} = [0, U_n V_n \widetilde{U}_n]_s.$$

Thus the continued fraction expansions of the numbers α and $B_{\ell_n+u_n-1}/B_{\ell_n+u_n}$ share the first u_n terms, hence by Lemma 3.4 we have

$$|B_{\ell_n+u_n}\alpha - B_{\ell_n+u_n-1}|_p \leq |B_{\ell_n+u_n}|_p |B_{u_n}|_p^{-2}. \quad (46)$$

We now apply Theorem 5.2 with $N = 4$ and the linear forms

$$L_{1,p}(X_1, X_2, X_3, X_4) = \alpha X_1 - X_3,$$

$$\begin{aligned}L_{2,p}(X_1, X_2, X_3, X_4) &= \alpha X_2 - X_4, \\L_{3,p}(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_2, \\L_{4,p}(X_1, X_2, X_3, X_4) &= X_2,\end{aligned}$$

and

$$L_{i,\infty}(X_1, X_2, X_3, X_4) = X_i,$$

for $1 \leq i \leq 4$. Evaluating them at the points

$$\mathbf{z}_n = (B_{\ell_n+u_n}, B_{\ell_n+u_n-1}, A_{\ell_n+u_n}, A_{\ell_n+u_n-1}),$$

by hypothesis (i) of Theorem 4.3 and inequalities (9) and (46), we find that

$$\prod_{1 \leq i \leq 4} |L_{i,p}(\mathbf{z}_n)|_p |L_{i,\infty}(\mathbf{z}_n)|_\infty \leq \frac{C_\infty^{4(\ell_n+u_n)}}{|B_{u_n}|_p^2} \leq \left(\frac{C_\infty^{4(2+c)}}{p^{2k}} \right)^{u_n}.$$

Notice now that

$$H(\mathbf{z}_n) \leq (C_\infty C_p)^{(2+c)u_n}.$$

But then, since by the definition of k in Theorem 4.3

$$k > (4 + 2c) \frac{\log C_\infty}{\log p},$$

Theorem 5.2 ensures that there exist $x_1, x_2, x_3, x_4 \in \mathbb{Z}$ not all zero and an infinite subsequence of positive integers n such that, for all n , we have

$$x_1 B_{\ell_n+u_n} + x_2 B_{\ell_n+u_n-1} + x_3 A_{\ell_n+u_n} + x_4 A_{\ell_n+u_n-1} = 0. \quad (47)$$

Dividing by $B_{\ell_n+u_n}$ we obtain

$$x_1 + x_2 \frac{B_{\ell_n+u_n-1}}{B_{\ell_n+u_n}} + x_3 \frac{A_{\ell_n+u_n}}{B_{\ell_n+u_n}} + x_4 \frac{A_{\ell_n+u_n-1}}{B_{\ell_n+u_n-1}} \cdot \frac{B_{\ell_n+u_n-1}}{B_{\ell_n+u_n}} = 0. \quad (48)$$

Recalling that by (46) we have

$$\lim_{n \rightarrow \infty} \frac{B_{\ell_n+u_n-1}}{B_{\ell_n+u_n}} = \alpha,$$

letting n tend to infinity in equation (48), we get

$$x_1 + (x_2 + x_3)\alpha + x_4\alpha^2 = 0.$$

Since α is not quadratic, we obtain $x_1 = x_4 = 0$ and $x_2 = -x_3$, hence, from (47) we have

$$A_{\ell_n+u_n} = B_{\ell_n+u_n-1}.$$

But then the matrix

$$\begin{pmatrix} B_{\ell_n+u_n} & B_{\ell_n+u_n-1} \\ A_{\ell_n+u_n} & A_{\ell_n+u_n-1} \end{pmatrix} = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{\ell_n+u_n} & 1 \\ 1 & 0 \end{pmatrix}$$

is symmetric. Hence, by the uniqueness of the continued fraction expansion for a given floor function, the word $a_1 \dots a_{\ell_n+u_n} = U_n V_n \widetilde{U}_n$ is palindromic, contradicting our assumption.

Case (3): the set $\{n \geq 1 \mid w_n = 0\}$ is finite

Let

$$A(t_0, \dots, t_m), \quad B(t_0, \dots, t_m)$$

be the functions defined by the continuants recurrences (2), that is

$$A(t_0) = t_0, \quad A(t_0, t_1) = t_0 t_1 + 1, \quad A(t_0, \dots, t_i) = t_i A(t_0, \dots, t_{i-1}) + A(t_0, \dots, t_{i-2})$$

$$B(t_0) = 1, \quad B(t_0, t_1) = t_1, \quad B(t_0, \dots, t_i) = t_i B(t_0, \dots, t_{i-1}) + B(t_0, \dots, t_{i-2}).$$

For any positive n , let

$$P_n = A(0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}_n), \quad Q_n = B(0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}_n)$$

and

$$P'_n = A(0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}'_n) \quad Q'_n = B(0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}'_n)$$

where W'_n denotes the word obtained by removing the first letter of W_n . We consider these as functions of $\ell_n + u_n + w_n$, respectively $\ell_n + u_n + w_n - 1$, parameters.

Observe that

$$\frac{P_n}{Q_n} = [0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}_n]_s$$

and

$$\frac{P'_n}{Q'_n} = [0, W_n, U_n, V_n, \tilde{U}_n, \tilde{W}'_n]_s.$$

Notice that the continued fraction expansions of both P_n/Q_n and P'_n/Q'_n have the first $\ell_n + u_n$ partial quotients in common with α . Therefore, by Sect. 3.2, we have

$$\begin{pmatrix} Q_n & Q'_n \\ P_n & P'_n \end{pmatrix} = \begin{pmatrix} B_{\ell_n+u_n} & B_{\ell_n+u_n-1} \\ A_{\ell_n+u_n} & A_{\ell_n+u_n-1} \end{pmatrix} \begin{pmatrix} B_{w_n} & A_{w_n} \\ B_{w_n-1} & A_{w_n-1} \end{pmatrix},$$

hence

$$\begin{aligned} Q_n &= B_{\ell_n+u_n} B_{w_n} + B_{\ell_n+u_n-1} B_{w_n-1}, \\ Q'_n &= B_{\ell_n+u_n} A_{w_n} + B_{\ell_n+u_n-1} A_{w_n-1}, \\ P_n &= A_{\ell_n+u_n} B_{w_n} + A_{\ell_n+u_n-1} B_{w_n-1}, \\ P'_n &= A_{\ell_n+u_n} A_{w_n} + A_{\ell_n+u_n-1} A_{w_n-1}. \end{aligned} \tag{49}$$

Again, since α , P_n/Q_n and P'_n/Q'_n share the same first $\ell_n + u_n$ partial quotients, from Lemma 3.4 we have

$$|Q_n \alpha - P_n|_p \leq \left| \frac{Q_n}{B_{\ell_n+u_n}^2} \right|_p \leq \left| \frac{B_{\ell_n+u_n} B_{w_n}}{B_{\ell_n+u_n}^2} \right|_p = \left| \frac{B_{w_n}}{B_{\ell_n+u_n}} \right|_p, \tag{50}$$

$$|Q'_n \alpha - P'_n|_p \leq \left| \frac{B_{\ell_n+u_n} A_{w_n}}{B_{\ell_n+u_n}^2} \right|_p = \left| \frac{A_{w_n}}{B_{\ell_n+u_n}} \right|_p \leq \left| \frac{B_{w_n}}{B_{\ell_n+u_n}} \right|_p. \tag{51}$$

Notice also that, by Lemma 3.6,

$$\frac{Q'_n}{Q_n} = [0, W_n, U_n, \tilde{V}_n, \tilde{U}_n, \tilde{W}_n]_s.$$

It follows from Lemma 3.4 that

$$|Q_n \alpha - Q'_n|_p < \left| \frac{Q_n}{B_{w_n+u_n}^2} \right|_p \leq \left| \frac{B_{\ell_n+u_n} B_{w_n}}{B_{w_n+u_n}^2} \right|_p \quad (52)$$

showing, in particular, that

$$\lim_{n \rightarrow \infty} \frac{Q'_n}{Q_n} = \alpha.$$

Furthermore, by (49),

$$|Q'_n|_p \leq |B_{\ell_n+u_n} A_{w_n}|_p. \quad (53)$$

We now apply Theorem 5.2 with $N = 4$ and the linear forms

$$\begin{aligned} L''_{1,p}(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_3, \\ L''_{2,p}(X_1, X_2, X_3, X_4) &= \alpha X_2 - X_4, \\ L''_{3,p}(X_1, X_2, X_3, X_4) &= \alpha X_1 - X_2, \\ L''_{4,p}(X_1, X_2, X_3, X_4) &= X_2, \end{aligned}$$

and

$$L''_{i,\infty}(X_1, X_2, X_3, X_4) = X_i$$

for $1 \leq i \leq 4$. Evaluating such forms at the points

$$\mathbf{z}''_n = (Q_n, Q'_n, P_n, P'_n)$$

it follows from (50), (51), (52) and (53) that

$$\prod_{1 \leq j \leq 4} |L''_{j,p}(\mathbf{z}''_n)|_p \leq \left| \frac{B_{w_n}^4}{B_{w_n+u_n}^2} \right|_p.$$

Moreover, by (49), we have

$$\max(|Q_n|_\infty, |Q'_n|_\infty, |P_n|_\infty, |P'_n|_\infty) \leq 2C_\infty^{2\ell_n-v_n} \ll C_\infty^{(3c+2)u_n} \quad (54)$$

and

$$\max(|Q_n|_p, |Q'_n|_p, |P_n|_p, |P'_n|_p) \ll C_p^{(3c+2)u_n}. \quad (55)$$

Since

$$\begin{aligned} \alpha^2 Q_n - \alpha Q'_n - \alpha P_n + P'_n &= \alpha(Q_n \alpha - P_n) - (Q'_n \alpha - P'_n) \\ &= \alpha Q_n \left(\alpha - \frac{P_n}{Q_n} \right) - Q'_n \left(\alpha - \frac{P'_n}{Q'_n} \right) \\ &= (\alpha Q_n - Q'_n) \left(\alpha - \frac{P_n}{Q_n} \right) + Q'_n \left(\frac{P'_n}{Q'_n} - \frac{P_n}{Q_n} \right) \end{aligned}$$

it follows from (49), (50), (51) and (52) that

$$\begin{aligned} |\alpha^2 Q_n - \alpha Q'_n - \alpha P_n + P'_n|_p &\leq \max \left(\left| \frac{B_{w_n}}{B_{w_n+u_n}^2 B_{\ell_n+u_n}} \right|_p, \left| \frac{1}{B_{\ell_n+u_n} B_{w_n}} \right|_p \right) \\ &= \left| \frac{1}{B_{\ell_n+u_n} B_{w_n}} \right|_p = \frac{1}{|Q_n|_p}. \end{aligned} \quad (56)$$

We now also consider the linear form

$$L''_{5,p}(X_1, X_2, X_3, X_4) = \alpha^2 X_1 - \alpha X_2 - \alpha X_3 + X_4.$$

From (51), (52), (53), and (56), we deduce that

$$\prod_{j=2}^5 |L''_{j,p}(\mathbf{z}''_n)|_p \leq \left| \frac{B_{w_n}^2}{B_{w_n+u_n}^2} \right|_p \leq \frac{1}{\prod_{j=w_n+1}^{w_n+u_n} |a_j|^2} \leq \frac{1}{p^{2ku_n}}. \quad (57)$$

From (49), (54), (55) and (57) we get

$$\prod_{j=1}^4 |L''_{j,\infty}(\mathbf{z}''_n)|_\infty \prod_{j=2}^5 |L''_{j,p}(\mathbf{z}''_n)|_p \leq \frac{2^4 C_\infty^{4(2\ell_n - v_n)}}{p^{2ku_n}} \leq 2^4 \left(\frac{C_\infty^{4(3c+2)}}{p^{2k}} \right)^{u_n}$$

and

$$H(\mathbf{z}''_n) \ll (C_p C_\infty)^{(3c+2)u_n}.$$

Therefore, since by the definition of k in Theorem 4.3

$$k > (4 + 6c) \cdot \frac{\log C_\infty}{\log p},$$

we conclude by Theorem 5.2 that there exist $t_1, t_2, t_3, t_4 \in \mathbb{Z}$ not all zero and an infinite subsequence of positive integers n such that

$$t_1 Q_n + t_2 Q'_n + t_3 P_n + t_4 P'_n = 0 \quad (58)$$

for all n . Dividing by Q_n and noting that

$$\lim_{n \rightarrow \infty} \frac{Q'_n}{Q_n} = \lim_{n \rightarrow \infty} \frac{P_n}{Q_n} = \lim_{n \rightarrow \infty} \frac{P'_n}{Q'_n} = \alpha, \quad (59)$$

by letting n go to infinity we obtain

$$t_1 + (t_2 + t_3)\alpha + t_4\alpha^2 = 0.$$

Since α is not quadratic by hypothesis, we get $t_1 = t_4 = 0$ and $t_2 = -t_3$, which, replaced in (58), imply $Q'_n = P_n$ for infinitely many n .

By (56), we then have, for infinitely many such n ,

$$|\alpha^2 Q_n - \alpha Q'_n - \alpha P_n + P'_n|_p = |\alpha^2 Q_n - 2\alpha Q'_n + P'_n|_p \ll \frac{1}{|Q_n|_p}.$$

We finally apply again Theorem 5.2 with $N = 3$ and the linear forms

$$\begin{aligned} L'''_{1,p}(X_1, X_2, X_3) &= \alpha^2 X_1 - 2\alpha X_2 + X_3 \\ L'''_{2,p}(X_1, X_2, X_3) &= \alpha X_2 - X_3 \\ L'''_{3,p}(X_1, X_2, X_3) &= X_1 \\ L'''_{i,\infty}(X_1, X_2, X_3) &= X_i \end{aligned}$$

for all $1 \leq i \leq 3$. By evaluating them at the points

$$\mathbf{z}'''_n = (Q_n, P_n, P'_n)$$

we find

$$\prod_{j=1}^3 |L'''_{j,\infty}(\mathbf{z}''')|_{\infty} \cdot |L'''_{j,p}(\mathbf{z}''')|_p \leq \frac{2^3 C_{\infty}^{3(2\ell_n - v_n)}}{p^{2ku_n}} \leq 2^3 \left(\frac{C_{\infty}^{3(c+2)}}{p^{2k}} \right)^{u_n}$$

and

$$H(\mathbf{z}''') \ll (C_p C_{\infty})^{(3c+2)u_n}.$$

Therefore, since the definition of k in Theorem 4.3 implies

$$k > \frac{3}{2}(3c+2) \frac{\log C_{\infty}}{\log p}$$

and (54) holds, we conclude by Theorem 5.2 that there exist $b_1, b_2, b_3, b_4 \in \mathbb{Z}$ not all zero and an infinite subsequence of positive integers n such that

$$b_1 Q_n + b_2 Q'_n + b_3 P'_n = 0$$

for all n . Diving by Q_n , using (59) and letting n go to infinity we obtain $b_1 + b_2 \alpha + b_3 \alpha^2 = 0$, contradicting the assumption that α is neither rational nor quadratic. $\square$

Appendix A Poking words: the ♠ and ♣ conditions

In this appendix we collect more results on ♠ and ♣ words and their combinatorial properties. In particular, we study how such words are related to other famous classes of words, such as automatic and Sturmian words. Moreover, we show how Corollary 4.6 can be specialized to these classes of words.

A.1 Link to automatic words

We briefly recall that, if $k \geq 2$ is an integer, given a set $\mathcal{A}$, a sequence $(a_n)_{n \geq 1}$ with $a_n \in \mathcal{A}$ is called k -automatic if, for every n , the term a_n can be computed by a finite state machine (or *automaton*). This machine takes in input the k -ary digits of n , starting from an initial state. Each digit of n corresponds to a state transition, and each state has an associated output. The output of the final state reached after reading all digits of n determines the value of a_n . For further details on automata theory, we refer to [11].

We can extend the notion of automaticity to words, by saying that a word $w = a_1 a_2 \dots$ is k -automatic if and only if the sequence $(a_n)_n$ is so.

As a finite automaton has only finitely many possible states and so finitely many possible outputs, any word taking on infinitely many different values is clearly not automatic. However, the finiteness of the values alone is not enough to ensure automaticity. An example of a word on a finite alphabet which is not k -automatic for any $k \geq 2$ is the *Fibonacci word*

$$\omega = 01001010010010100101001010010100 \dots$$

whose letters are given by the sequence $(a_n)_{n \geq 1}$ where $a_n = 2 + \lfloor n\varphi \rfloor - \lfloor (n+1)\varphi \rfloor$ where φ is the golden ratio. The Fibonacci word and, more generally, Sturmian words (see Sect. A.2), are not k -automatic [28, §1].

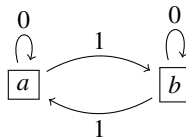
A result by Cobham [20] shows that, for any automatic word ω , there exists a constant $C > 0$ such that $p_{\omega}(n) \leq Cn$ for all $n \geq 1$. This, together with Theorem 2.1 gives:

Proposition A.1 *Every automatic word satisfies condition ♠.*

We now discuss some examples of automatic words.

Thue–Morse words

One of the most classical and simplest examples of a 2-automatic word is the Thue–Morse word mentioned earlier. Put $\mathcal{A} = \{a, b\}$. The *Thue–Morse sequence on the alphabet $\mathcal{A}$* is the sequence $(a_n)_{n \geq 0}$ where $a_n = a$ if the number of 1 in the binary writing of n is even and $a_n = b$ otherwise. It is produced by the following automaton which has 2 states, where the arrows indicate the transition functions depending on the lecture of a binary digit of n .



The corresponding *Thue–Morse word* is

$$w = abbabaabbaababbabaababba \dots$$

We have the following:

Proposition A.2 *Thue–Morse words satisfy:*

- (i) condition ♠ with constant $c = 2$;
- (ii) condition ♣ with constant $c = 0$.

Proof Let w be a Thue–Morse word. Notice that, for each n , any initial block of w of length 2^n has the form $U_n V_n U_n$ with $|V_n| = 2|U_n| = 2^{n-1}$. This proves (i). As for (ii), as remarked in [2], denoting by U_n the word consisting of the first 2^n letters of w , then $U_n \widetilde{U_n}$ is a prefix of w for every n . Hence w satisfies condition ♣ with $|W_n| = |V_n| = 0$. $\square$

Rudin–Shapiro words

Put $\mathcal{A} = \{a, b\}$. The *Rudin–Shapiro sequence on the alphabet $\mathcal{A}$* is the sequence $(a_n)_{n \geq 0}$ defined as: for every $n \geq 1$ in which case $a_n = a$ if the number of occurrences (possibly overlapping) of the block 11 in the base-2 expansion of n is even, or $a_n = b$ otherwise.

The corresponding *Rudin–Shapiro word* is

$$w = aaabaabaaaabbbabaaabaababbbbaaabaabaa \dots$$

and it is 2-automatic.

Proposition A.3 *Rudin–Shapiro words satisfy condition ♠ with constant $c = 25$.*

Proof It has been proved that $p_w(n) = 8(n-1)$ for every $n \geq 8$ (see [12] for a more general result). Thus the result follows from Theorem 2.1 with $C = 8$. $\square$

Paperfolding words

Another classical example of 2-automatic word is the *paperfolding word*. Put $\mathcal{A} = \{a, b\}$. The *paperfolding sequence on the alphabet $\mathcal{A}$* is the sequence $(a_n)_n$ defined as it follows: for every $n \geq 1$, write $n = 2^k m$ where m is odd and set $a_n = a$ if $m \equiv 1 \pmod{4}$ and $a_n = b$ otherwise. The corresponding *paperfolding word* is

$$w = aabaabbaaabbabbaaabaabbaabbabbaaabaabbaa \dots$$

Proposition A.4 *Paperfolding words satisfy condition ♠ with constant $c = 13$.*

Proof By a result of Allouche (see [8]) $p_\omega(n) = 4n$ for all $n \geq 7$. Hence, the result follows by Theorem 2.1 with $C = 4$. $\square$

Remark A.5 To the authors' knowledge, it is not known whether Rudin–Shapiro and paperfolding words satisfy condition ♣. However, if this is the case we must have that the sequence $(V_n)_n$ defining the ♣ condition can contain only finitely many empty words. Indeed, as proved in [9], the longest palindromes in a Rudin–Shapiro word have length 14 while they have length 13 in a paperfolding word.

A.2 Link to Sturmian words

A word ω is said to be *Sturmian* if $p_\omega(n) = n + 1$ for all $n \geq 1$. Equivalently (see for instance [7, Section 4.2]) a word $w = a_0a_1a_2 \dots$ is Sturmian if and only if there exists a set $\mathcal{A} = \{a, b\}$ such that the sequence $(a_n)_{n \geq 0}$ satisfies, for some irrational $\alpha \in (0, 1)$ and some real number β , one of the following properties

- (i) for all $n \geq 1$, $a_n = a$ if $\lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor = 0$ and $a_n = b$ otherwise;
- (ii) for all $n \geq 1$, $a_n = a$ if $\lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil = 0$ and $a_n = b$ otherwise.

A word is called *characteristic Sturmian* if it satisfies the above conditions with $\beta = 0$. A classical example of characteristic Sturmian word is the Fibonacci word introduced above.

Proposition A.6 *We have that:*

- (i) *Sturmian words satisfy condition ♠ with constant $c = 0$;*
- (ii) *Characteristic Sturmian words satisfy condition ♣ with constant $c = 0$.*
- (iii) *Sturmian words satisfy condition ♣ with constant $c = 2$.*

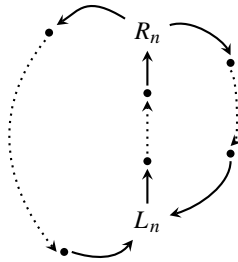
Proof Point (i) follows from the fact that any Sturmian word begins with arbitrarily long squares, as proved in [7, Proposition 2].

Point (ii) follows from the fact that characteristic Sturmian words starts with arbitrarily long palindromes (see for instance [21]).

Point (iii) requires some further work. Let α be a Sturmian word. Following [24, §1.3.4], for each $n \geq 1$, we define the *factor graph* of α as the directed graph $G_n = G_n(\alpha)$ whose vertices are the distinct blocks of length n appearing in α (there are $n + 1$ distinct blocks in the case of Sturmian words), and such that there is a directed edge from a block $w_1 \dots w_n$ to $w'_1 \dots w'_n$ if and only if $w_2 \dots w_n = w'_1 \dots w'_{n-1}$ and $w_1 \dots w_n w'_n$ is a block of α of length $n + 1$. Notably, α is completely described by some path on $G_n(\alpha)$, namely the one starting on the vertex representing $a_1 \dots a_n$, and such that the i -th step lands exactly the block corresponding to $a_{i+1} \dots a_{n+i+1}$ for each $i \in \mathbb{N}$.

By [24, p. 63], there exists a unique subword of length n , say $w_1 \dots w_n$, called the *n -th right special factor* of α and denoted by R_n , such that both $w_1 \dots w_n a$ and $w_1 \dots w_n b$ are subwords of α . Similarly, there exists a unique subword of length n , say $w'_1 \dots w'_n$, called the *n -th left special factor* of α and denoted by L_n , such that both $a w_1 \dots w_n$ and $b w_1 \dots w_n$ are subwords of α . In particular, for every $n \geq 1$, R_n (resp. L_n) is the unique vertex of outdegree (resp. indegree) 2 in the graph G_n .

Therefore, G_n is as in the following picture (see also [24, p. 82]).



Moreover, one can check that $R_n = \widetilde{L}_n$.

To prove the $\clubsuit$ property, we claim that one can consider the initial block of length $4n + 1$ and set $U_n = L_n$. Indeed, L_n is reached at least once within the first n steps of any path on G_n . Next, to avoid overlapping, we still take $n - 1$ steps on G_n . Starting from the reached vertex, R_n (which is $\widetilde{L}_n$) is reached within $n + 1$ steps. Thus a path of length $3n$ hits L_n and $\widetilde{L}_n$ with no overlapping, which is equivalent to our claim. $\square$

A.3 Explicit estimates in Corollary 4.6 for some families of words

We now show how condition (iii) of Corollary 4.6 can be made explicit, or even improved, for some families of words introduced in Sects. A.1 and A.2 on the alphabet $\{a, b\} \subset \mathbb{Z}[1/p]$.

Suppose that (i) and (ii) hold, i.e. $|a - b|_p \geq 1$ and $\min(|a|_p, |b|_p) \geq p$. Then:

- If ω is a Rudin–Shapiro, (iii) gives

$$(\max(|a|_\infty, |b|_\infty) + 1)^{152} < p$$

(taking $C = 8$ in Corollary 4.6, as computed in the proof of Proposition A.4).

- If ω is a paperfolding word, (iii) gives

$$(\max(|a|_\infty, |b|_\infty) + 1)^{80} < p$$

(taking $C = 4$ in Corollary 4.6 as computed in the proof of Proposition A.3 and Proposition A.4).

- If ω is Sturmian, (iii) can be improved as

$$(\max(|a|_\infty, |b|_\infty) + 1)^3 < p$$

. This is in fact equivalent to (12) from Corollary 4.5 with $c = 0$ (the latter constant being deduced from Proposition A.6. (ii)).

- If ω is Thue–Morse or characteristic Sturmian, then one can also improve (iii) as

$$(\max(|a|_\infty, |b|_\infty) + 1) < p.$$

This is in fact equivalent to (13) from Corollary 4.5 with $c = 0$ (the latter constant is deduced, respectively, from Proposition A.2. (ii) and Proposition A.6 (ii)).

A.4 Some negative examples

As for words that do *not* satisfy condition $\spadesuit$ or $\clubsuit$, an obvious example is given by those having all distinct letters from an infinite alphabet. However, words containing only two letters might

also fail to satisfy these conditions. An example is the word α obtained adjoining the words

$$w_i = \underbrace{0 \cdots 0}_{i \text{ times}} \underbrace{1 \cdots 1}_{i \text{ times}}.$$

Given any finite prefix β of α , let j be the largest index such that w_j is contained in β . Then, the longest repeated subword of β is w_{j-1} . Equivalently, the largest possible choice for U_j in β would have length $2(j-1)$. However, even in this case

$$|W_j| = \sum_{k=1}^{j-2} 2k = (j-2)(j-1),$$

so that the ratio $|W_n|/|U_n|$ tends to infinity. A similar argument shows that α also fails to satisfy condition $\clubsuit$.

The periodic word $\overline{101001000}$ is trivially $\spadesuit$ but not $\clubsuit$, as the longest palindrome it contains is 0100010 .

Finally, we construct a word that is $\clubsuit$ but not $\spadesuit$. Defining the finite words

$$\beta_i = 0 \underbrace{1 \cdots 1}_{i \text{ times}},$$

we claim that

$$\alpha = \beta_1 \widetilde{\beta_1} \beta_2 \beta_3 \widetilde{\beta_3} \widetilde{\beta_2} \cdots \beta_{2^n} \cdots \beta_{2^{n+1}-1} \widetilde{\beta_{2^{n+1}-1}} \cdots \widetilde{\beta_{2^n}} \cdots$$

is the word we are looking for. Indeed, it is $\clubsuit$ by setting $U_n = \beta_{2^n} \cdots \beta_{2^{n+1}-1}$ and $V_n = \emptyset$, so that

$$|U_n| = \sum_{i=2^n+1}^{2^{n+1}} i = \frac{(2^n+1+2^{n+1})(2^{n+1}-2^n)}{2} \geq 2^{2n}$$

and

$$|W_n| = (2+2^n)(2^n-1) \leq 2^{2n+1},$$

proving that α is $\clubsuit$ with constant 2. To conclude, suppose by contradiction that α were $\spadesuit$, and let $(U_n)_n$, $(V_n)_n$, and $(W_n)_n$ be as in Definition 1.2. From the structure of α one can easily deduce that U_n contains at most two zeroes. Let k_n be the maximum number of consecutive 1-s in U_n , so that $|U_n| \leq 3k_n + 2$. Then $|W_n| \geq (2+2^{k_n-2})(2^{k_n-2}-1) \geq 2^{2(k_n-2)}$. Since $(k_n)_n$ must be unbounded, also $(|W_n|/|U_n|)_n$ is, contradicting part (ii) of Definition 1.2.

A.5 Some more remarks on condition $\spadesuit$

Words satisfying $\spadesuit$ condition are sometimes called *stammering words* (see [19, Paragraph after Theorem 1.3]). Indeed, they are an analogue (and, to some extent, a generalisation) of the stammering words defined in [3], which are words beginning with arbitrarily large repetitions. Following [3], a word ω is called *stammering* if there exists a real number $w > 1$ and two sequences $(U_n)_n$, $(W_n)_n$ of finite words and such that conditions (ii) and (iii) hold and condition (i) is replaced by requiring that, for every $n \geq 1$, the word $W_n U_n^w$ is a prefix of ω (here U^w denotes the words obtained repeating $\lfloor w \rfloor$ times the word U , followed by the prefix of U of length $\lceil (w - \lfloor w \rfloor)|U| \rceil$). In particular, $\spadesuit$ words having $|V_n| = 0$ for all $n \geq 1$ are stammering words with $w = 2$. On the other hand, $\spadesuit$ words are a generalisation of the stammering ones for which w is an even integer. Indeed, if ω is stammering with $w = 2w'$

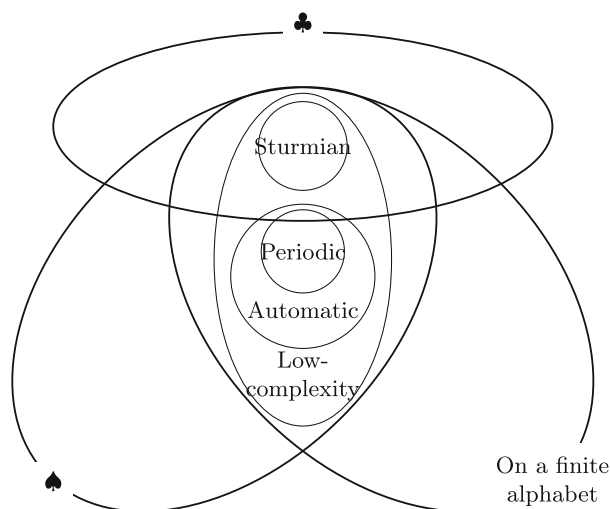


Fig. 1 Families of special words

for some $w' \in \mathbb{N}$, then there are sequences W_n and U'_n of finite words such that $W_n U_n'^{2w}$ is a prefix of ω for every n , hence ω satisfies condition ♠ with V_n the empty word and $U_n = U_n'^w$.

A.6 Some more remarks on condition ♣

As we have seen, there are only few examples and results on words satisfying condition ♣.

However, in [5] the authors proposed a general construction for the class of words satisfying condition ♣ with constant $c = 0$ i.e. sequences beginning with arbitrarily large palindromes.

Given an arbitrary sequence $(R_n)_{n \geq 0}$ of nonempty finite words whose letters are positive integers, they define a sequence of finite words $(T_n)_{n \geq 0}$ by setting $T_0 = R_0$ and

$$T_{n+1} = T_n R_{n+1} \widetilde{T_n R_{n+1}},$$

for $n \geq 0$. Thus, T_{n+1} begins with T_n , and the sequence of finite words $(T_n)_{n \geq 0}$ converges to an infinite word ω such that, if $U_n = T_n R_{n+1}$, then $U_n \widetilde{U_n}$ is a prefix of ω for every $n \geq 0$ and $|U_n|$ tends to infinity with n . They also remark that such a construction actually gives every sequence beginning with arbitrarily large palindromes.

We end this section with the following Fig. 1, which illustrates the relations between the families of words we have considered so far.

Acknowledgements The authors warmly thank the anonymous referee for their exceptionally careful reading of the manuscript and for their detailed comments, which have significantly improved the clarity, precision, and overall quality of this work. They are also deeply grateful to Yann Bugeaud for many enlightening discussions and for his precious comments on our work during his visit to the Dipartimento di Matematica G. Peano in Turin. The authors also all acknowledge the IEA (International Emerging Action) project PARIAPP (Problèmes sur l'Arithmétique et l'Algèbre des Petits Points) of the CNRS for the financial support. Laura Capuano, Marzio Mula, and Lea Terracini are members of the INdAM group GNSAGA.

Funding Open access funding provided by Università degli Studi Roma Tre within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamczewski, B., Bugeaud, Y.: On the complexity of algebraic numbers, II. Continued fractions. *Acta Math.* **195**(1), 1–20 (2005)
- Adamczewski, B., Bugeaud, Y.: A short proof of the transcendence of Thue-Morse continued fractions. *Am. Math. Mon.* **114**(6), 536–540 (2007)
- Adamczewski, B., Bugeaud, Y.: On the complexity of algebraic numbers. I: expansions in integer bases. *Ann. Math. (2)* **165**(2), 547–565 (2007)
- Adamczewski, B., Bugeaud, Y.: On the Maillet-Baker continued fractions. *J. Reine Angew. Math.* **2007**(606), 105–121 (2007)
- Adamczewski, B., Bugeaud, Y.: Palindromic continued fractions. *Ann. Inst. Fourier* **57**(5), 1557–1574 (2007)
- Adamczewski, B., Bugeaud, Y., Davison, L.: Continued fractions and transcendental numbers. *Ann. Inst. Fourier* **56**(7), 2093–2113 (2006)
- Allouche, J.-P., Davison, J., Queffélec, M., Zamboni, L.: Transcendence of Sturmian or morphic continued fractions. *J. Number Theory* **91**(1), 39–66 (2001)
- Allouche, J.-P.: The number of factors in a paperfolding sequence. *Bull. Aust. Math. Soc.* **46**, 23–32 (1992)
- Allouche, J.-P.: Schrödinger operators with Rudin-Shapiro potentials are not palindromic. *J. Math. Phys.* **38**, 1843–1848 (1997)
- Angell, D.: *Irrationality and Transcendence in Number Theory*, 1st edn. Chapman and Hall/CRC (2021)
- Allouche, J.-P., Shallit, J.: *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press (2003)
- Allouche, J.-P., Shallit, J.O.: Complexité des suites de Rudin-Shapiro généralisées. *J. Théorie Nombres Bordeaux* **5**(2), 283–302 (1993)
- Baker, A.: Continued fractions of transcendental numbers. *Mathematika* **9**(1), 1–8 (1962)
- Bombieri, E., Gubler, W.: *Heights in Diophantine Geometry*. New Mathematical Monographs, vol. 4. Cambridge University Press, Cambridge (2006)
- Bilu, Y.: The many faces of the Subspace Theorem [after Adamczewski, Bugeaud, Corvaja, Zannier...]. *Astérisque Exp. No. 967*, vii. Séminaire Bourbaki, vol. 2006/2007, pp. 1–38. (2008)
- Berthé, V., Rigo, M. (eds.): *Combinatorics, Automata and Number Theory*, vol. 135. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (2010)
- Browkin, J.: Continued fractions in local fields II. *Math. Comput.* **70**, 1281–1292 (2000)
- Browkin, J.: Continued fractions in local fields I. *Demonstratio Math.* **11**, 67–82 (1978)
- Bugeaud, Y.: Automatic continued fractions are transcendental or quadratic. *Ann. Sci. Éc. Norm. Supér. Ser. 4* **46**(6), 1005–1022 (2013)
- Cobham, A.: Uniform tag sequences. *Math. Syst. Theory* **6**(1), 164–192 (1972)
- Fischler, S.: Palindromic prefixes and episturmian words. *J. Comb. Theory A* **113**, 1281–1304 (2005)
- Laohakosol, V.: A characterization of rational numbers by p -adic Ruban continued fractions. *J. Austral. Math. Soc. Ser. A* **39**, 300–305 (1985)
- Longhi, I., Murru, N., Saettone, F.M.: Heights and Transcendence of p -adic Continued Fractions. *Ann. Mat. Pura Appl. (1923 -)* **204**, 129–145 (2025)
- Lothaire, M.: *Algebraic Combinatorics on Words*. Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press (2002)
- Mahler, K.: On a geometrical representation of p -adic numbers. *Ann. Math.* **41**, 8–56 (1940)
- Maillet, E.: *Introduction à la Théorie des Nombres Transcendants: et des Propriétés Arithmétiques des Fonctions*. Gauthier-Villars (1906)
- Morse, M., Hedlund, G.A.: Symbolic dynamics. *Am. J. Math.* **60**(4), 815–866 (1938)

28. Mol, L., Rampersad, N., Shallit, J., Stipulanti, M.: Cobham's theorem and automaticity. *Int. J. Found. Comput. Sci.* **30**(08), 1363–1379 (2019)
29. Ooto, T.: Transcendental p -adic continued fractions. *Math. Z.* **287**(3), 1053–1064 (2017)
30. Queffélec, M.: Irrational numbers with automaton-generated continued fraction expansion. In: Gambaudo, J.-M., Hubert, P., Tisseur, P., Vaienti, S. (eds.) *Dynamical Systems: From Crystal to Chaos*, pp. 190–198. World Scientific (2000)
31. Queffélec, M.: Transcendance des fractions continues de Thue-Morse. *J. Number Theory* **73**(2), 201–211 (1998)
32. Romeo, G.: Continued Fractions in the Field of p -adic Numbers. *Bull. Am. Math. Soc.* **61**, 343–371 (2024)
33. Rockett, A.M., Szűsz, P.: *Continued Fractions*. World Scientific Publishing Co., Inc, River Edge, NJ (1992)
34. Ruban, A.A.: Certain metric properties of the p -adic numbers. *Sibirsk. Mat. Ž.* **11**, 222–227 (1970)
35. Schneider, T.: Über p -adische Kettenbrüche. *Symposia Mathematica (INDAM, Rome, 1968/69)*, vol. IV, pp. 181–189. Academic Press, London (1970)
36. Schlickewei, H.P.: Die p -adische Verallgemeinerung des Satzes von Thue-Siegel-Roth-Schmidt. *J. Reine Angew. Math.* **228**, 86–105 (1976)
37. Schlickewei, H.P.: On products of special linear forms with algebraic coefficients. *Acta Arith.* **31**, 389–398 (1976)
38. Shallit, J.: Real numbers with bounded partial quotients: a survey. *Enseign. Math. (2)* **38**(1–2), 151–187 (1992)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.