Morphic and Automatic Words: Maximal Blocks and Diophantine Approximation

Yann Bugeaud
Université de Strasbourg
Mathématiques
7 rue René Descartes, 67084 Strasbourg
France
bugeaud@math.unistra.fr

Dalia Krieger
Faculty of Mathematics and Computer Science
The Weizmann Institute of Science
POB 26, Rehovot 76100
Israel
dalia.krieger@gmail.com

Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
shallit@cs.uwaterloo.ca

February 14, 2011

Abstract

Let $\mathbf{w}$ be a morphic word over a finite alphabet Σ, and let Δ be a nonempty subset of Σ. We study the behavior of maximal blocks consisting only of letters from Δ in $\mathbf{w}$, and prove the following: let (i_k, j_k) denote the starting and ending positions, respectively, of the k'th maximal Δ-block in $\mathbf{w}$. Then $\limsup_{k \to \infty} (j_k / i_k)$ is algebraic if $\mathbf{w}$ is morphic, and rational if $\mathbf{w}$ is automatic. As a result, we show that the same conclusion holds if (i_k, j_k) are the starting and ending positions of the k'th maximal zero

1

block, and, more generally, of the k'th maximal x-block, where x is an arbitrary word. This enables us to draw conclusions about the irrationality exponent of automatic and morphic numbers. In particular, we show that the irrationality exponent of automatic (resp., morphic) numbers belonging to a certain class that we define is rational (resp., algebraic).

1   Introduction

The irrationality exponent $\mu(\xi)$ of an irrational number ξ is the supremum of the real numbers μ such that the inequality

$$\left| \xi - \frac{p}{q} \right| < \frac{1}{q^\mu}$$

has infinitely many solutions in rational numbers p/q. It follows from the theory of continued fractions that $\mu(\xi) \geq 2$ for all irrational numbers ξ, and from the Borel-Cantelli lemma that $\mu(\xi) = 2$ for almost all real numbers (with respect to Lebesgue measure). However, to determine $\mu(\xi)$ for a given real number ξ is often a very difficult problem. Naïvely, we could hope to be able to read it off from the expansion of ξ in some integer base b, but this is almost never the case (see [8] for a thorough discussion).

Let $b \geq 2$ be an integer. Recently, Bugeaud [10] constructed a class $\mathcal{C}$ of real numbers whose irrationality exponent can be read off from their base-b expansion. The class $\mathcal{C}$ consists of precisely those numbers

$$\xi_{\mathbf{n}} = \sum_{j \geq 1} b^{-n_j},$$

for a strictly increasing sequence $\mathbf{n} = (n_j)_{j \geq 1}$ of positive integers satisfying $n_{j+1}/n_j \geq 2$ for every large integer j. To obtain good rational approximations to $\xi_{\mathbf{n}}$, we simply truncate the above sum. Thus, we set

$$\xi_{\mathbf{n},J} = \sum_{j=1}^{J} b^{-n_j} = \frac{p_J}{b^{n_J}}, \quad J \geq 1.$$

It then follows from

$$\left| \xi_{\mathbf{n}} - \frac{p_J}{b^{n_J}} \right| < \frac{2}{(b^{n_J})^{n_{J+1}/n_J}}, \quad J \geq 1,$$

that

$$\mu(\xi_{\mathbf{n}}) \geq \limsup_{j \to \infty} \frac{n_{j+1}}{n_j} .$$

2010 *Mathematics Subject Classification*: Primary 68R15; Secondary 11J04.
Key words and phrases: Automatic sequence, morphic sequence, irrationality exponent.

Shallit [19] proved that the continued fraction expansions of some rational translate of any such $\xi_{\mathbf{n}}$ can be given explicitly, and Bugeaud [10] proved that its irrationality exponent is given by

$$(1) \qquad \mu(\xi_{\mathbf{n}}) = \limsup_{j \to \infty} \frac{n_{j+1}}{n_j} \ ,$$

and hence can be read off from its base-b expansion; see Section 5 for a brief explanation. This means that the denominators of the best rational approximations to $\xi_{\mathbf{n}}$ are, with finitely many exceptions, powers of b. In this paper, among other results, we use this method to study the irrationality exponents of *automatic and morphic numbers*.

Let Σ be a finite alphabet, let ε denote the empty word. As usual, let Σ^* denote the set of finite words on Σ and write Σ^+ for Σ^* deprived of the empty word. Let $h : \Sigma^* \to \Sigma^*$ be a morphism. If there exists a letter $a \in \Sigma$ such that $h(a) = ax$ for some $x \in \Sigma^+$, and furthermore, $h^n(x) \neq \varepsilon$ for all $n \geq 0$ (this means that h is *prolongable* on a), then the sequence $a, h(a), h^2(a), \ldots$ converges as n tends to infinity to the infinite word $h^\omega(a) = axh(x)h^2(x)\cdots$, which is a fixed point of the morphism h. Such infinite fixed points are called *pure morphic words*. An infinite word is *morphic* if it is the image under a *coding* (that is, a letter-to-letter morphism) of a pure morphic word; it is *automatic* if it is morphic, and the underlying pure morphic word can be generated by a *uniform* morphism, that is, a morphism that maps all letters to words of equal length. (Note: the standard definition of automatic words, or sequences, uses finite automata. See, e.g., [7, Chapter 5].) A real number is *automatic* (resp., *morphic*) if its expansion in some integer base $b \geq 2$ is an automatic (resp., morphic) word over the alphabet $\Sigma_b = \{0, 1, \ldots, b-1\}$.

Recall that a Liouville number is a real number ξ satisfying $\mu(\xi) = \infty$ and that any real number greater than or equal to 2 is the irrationality exponent of some real number. Adamczewski and Cassaigne [3] established in 2006 that the irrationality exponent of an automatic number is always finite, that is, automatic numbers are not Liouville numbers. In [10], Eq. (1) was used to show that any rational number $\mu \geq 2$ is the irrationality exponent of some automatic number. These two results motivate the following question, also formulated in [4]:

Problem 1. *Determine the set of irrationality exponents of automatic numbers. In particular, is the irrationality exponent of an automatic number always rational?*

Unfortunately, we are unable to settle this problem. However, Theorem 1 below implies that for any automatic number ξ in the class $\mathcal{C}$, the quantity $\mu(\xi)$ is rational. Consequently, to give a negative answer to Problem 1 we would need to use a radically different method to construct automatic numbers than in [10]. Recently, Adamczewski and Rivoal [4] also formulated Problem 1 and they gave upper bounds for the irrationality exponent of some classical automatic numbers. (moved to Section 2)

As automatic numbers form a subclass of the morphic numbers, Problem 1 can be naturally extended as follows:

Problem 2. *Determine the set of irrationality exponents of morphic numbers.*

3

Theorem 1 implies that the irrationality exponent of every morphic number in the class $\mathcal{C}$ is always an algebraic number. Moreover, using the same method as in [10], we are able to show that every Perron number $\mu \geq 2$ is the irrationality exponent of some morphic number (recall that a *Perron number* is a positive real algebraic integer that is greater in absolute value than all of its conjugates). However, Problem 2 remains unsolved.

Theorem 1 is proved through a combinatorial study of maximal Δ-blocks in automatic and morphic words, where $\Delta \subseteq \Sigma$ is a subalphabet. We find the combinatorial results interesting in their own right.

Our paper is organized as follows. In Section 2 we give some definitions and state the main theorems, as well as some open problems. In Section 3 we analyze the structure of Δ-blocks in pure morphic words. In Section 4 we apply the results of Section 3 to morphic words in general. In Section 5 we construct, for a given Perron number $\mu \geq 2$, a morphic number ξ in $\mathcal{C}$ such that $\mu(\xi) = \mu$.

2 Exponents of Diophantine approximation and maximal blocks

To carefully investigate the question whether one can read off the irrationality exponent of a real number from its expansion in some integer base, Amou and Bugeaud [8] introduced new exponents of Diophantine approximation. Throughout the present paper, $|| \cdot ||$ denotes the distance to the nearest integer and $\lfloor \cdot \rfloor$ denotes the greatest integer function.

Definition 1. Let ξ be an irrational real number, and let $b \geq 2$ be an integer. We let $v_b(\xi)$ denote the supremum of the real numbers v for which the inequality

$$||b^n \xi|| < (b^n)^{-v}$$

has infinitely many solutions in positive integers n. We let $v_b'(\xi)$ denote the supremum of the real numbers v for which the inequality

$$||b^r(b^s - 1)\xi|| < (b^{r+s})^{-v}$$

has infinitely many solutions in positive integers r and s.

The exponent v_b measures the accuracy with which a real number is approximable by rationals obtained by truncating its base-b expansion (thus, whose denominator is a power of b), while v_b' measures the accuracy with which a real number is approximable by rationals obtained by truncating its base-b expansion and completing by periodicity. Furthermore, $v_b'(\xi)$ is the Diophantine exponent of the sequence of base-b digits of ξ, as defined in [2].

The exponents v_b, v_b' satisfy
$$v_b'(\xi) \geq v_b(\xi) \geq 0$$
and
$$\mu(\xi) \geq 1 + \max\{v_b'(\xi), 1\} \geq 1 + \max\{v_b(\xi), 1\}$$

for all irrational numbers ξ and integers $b \geq 2$. Furthermore, any real number $\xi_{\mathbf{n}} = \sum_{j \geq 1} b^{-n_j}$ belonging to the class $\mathcal{C}$ satisfies

$$(2) \qquad \mu(\xi) = 1 + v_b(\xi) = 1 + v_b'(\xi) = \limsup_{j \to \infty} \frac{n_{j+1}}{n_j}.$$

To understand the shift by 1, just observe that $\mu(\xi) - 1$ is the supremum of the real numbers μ such that $\|q\xi\| < q^{-\mu}$ has infinitely many solutions in positive integers q.

Let $b \geq 2$ be an integer. Recall that $v_b(\xi) = v_b'(\xi) = 0$ for almost all real numbers ξ. Since $v_b(\sum_{j \geq 1} b^{-\lfloor v^j \rfloor}) = v - 1$ for every real number $v > 1$, the set of values taken by the function v_b evaluated at real numbers is the interval $[0, +\infty]$.

The main result of the present paper is the following:

Theorem 1. *Let ξ be an irrational real number, and suppose the expansion of ξ in some integer base $b \geq 2$ is an automatic (resp., morphic) word over the alphabet $\{0, 1, \dots, b-1\}$. Then the number $v_b(\xi)$ is finite and rational (resp., algebraic).*

Conversely, we do not know whether for every positive algebraic number v there exist $b \geq 2$ and a morphic number ξ such that $v_b(\xi) = v$. The next theorem provides a partial result towards the resolution of this problem (see also below Corollary 3).

Theorem 2. *For every rational number $v \geq 1$ (resp., Perron number $v > 1$) and every integer $b \geq 2$ there exists a real number ξ, such that the base-b expansion of ξ is an automatic (resp., morphic) word over the alphabet $\{0, 1, \dots, b-1\}$ and $v_b(\xi) = v - 1$.*

Theorem 2 asserts that the set of values taken by the exponent v_b at automatic irrational real numbers is precisely the set of nonnegative rational numbers. The real numbers ξ constructed in the proof of Theorem 2 satisfy $\mu(\xi) = 1 + v_b(\xi)$ when $v_b(\xi) \geq 1$, but we do not know their irrationality exponent if $v_b(\xi)$ is less than 1.

We use the same method as in [10] to show the following:

Corollary 3. *For every Perron number $\mu \geq 2$ there exists a morphic number ξ such that $\mu(\xi) = \mu$.*

begin We do not know whether every real algebraic number greater than 2 is the irrationality exponent of some morphic number, but there exist morphic numbers whose irrationality exponent is algebraic but not a Perron number, as we discuss now (see also Remark 2 at the end of Section 5). In 1927 Böhmer [9] found a class of real numbers with the property that both their b-ary expansion and their continued fraction expansion are explicitly determined (this was rediscovered nearly fifty years later by Danilov [13] and, independently, by Adams and Davison [5], extending a result of Davison [14]). For a positive real number $\alpha = [0; a_1, a_2, \dots]$ in $(0, 1)$ and an integer $b \geq 2$, set

$$S_b(\alpha) = (b - 1) \sum_{j=1}^{+\infty} \frac{1}{b^{\lfloor j/\alpha \rfloor}}.$$

5

For $n \geq 1$, let p_n/q_n denote the n-th convergent to α and set

$$t_n := \frac{b^{q_n} - b^{q_{n-2}}}{b^{q_{n-1}} - 1}.$$

Then, as proved in [13, 5], we have

$$S_b(\alpha) = [0; t_1, t_2, t_3, \ldots]$$

and the irrationality exponent $\mu(S_b(\alpha))$ is given by

$$\mu(S_b(\alpha)) = 1 + \limsup_{n \to +\infty} [a_n; a_{n-1}, \ldots, a_0],$$

the latter computation being done in [1].

Following Section 2.3.6 of [18], a Sturm number is a real number whose continued fraction expansion is either of the form

$$[0; 1, a_0, \overline{a_1, \ldots, a_k}], \quad \text{with } a_k \geq a_0$$

or

$$[0; 1 + a_0, \overline{a_1, \ldots, a_k}], \quad \text{with } a_k \geq a_0 \geq 1.$$

As usual, the notation $\overline{a_1, \ldots, a_k}$ is an abbreviation for the periodic sequence with period $a_1, \ldots, a_k$. It was proved by Allauzen [6] that the sequence of base-b digits of $S_b(\alpha)$ is a fixed point of some nontrivial morphism if and only if α is a Sturm number. Putting everything together, we see that, for every finite sequence $a_1, \ldots, a_k$ of positive integers, the quadratic number

$$1 + \max_{1 \leq j \leq k} [a_j; \overline{a_{j+1}, \ldots, a_{j+k-1}, a_j}]$$

is the irrationality exponent of some morphic number, with the convention that $a_{k+h} = a_h$ for $h = 1, \ldots, k$. In particular, $[4; 2, 3, 2, 3, 2, 3, 2, \ldots] = (5 + \sqrt{15})/2$ is the irrationality exponent of a morphic number but this number is not an algebraic integer, in particular not a Perron number.

Recently, Adamczewski and Rivoal [4] gave upper bounds for the irrationality exponent of some classical automatic numbers. Furthermore, it is proved in [11] that, for $b \geq 2$, the irrationality exponent of the real number whose sequence of base-b digits is the Thue–Morse sequence on $\{0, 1\}$ is equal to 2.

end

Theorem 1 and Corollary 3 can be phrased in combinatorial terms. Let $\mathbf{w} = w_0 w_1 w_2 \cdots$ be an infinite word over $\Sigma_b = \{0, 1, \ldots, b-1\}$, and let $0 \leq i \leq j$. We say that $w_i \cdots w_j$ is a *maximal zero block* in $\mathbf{w}$ if $w_i = w_{i+1} = \cdots = w_j = 0$, $w_{j+1} \neq 0$, and either $i = 0$ or $w_{i-1} \neq 0$. Theorem 1 and Corollary 3 can be phrased in terms of the maximal zero blocks in the base-b expansion of ξ:

Theorem 4. *Let* $\mathbf{w} = w_0 w_1 w_2 \cdots$ *be an automatic (resp., morphic) word over* $\{0, 1, \ldots, b-1\}$, *that does not have a suffix of the form* 0^ω. *For* $k \geq 0$, *let* (i_k, j_k) *denote the starting and ending positions, respectively, of the k'th maximal zero block in* $\mathbf{w}$. *Then* $\limsup_{k \to \infty} j_k/i_k$ *is finite and rational (resp., algebraic).*

6

Strictly speaking, Theorem 1 is not a restatement of Theorem 4. Indeed, the exponent v_b measures not only the occurrences of zero blocks, but also the occurrences of blocks composed only of the digit $b-1$, since its definition involves the function distance to the nearest integer.

Theorem 5. *For every Perron number $\mu \geq 2$ there exist a binary morphic word $\mathbf{w} = w_0w_1w_2\cdots$, such that the sequence of indices $\{n_j\}_{j\geq 0} = \{n : w_n = 1\}$ satisfies $n_{j+1}/n_j \geq 2$ for every large integer j, and $\limsup_{j\to\infty} n_{j+1}/n_j = \mu$.*

Maximal zero blocks are a special case of *maximal x-blocks*, where $x \in \Sigma^+$ is an arbitrary word. We say that $y = w_i \cdots w_j$ is an x-block in $\mathbf{w}$ if there exist some proper suffix x' and proper prefix x'' of x, such that $y = x'x^nx''$ for some integer $n \geq 1$; if $x''w_{j+1}$ is not a prefix of x, and either $i = 0$ or $w_{i-1}x'$ is not a suffix of x, then the x-block is maximal. Theorem 4 can be generalized as follows:

Theorem 6. *Let $\mathbf{w} = w_0w_1w_2\cdots$ be an automatic (resp., morphic) word over $\{0, 1, \ldots, b-1\}$, and let $x \in \Sigma_b^+$. Assume $\mathbf{w}$ does not have a suffix of the form x^ω. For $k \geq 0$, let (i_k, j_k) denote the starting and ending positions, respectively, of the k'th maximal x-block in $\mathbf{w}$. Then $\limsup_{k\to\infty} j_k/i_k$ is finite and rational (resp., algebraic).*

Theorems 4, 6 are proved in Section 4. Theorems 5, 2 and Corollary 3 are proved in Section 5.

In light of Theorem 6, it seems plausible that the method used in the proof would allow us to say something about the exponent v_b'. This is not the case, however, since we then have to consider possible repetitions of every finite word x in Σ_b^+, and hence, to take the supremum of an infinite set of rational (resp., algebraic) numbers that may not be all distinct. We cannot guarantee that this supremum is rational (resp., algebraic), nor even that it is finite. Thus, we are unfortunately unable to establish the following statement:

Unproven Assertion (i). *Let ξ be an irrational real number, and suppose the expansion of ξ in some integer base $b \geq 2$ is an automatic (resp., morphic) word over the alphabet $\Sigma_b = \{0, 1, \ldots, b-1\}$. Then the number $v_b'(\xi)$ is finite and rational (resp., algebraic).*

If we could prove that for every irrational real number ξ, such that the expansion of ξ in some integer base $b \geq 2$ is morphic and has sublinear complexity, the number $v_b'(\xi)$ is finite, then we could extend Theorem 2.1 of Adamczewski and Cassaigne [3], asserting that the irrationality exponent of an automatic number is always finite, as follows:

Unproven Assertion (ii). *A morphic number of sublinear complexity cannot be a Liouville number.*

Lemma 5.1 of [3] states that $v_b'(\xi)$ is finite for every irrational automatic number, and is a key step in the proof of Theorem 2.1 of [3]. Here, the assumption that ξ is automatic is crucial. The other steps of the proof do not require such a strong condition on ξ and can be easily adapted to the case where ξ is morphic with sublinear complexity.

However, for numbers in the class $\mathcal{C}$, Theorem 1 and Eq. (2) imply the following corollary:

Corollary 7. *Let ξ be an automatic (resp., morphic) number in the class $\mathcal{C}$. Then $\mu(\xi)$ is finite and rational (resp., algebraic).*

3 Δ-blocks in pure morphic words

Definition 2. Let $\mathbf{w} = w_0 w_1 w_2 \cdots \in \Sigma^\omega$. Let $\mathrm{Sub}(\mathbf{w})$ denote the set of finite subwords of $\mathbf{w}$. An *occurrence* of $\mathbf{w}$ is a triple (u, i, j), where $\varepsilon \neq u \in \mathrm{Sub}(\mathbf{w})$ and $0 \leq i \leq j$, such that $w_i \cdots w_j = u$. The set of all occurrences of $\mathbf{w}$ is denoted by $\mathrm{Occ}(\mathbf{w})$. An occurrence $(u, i, j) \in \mathrm{Occ}(\mathbf{w})$ *contains* an occurrence $(u', i', j') \in \mathrm{Occ}(\mathbf{w})$, denoted $u' \prec u$, if $i \leq i'$ and $j \geq j'$.

Throughout this paper, we abuse the notation and identify an occurrence (u, i, j) with the word u. Thus, for example, if $\mathbf{w} = w_0 w_1 w_2 \cdots = h^\omega(w_0)$ for some morphism h, and (u, i, j) is an occurrence, we write simply $u = w_i \cdots w_j$. Similarly, we write $h(u)$ to denote the occurrence (v, k, m), where $v = h(u)$, and when applying h to the subword occurring at positions $i, \ldots, j$ in $\mathbf{w}$ we get a subword occurring at positions $k, \ldots, m$.

Definition 3. Let $\mathbf{w} = w_0 w_1 w_2 \cdots \in \Sigma^\omega$. Let $\Delta \subseteq \Sigma$, and let $\overline{\Delta} = \Sigma \setminus \Delta$. An occurrence $(u, i, j) \in \mathrm{Occ}(\mathbf{w})$ is a Δ-*block* if $u \in \Delta^+$. A Δ-block $(u, i, j) \in \mathrm{Occ}(\mathbf{w})$ is *maximal* if $w_{j+1} \in \overline{\Delta}$, and either $i = 0$ or $w_{i-1} \in \overline{\Delta}$.

Our goal in this section is to prove the following theorem:

Theorem 8. *Let $h : \Sigma^* \to \Sigma^*$ be a nonerasing morphism, and let $\mathbf{w} = w_0 w_1 w_2 \cdots = h^\omega(w_0)$. Let $\Delta \subsetneq \Sigma$ be a nonempty subalphabet, such that $\mathbf{w}$ contain infinitely many letters of $\overline{\Delta}$ and Δ-blocks of unbounded length. For $k = 0, 1, 2, \ldots$, let (i_k, j_k) denote the starting and ending positions, respectively, of the k'th maximal Δ-block in $\mathbf{w}$. Then $\limsup_{k \to \infty} j_k / i_k$ is an algebraic number of degree at most $|\Sigma|$. If h is also uniform, then $\limsup_{k \to \infty} j_k / i_k$ is rational.*

We require $\mathbf{w}$ to contain Δ-blocks of unbounded length because otherwise $\limsup_{k \to \infty} j_k / i_k$ is trivially rational. This condition implies in particular that $\mathbf{w}$ is not *ultimately periodic*, that is, $\mathbf{w}$ is not of the form $\mathbf{w} = xy^\omega$ for some $x \in \Sigma^*$ and $y \in \Sigma^+$.

The technique we use to prove Theorem 8 is very similar to the technique used to prove the algebraicity of critical exponents in pure morphic words [15, 16]. The idea is as follows:

1. The sequence of maximal Δ-blocks can be partitioned into subsequences, where for each subsequence, every element is an image under h of the previous element, up to a small change at the edges.

2. There are only finitely many different such subsequences in $\mathbf{w}$. Since we are interested in $\limsup$, it is enough to consider only the first of each of the different subsequences.

3. The $\limsup$ of a subsequence can be computed using the incidence matrix of h (see Definition 4). In particular, the $\limsup$ is a rational expression of the eigenvalues of the said matrix, which are algebraic numbers of degree at most $|\Sigma|$. When h is uniform, the expression turns out to be rational.

Definition 4. Let $\Sigma = \Sigma_n = \{0, 1, \ldots, n-1\}$, let $h : \Sigma_n^* \to \Sigma_n^*$, and let $u \in \Sigma_n^*$. The *Parikh vector* of u, denoted by $[u]$, is a vector of size n that counts how many times different letters occur in u: $[u] = (|u|_0, |u|_1, \ldots, |u|_{n-1})^T$. The *incidence matrix* associated with h, denoted by $A(h)$, is an $n \times n$ matrix, whose jth column is the Parikh vector of $h(j)$:

$$A(h) = (a_{i,j})_{0 \leq i,j < n} \; ; \;\; a_{i,j} = |h(j)|_i \; .$$

It is an easy induction to show that $[h(u)] = A[u]$ for all $u \in \Sigma^*$, and that $A(h^n) = A^n$ for all $n \in \mathbb{N}$. See, e.g., [7, Section 8.2].

Notation: for a word w (finite or not), $\mathrm{alph}(w)$ denotes the set of letters occurring in w.

Lemma 9. *Let $h : \Sigma^* \to \Sigma^*$. Then there exists some power g of h such that for all $a \in \Sigma$ and for all $n \geq 1$, $\mathrm{alph}(g^n(a)) = \mathrm{alph}(g(a))$.*

Proof. Let $A = A(h) = (a_{i,j})$, and denote $A^n = A(h^n) = (a_{i,j}^{(n)})$. Then for all letters $a, b \in \Sigma$ and for all $n \geq 1$, $b \in \mathrm{alph}(h^n(a))$ if and only if $|h^n(a)|_b > 0$, that is, if and only if $a_{b,a}^{(n)} > 0$. Since we care only about the zero pattern of A^n and not about the value of the non-zero entries, it is enough to consider A as a boolean matrix. Let B be a $|\Sigma| \times |\Sigma|$ boolean matrix, such that $b_{i,j} = 0$ if and only if $a_{i,j} = 0$. Then it is enough to prove the following: there exists some power B' of B, such that $B'^n = B'$ for all $n \geq 1$.

Since there are only finitely many boolean matrices of a given size, there exist some integers $t \geq 0$ and $c \geq 1$ such that $B^t = B^{t+c}$, and so $B^{t+k} = B^{t+k+nc}$ for all $k \in \{0, 1, \ldots, c-1\}$ and for all $n \geq 0$. Choose a $k \in \{0, 1, \ldots, c-1\}$ such that c divides $t+k$, and let $B' = B^{t+k}$. Then $t + k = mc$ for some integer m, and for all $n \geq 1$,

$$B'^n = B^{n(t+k)} = B^{t+k+(n-1)mc} = B^{t+k} = B'.$$

By setting $g = h^{t+k}$ we get the desired morphism. $\qquad\square$

Let $\mathbf{w} = h^\omega(a)$ be a pure morphic word over Σ. Then $\mathbf{w} = (h^t)^\omega(a)$ for all $t \geq 1$, and so we can replace h by some convenient power. Therefore, by Lemma 9, we can assume the following:

Assumption 1. For all $a \in \Sigma$ and for all $n \geq 1$, $\mathrm{alph}(h^n(a)) = \mathrm{alph}(h(a))$.

Notation: For the rest of this section, $h : \Sigma^* \to \Sigma^*$ is a nonerasing morphism satisfying Assumption 1; $M = \max\{|h(a)| : a \in \Sigma\}$; $\mathbf{w} = w_0 w_1 w_2 \cdots = h^\omega(w_0)$ is a non-ultimately periodic pure morphic word over Σ; and $\Delta \subsetneq \Sigma$ is a nonempty subalphabet, such that $\mathbf{w}$ contains infinitely many letters of $\overline{\Delta}$ and Δ-blocks of unbounded length.

Definition 5. The *inverse image* under h of an occurrence $u \in \mathrm{Occ}(\mathbf{w})$, denoted $h^{-1}(u)$, is the shortest occurrence $v \in \mathrm{Occ}(\mathbf{w})$ such that $h(v)$ contains u.

In the next two lemmas, we want to establish the following idea: if $\mathbf{w}$ contains Δ-blocks of unbounded length and infinitely many letters of $\overline{\Delta}$, then sufficiently long Δ-blocks are images under h of other Δ-blocks, except perhaps for edges of bounded length.

Lemma 10. *Let $u = w_r \cdots w_s \in \mathrm{Occ}(\mathbf{w})$ be a maximal Δ-block, with $|u| > M^2$ and $r > M$, and let $h^{-1}(u) = w_i \cdots w_j$. Then*

    1. $w_{i+M} \cdots w_{j-M}$ is a (not necessarily maximal) Δ-block;

    2. both $w_{i-M+1} \cdots w_{i+M-1}$ and $w_{j-M+1} \cdots w_{j+M-1}$ contain a letter of $\overline{\Delta}$.

Proof. Suppose there is a letter a occurring in u such that $b := h^{-1}(a) \in \overline{\Delta}$. Let $c = h^{-1}(b)$ (since a occurs at a position $k > M \geq |h(w_0)|$, $h^{-1}(b)$ is well-defined). Then by Assumption 1, $h^2(c)$ contains b, and so $h^2(c) = h^2(h^{-2}(a))$ is not contained in u. This implies that b (which is contained in $h(c)$) occurs at a distance of at most M from the edges of $h^{-1}(u)$. Therefore, $w_{i+M} \cdots w_{j-M}$ is a Δ-block.

    Now consider $w_{i-M+1} \cdots w_{i+M-1}$. Since u is maximal, $d := w_{r-1} \in \overline{\Delta}$. Let $e = h^{-2}(d)$ (again, $h^{-2}(d)$ is well-defined, since $r - 1 \geq M$). Then $h(e)$ contains d by Assumption 1. But $h(e)$ is contained in $w_{i-M+1} \cdots w_{i+M-1}$, and so $w_{i-M+1} \cdots w_{i+M-1}$ contains a letter of $\overline{\Delta}$. Similarly, $w_{j-M+1} \cdots w_{j+M-1}$ contains a letter of $\overline{\Delta}$. $\qquad\square$

Lemma 11. *Let $u = w_i \cdots w_j \in \mathrm{Occ}(\mathbf{w})$ be a maximal Δ-block, with $|u| > M^2$ and $i > M$. Then*

    1. $h(w_{i+M} \cdots w_{j-M})$ is a (not necessarily maximal) Δ-block;

    2. both $h(w_{i-M+1} \cdots w_{i+M-1})$ and $h(w_{j-M+1} \cdots w_{j+M-1})$ contain a letter of $\overline{\Delta}$.

Proof. Suppose $h(u)$ contains a letter $b \in \overline{\Delta}$. Then there exists a letter $a \in \Delta$ such that $h(a)$ contains b. By the same argument as in the proof of Lemma 10, $h(h^{-1}(a))$ contains b, and cannot be contained in u. We get that a occurs at a distance of at most M from the edges of u, and so $h(w_{i+M} \cdots w_{j-M})$ is a Δ-block. The rest is proved similarly. $\qquad\square$

Corollary 12. *The set of maximal Δ-blocks $u = w_i \cdots w_j$ that satisfy $i > M$ and $|u| > M^2$ can be partitioned into (infinitely many) sequences, each of which has the form $u^{(0)}, u^{(1)}, u^{(2)}, \cdots$, where for all $k \geq 0$, if $u^{(k)} = w_i \cdots w_j$, then*

$$h(w_{i+M} \cdots w_{j-M}) \prec u^{(k+1)} \prec h(w_{i-M+1} \cdots w_{j+M-1}).$$

Definition 6. Let $\mathbf{w} = w_0 w_1 w_2 \cdots = h^{\omega}(w_0)$ be an aperiodic pure morphic word over an alphabet Σ, let $M = \max\{|h(a)| : a \in \Sigma\}$, and let $\Delta \subsetneq \Sigma$. A Δ-*sequence* in $\mathbf{w}$ is a sequence $u^{(k)} = w_{i_k} \cdots w_{j_k}$ of maximal Δ-blocks, where for all $k \geq 0$

    • $i_k > M$;

    • $|u^{(k)}| > M^2$; and

    • $h(w_{i_k+M} \cdots w_{j_k-M}) \prec u^{(k+1)} \prec h(w_{i_k-M+1} \cdots w_{j_k+M-1})$.

Definition 7. Let $\mathbf{w} = w_0 w_1 w_2 \cdots = h^{\omega}(w_0)$ and let $\{u^{(k)}\}_{k \geq 0}$ be a Δ-sequence. For $k \geq 0$, let $h(w_{i_k}) = w_{r_{k+1}} \cdots w_{s_{k+1}}$ and $h(w_{j_k}) = w_{m_{k+1}} \cdots w_{n_{k+1}}$. Then

- $u^{(k+1)}$ is *growing* on the left if $i_{k+1} < r_{k+1}$;

- $u^{(k+1)}$ is *shrinking* on the left if $i_{k+1} > r_{k+1}$;

- $u^{(k+1)}$ is *stationary* on the left if $i_{k+1} = r_{k+1}$.

Similarly, $u^{(k+1)}$ is growing on the right if $j_{k+1} > n_{k+1}$, shrinking on the right if $j_{k+1} < n_{k+1}$, and stationary on the right if $j_{k+1} = n_{k+1}$. The *left stretch* of $u^{(k+1)}$, denoted by $\sigma^{(k+1)}$, is the word that occurs between the left edge of $u^{(k+1)}$ and the left edge of $h(u^{(k)})$. That is, if $u^{(k+1)}$ is shrinking on the left, then $\sigma^{(k+1)} := w_{r_{k+1}} \cdots w_{i_{k+1}-1}$ (in this case we say that the left stretch is *negative*); if $u^{(k+1)}$ is growing on the left, then $\sigma^{(k+1)} := w_{i_{k+1}} \cdots w_{r_{k+1}-1}$ (in this case we say that the left stretch is *positive*). Note that if $\sigma^{(k+1)}$ is positive then it is contained in $u^{(k+1)}$, and if it is negative then it borders $u^{(k+1)}$ on the left. If $r_{k+1} = i_{k+1}$ then $\sigma^{(k+1)} := \epsilon$. The *right stretch*, denoted by $\rho^{(k)}$, is defined similarly.

The k'th *left pivot*, denoted by $p_L^{(k)}$, is the rightmost letter in $w_{i_k - M+1} \cdots w_{i_k + M - 1}$ such that $h(p_L^{(k)})$ contains a letter of $\overline{\Delta}$; that is, $p_L^{(k)} := h^{-1}(w_{i_{k+1}-1})$. The *right pivot*, denoted by $p_R^{(k)}$, is defined similarly.

Figure 1 illustrates Definition 7.

Figure 1: Maximal Δ-blocks. $u^{(k+1)}$ is growing on the left and shrinking on the right. The k'th right pivot is exactly w_{j_k}. The black circles are the $\overline{\Delta}$ letters that terminate $u^{(k+1)}$.

Lemma 13. *Let $\{u^{(k)}\}_{k \geq 0}$ be a Δ-sequence in $\mathbf{w}$. Then the sequences $\{\sigma^{(k)}\}_{k \geq 1}$, $\{\rho^{(k)}\}_{k \geq 1}$ are ultimately periodic.*

Proof. We prove the lemma for the left stretch. The proof for the right stretch is similar.

For $k \geq 0$, let $p_L^{(k)} = w_{p_k}$. Then $h(w_{p_0+1} \cdots w_{i_0+M-1}) \in \Delta^+$, and so $h^n(w_{p_0+1} \cdots w_{i_0+M-1}) \in \Delta^+$ for all $n > 0$ (recall Assumption 1). In particular, $h^2(w_{p_0+1} \cdots w_{i_0+M-1}) \in \Delta^+$, and so $p_L^{(1)}$ cannot occur in $h(w_{p_0+1} \cdots w_{i_0+M-1})$. On the other hand, $\text{alph}(h^2(w_{p_0})) = \text{alph}(h(w_{p_0}))$, and so $h(w_{p_0})$ contains a letter a such that $h(a)$ contains a letter of $\overline{\Delta}$. In particular, $p_L^{(1)} \prec h(p_L^{(0)})$. More generally, for all $k > 0$, $p_L^{(k)}$ is the rightmost letter a in $h(p_L^{(k-1)})$ such

11

that $h(a)$ contains a letter of $\overline{\Delta}$. This implies that the sequence of left pivots, $\{p_L^{(k)}\}_{k\geq 0}$, is ultimately periodic: since Σ is finite, there exist some $k \neq m$ such that $p_L^{(k)} = p_L^{(m)}$, and so $p_L^{(k+n)} = p_L^{(m+n)}$ for all $n \geq 1$.

Now consider the left stretch. Let $h(p_L^{(0)}) = w_{x_1} \cdots w_{y_1}$. By definition, either $i_1 = y_1 + 1$ (if the rightmost $\overline{\Delta}$ letter is the last letter of $h(p_L^{(0)})$), or $x_1 < i_1 \leq y_1$. In the first case, $w_{i_1} \prec h(w_{p_0+1})$, and so $h(w_{i_1}) \in \Delta^+$, and $u^{(2)}$ is either growing or stationary. Since $h(p_L^{(0)})$ contains $p_L^{(1)}$ (that is, $h^2(p_L^{(0)})$ contains a letter of $\overline{\Delta}$), we get that $\sigma^{(2)}$ is a suffix of $h^2(p_L^{(0)})$.

If $x_1 < i_1 \leq y_1$, then $u^{(2)}$ can be also shrinking. However, in this case both $h(w_{i_1})$ and w_{i_2} are contained in $h^2(p_L^{(0)})$, and so $\sigma^{(2)} \prec h^2(p_L^{(0)})$. Similarly, $\sigma^{(k)} \prec h(p_L^{(k-2)})$ for all $k \geq 2$. This implies that $\{\sigma^{(k)}\}_{k\geq 1}$ is ultimately periodic. $\qquad\square$

Let $\{u^{(k)}\}_{k\geq 0}$ be a Δ-sequence. Then the sequence $\{(\sigma^{(k)}, \rho^{(k)})\}_{k\geq 1}$ is ultimately periodic. By ignoring the first few elements we can assume it is purely periodic; by replacing h by h^p, where p is the period, we can partition $\{(\sigma^{(k)}, \rho^{(k)})\}_{k\geq 1}$ into p subsequences, where each subsequence has period 1. We now compute i_k and j_k for a sequence of maximal Δ-blocks, assuming that $\sigma^{(k)}$ and $\rho^{(k)}$ are fixed.

Lemma 14. *Let $\{u^{(k)}\}_{k\geq 0}$ be a Δ-sequence, and assume that $\sigma^{(k)} = \sigma$ and $\rho^{(k)} = \rho$ for all $k \geq 0$. Let A be the incidence matrix of h, and let $\mathbf{1}$ be the all ones vector of size $1 \times |\Sigma|$. Then there exist integral vectors U, V, X, Y of size $|\Sigma| \times 1$, where U and V are nonnegative and nonzero, and a constant c, such that for all $k \geq 0$, $j_k/i_k < c$, and*

- $i_k = \mathbf{1}\left(A^k V + (\sum_{n=0}^{k-1} A^n)X \right);$

- $j_k = i_k + \mathbf{1}\left(A^k U + (\sum_{n=0}^{k-1} A^n)Y \right) - 1.$

Proof. Let $v^{(k)} = w_0 \cdots w_{i_k-1}$. Then $i_k = |v^{(k)}| = \mathbf{1}[v^{(k)}]$ and $j_k = i_k + |u^{(k)}| - 1 = i_k + \mathbf{1}[u^{(k)}] - 1$. To compute i_k and j_k we need to compute $[u^{(k)}]$ and $[v^{(k)}]$.

Let $U = [u^{(0)}]$ and $V = [v^{(0)}]$. Since $i_0 > M$, $v^{(0)}$ is a nonempty word, and so both U and V are nonnegative, nonzero vectors. Assume that $\rho = \varepsilon$. Then there are two possible situations:

1. $h(u^{(k)}) = \sigma u^{(k+1)}$ (σ is negative);

2. $\sigma h(u^{(k)}) = u^{(k+1)}$ (σ is positive).

Suppose σ is negative. Then $h(u^{(0)}) = \sigma u^{(1)}$, $h^2(u^{(0)}) = h(\sigma u^{(1)}) = h(\sigma)h(u^{(1)}) = h(\sigma)\sigma u^{(2)}$, and by induction, $h^k(u^{(0)}) = h^{k-1}(\sigma)h^{k-2}(\sigma) \cdots h(\sigma)\sigma u^{(k)}$. Then for all $k \geq 0$,

$$[u^{(k)}] = [h^k(u^{(0)})] - [h^{k-1}(\sigma)] - [h^{k-2}(\sigma)] - \cdots - [\sigma] = A^k U - \left(\sum_{\ell=0}^{k-1} A^\ell\right)[\sigma].$$

12

Now suppose that σ is positive. Then $u^{(1)} = \sigma h(u^{(0)})$, $u^{(2)} = \sigma h(u^{(1)}) = \sigma h(\sigma) h^2(u^{(0)})$, and by induction, $u^{(k)} = \sigma h(\sigma) \cdots h^{k-1}(\sigma) h^k(u^{(0)})$. Then for all $k \geq 0$,

$$[u^{(k)}] = [h^k(u^{(0)})] + [h^{k-1}(\sigma)] + [h^{k-2}(\sigma)] + \cdots + [\sigma] = A^k U + \left(\sum_{\ell=0}^{k-1} A^\ell \right) [\sigma].$$

If $\rho \neq \varepsilon$, then, depending on its sign, we get that

$$[u^{(k)}] = A^k U \pm (\sum_{\ell=0}^{k-1} A^\ell)[\sigma] \pm (\sum_{\ell=0}^{k-1} A^\ell)[\rho].$$

Similarly, $[v^{(k)}] = A^k V + (\sum_{\ell=0}^{k-1} A^\ell)[\sigma]$ if σ is negative, and $[v^{(k)}] = A^k V - (\sum_{\ell=0}^{k-1} A^\ell)[\sigma]$ if σ is positive (here the roles are inverted: if σ is negative then it is positive with respect to $v^{(k)}$, and vice versa). Let $Y = \pm[\sigma] \pm [\rho]$, $X = \mp[\sigma]$. Then for all $k \geq 0$,

$$[u^{(k)}] = A^k U + \left(\sum_{\ell=0}^{k-1} A^\ell \right) Y, \quad [v^{(k)}] = A^k V + \left(\sum_{\ell=0}^{k-1} A^\ell \right) X.$$

It remains to show that the sequence j_k/i_k is bounded by a constant. Let $w^{(k)} = w_{j_k+1} \cdots w_{i_{k+1}-1}$. Since maximal Δ-blocks are disjoint and separated by at least one letter from $\overline{\Delta}$, $|w^{(k)}| \geq 1$ for all $k \geq 0$. Now, for all $k \geq 1$,

$$\begin{aligned} |u^{(k)}| &= |h(u^{(k-1)})| \pm |\sigma| \pm |\rho| &< M|u^{(k-1)}| + 2M, \\ |v^{(k)}| &= |v^{(k-1)} u^{(k-1)} w^{(k-1)}| &> |u^{(k-1)}|, \end{aligned}$$

and so

$$\frac{j_k}{i_k} = 1 + \frac{|u^{(k)}| - 1}{|v^{(k)}|} < 1 + M + \frac{2M}{|u^{(0)}|}.$$

This completes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

The following theorem was proved in [15, 16]:

Theorem 15. *Let A be an $n \times n$ nonnegative integral matrix with no zero columns, and let U, V, W be nonnegative integral column vectors of size n, with $W \neq 0$. Let*

$$\mathcal{F}(k) = \frac{\mathbf{1} \left(A^k U + (\sum_{i=0}^{k-1} A^i) V \right)}{\mathbf{1} (A^k W)}, \quad k \geq 0.$$

Then $\{\mathcal{F}(k)\}_{k \geq 0}$ has finitely many accumulation points. Moreover, any finite accumulation point is a rational expression of the eigenvalues of A. In particular, it is algebraic of degree at most n.

The proof of Theorem 15 can be adapted, with slight changes, to the case of the sequence $\{j_k/i_k\}_{k\geq 0}$. Here we have a sequence of the form

$$\frac{j_k}{i_k} = 1 + \frac{|u^{(k)}| - 1}{|v^{(k)}|} = 1 + \frac{\mathbf{1}\left(A^k U + (\sum_{\ell=0}^{k-1} A^\ell)Y\right) - 1}{\mathbf{1}\left(A^k V + (\sum_{\ell=0}^{m-1} A^\ell)X\right)} \ , \quad k \geq 0 \ ,$$

where A is a $|\Sigma| \times |\Sigma|$ nonnegative integral matrix with no zero columns (recall that h is nonerasing), U and V are nonnegative integral vectors, both nonzero, and X and Y are integral vectors, with possibly negative entries. However, since both $|u^{(k)}|$ and $|v^{(k)}|$ are tending to infinity as k tends to infinity, both numerator and denominator are always positive. In particular, the fact that X and Y may contain negative entries does not alter the result. Also, Lemma 14 implies that all accumulation points are finite.

Corollary 16. *Under the conditions of Lemma 14, $\limsup_{k\to\infty} j_k/i_k$ is an algebraic number of degree at most $|\Sigma|$.*

Lemma 17. *Under the conditions of Lemma 14, if h is uniform then $\lim_{k\to\infty} j_k/i_k$ exists and is rational.*

Proof. If h is an m-uniform morphism, then $|h(w)| = m|w|$ for all $w \in \Sigma^*$. Let $u = |u^{(0)}|$, $v = |v^{(0)}|$, $y = \pm|\sigma| \pm |\rho|$, and $x = \mp|\sigma|$. Then the expressions for $|u^{(k)}|$ and $|v^{(k)}|$ are reduced to

$$|u^{(k)}| = m^k|u| + (\sum_{\ell=0}^{k-1} m^\ell)y = m^k|u| + y\frac{m^k - 1}{m - 1} \ ,$$

$$|v^{(k)}| = m^k|v| + (\sum_{\ell=0}^{k-1} m^\ell)x = m^k|v| + x\frac{m^k - 1}{m - 1} \ .$$

Therefore,

$$\frac{j_k}{i_k} = 1 + \frac{|u^{(k)}| - 1}{|v^{(k)}|} = 1 + \frac{m^k|u| + y\frac{m^k - 1}{m-1} - 1}{m^k|v| + x\frac{m^k-1}{m-1}} \ \xrightarrow{k \to \infty} \ 1 + \frac{(m-1)|u| + y}{(m-1)|v| + x} \ .$$

$\square$

Proof of Theorem 8. Since we are interested in $\limsup j_k/i_k$, it is enough to consider only Δ-blocks of size larger than M^2 that occur at an index $i > M$. By Corollary 12, these Δ-blocks can be partitioned into sequences, where for each sequence, an element is the image under h of the previous element, save perhaps for edges of a bounded length. Let $u = w_i \cdots w_j$ be the first element of such a sequence. Then $|u| > M^2$ and $i > M$, and so by Lemma 10, $v := h^{-1}(u)$ is a maximal Δ-block (up to the edges); however, $|v| \leq M^2$, or it would be part of the sequence itself. Taking into account the occurrences of size M on both sides of v, we get that each sequence is uniquely determined by a subword of $\mathbf{w}$ of

14

length at most $M^2 + 2M$. Since there are only finitely many such subwords, there are only finitely many different such sequences. To compute the lim sup, it is enough to consider only the first of each of the different sequences, where i_k is the smallest. Therefore, we need to consider only finitely many sequences. Each sequence can be further partitioned into finitely many subsequences, where for each of those, $\limsup j_k/i_k$ is algebraic of degree at most $|\Sigma|$ (Lemma 14). For uniform morphisms, $\limsup j_k/i_k$ is rational (Lemma 17). $\quad\square$

4 $\quad \Delta$-blocks and x-blocks in morphic words

In this section we extend Theorem 8 to morphic words in general, as described in the beginning of Section 3. First, the next theorem shows that we lose no generality by restricting ourself to nonerasing morphisms:

Theorem 18 ([7, Theorem 7.5.1]). *Every pure morphic word is the image under a coding of a pure morphic word generated by a nonerasing morphism.*

Theorem 19. *Let $\mathbf{w}$ be a morphic word over a finite alphabet Σ. Let $\Delta \subset \Sigma$ be a nonempty proper subalphabet, such that $\mathbf{w}$ contains infinitely many letters of $\overline{\Delta}$ and Δ-blocks of unbounded length. For $k = 0, 1, 2, \ldots$, let $u^{(k)} = w_{i_k} \cdots w_{j_k}$ be the k'th maximal Δ-block in $\mathbf{w}$. Then $\limsup_{k\to\infty} j_k/i_k$ is algebraic. If $\mathbf{w}$ is also automatic then $\limsup_{k\to\infty} j_k/i_k$ is rational.*

Proof. Since $\mathbf{w}$ is morphic, there exists some alphabet Σ', a morphism $h : \Sigma'^* \to \Sigma'^*$, and a coding $\tau : \Sigma'^* \to \Sigma^*$, such that $\mathbf{w} = \tau(h^\omega(a))$ for some $a \in \Sigma'$. By Theorem 18, we can assume that h is nonerasing. Let $\Delta' = \tau^{-1}(\Delta)$. Then every maximal Δ-block in $\mathbf{w}$ is the image under τ of a maximal Δ'-block in $h^\omega(a)$, and every maximal Δ'-block in $h^\omega(a)$ is mapped by τ to a maximal Δ-block in $\mathbf{w}$. The result follows from Theorem 8. $\quad\square$

Proof of Theorem 4. Set $\Delta = \{0\}$ and apply Theorem 19. $\quad\square$

Proof of Theorem 6. Let $|x| = d$, and let $X = \{(u^{(k)}, i_k, j_k) \in \mathrm{Occ}(\mathbf{w}) : k \geq 0\}$. First, we partition X into d subsequences, $X_0, \ldots X_{d-1}$, where

$$X_m = \{(u, r, s) \in X : u = x'x^n x'', \ x' \text{ is a proper suffix of } x, \text{ and } r + |x'| \equiv m \pmod{d}\}.$$

That is, X_m is the sequence of maximal x-block for which x itself begins at an index equivalent to $m \pmod{d}$. For $m = 0, 1, \ldots, d - 1$ and for $k = 0, 1, 2, \ldots$, let $(i_{k,m}, j_{k,m})$ denote the starting and ending positions of the k'th element of X_m. Then

$$(3) \qquad \limsup_{k\to\infty} \frac{j_k}{i_k} = \max \left\{ \limsup_{k\to\infty} \frac{j_{k,m}}{i_{k,m}} : 0 \leq m < d \right\}.$$

Let $\Sigma' = \{\langle a_0 \cdots a_{d-1}\rangle : a_0 \cdots a_{d-1} \in \mathrm{Sub}(\mathbf{w})\}$, and define d infinite words over Σ' by

$$\mathbf{w}_i = \langle w_i \cdots w_{i+d-1}\rangle\langle w_{i+d} \cdots w_{i+2d-1}\rangle \cdots, \quad i = 0, 1, \ldots, d - 1.$$

By [7, Theorem 7.9.1], if $\mathbf{w}$ is morphic then $\mathbf{w}_m$ is morphic for all m; by [12], if $\mathbf{w}$ is automatic then $\mathbf{w}_m$ is automatic for all m. Let $\Gamma = \Sigma \cup \{\alpha\}$, where $\alpha \notin \Sigma$, and define a d-uniform morphism $\tau : \Sigma'^* \to \Gamma^*$ by

$$\tau(\langle a_0 \cdots a_{d-1}\rangle) = \begin{cases} \alpha^d, & \text{if } a_0 \cdots a_{d-1} = x; \\ a_0 \cdots a_{d-1}, & \text{if } a_0 \cdots a_{d-1} \neq x. \end{cases}$$

Let $\mathbf{v}_m = \tau(\mathbf{w}_m)$, $m = 0, \ldots, d-1$. By [7, Corollary 7.7.5, Corollary 6.8.3], if $\mathbf{w}_m$ is morphic (resp., automatic), then so is $\mathbf{v}_m$. Let $\Delta = \{\alpha\}$, and let $(r_{k,m}, s_{k,m})$ denote the starting and ending positions of the k'th maximal Δ-block in $\mathbf{v}_m$. Then for all $k \geq 0$, $|r_{k,m} - i_{k,m}| < d$ and $|s_{k,m} - j_{k,m}| < d$, and so $\limsup_{k\to\infty} s_{k,m}/r_{k,m} = \limsup_{k\to\infty} j_{k,m}/i_{k,m}$. By Theorem 19, $\limsup_{k\to\infty} s_{k,m}/r_{k,m}$ is algebraic (resp., rational) if $\mathbf{v}_m$ is morphic (resp., automatic), and so $\limsup_{k\to\infty} j_{k,m}/i_{k,m}$ is algebraic (resp., rational) if $\mathbf{w}$ is morphic (resp., automatic). By (3), the result follows. $\qquad\square$

5 Perron numbers as irrationality exponents of morphic numbers

Proof of Theorem 5. Let $\mu > 1$ be a Perron number. Then there exists a primitive integral square matrix A, of size $k \times k$ for some positive integer k, such that $r(A) = \mu$, where $r(A)$ is the Perron-Frobenius eigenvalue of A [17, Theorem 11.1.4]. We may assume that $k \geq 2$: if μ is not integral then necessarily $k \geq 2$, and if μ is integral we can set A to be the 2×2 matrix $\left(\begin{smallmatrix} \mu-1 & \mu-1 \\ 1 & 1 \end{smallmatrix}\right)$. Let $\Sigma = \Sigma_k = \{0, 1, \ldots, k-1\}$, and let $h : \Sigma^* \to \Sigma^*$ be a morphism such that $A(h) = A$. Then

$$\frac{|h^n(0)|}{|h^{n-1}(0)|} = \frac{\mathbf{1}A^n[0]}{\mathbf{1}A^{n-1}[0]}, \quad n \geq 0 .$$

Since μ is a Perron number, the Jordan decomposition of A has one block of size 1 associated with μ, and it is easy to check that

$$\lim_{n\to\infty} \frac{|h^n(0)|}{|h^{n-1}(0)|} = \mu.$$

For $\mu > 2$, this equation implies that $|h^n(0)|/|h^{n-1}(0)| > 2$ for n sufficiently large; for $\mu = 2$, we let $A = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, and get that $|h^n(0)|/|h^{n-1}(0)| = 2$ for all n. Note that we do not require h to be prolongable on 0; we consider only the finite words $\{h^n(0)\}_{n\geq 0}$.

Let $\Gamma = \{\alpha, \beta\} \cup \Sigma$. Define a morphism $g : \Gamma^* \to \Gamma^*$ by $g(\alpha) = \alpha\beta 0$, $g(\beta) = \beta$, and $g(i) = h(i)$ for $i \in \Sigma$. Let $\mathbf{u} = g^\omega(\alpha)$. It is an easy induction to show that

$$\mathbf{u} = \alpha \cdot \beta \cdot 0 \cdot \beta \cdot h(0) \cdot \beta \cdot h^2(0) \cdot \beta \cdots .$$

Now let $\tau : \Gamma^* \to \{0,1\}^*$ be the coding that maps β to 1 and all other letters to 0, and let $\mathbf{w} = \tau(\mathbf{u})$. Then

$$\mathbf{w} = 0 \cdot 1 \cdot 0^{x_0} \cdot 1 \cdot 0^{x_1} \cdot 1 \cdot 0^{x_2} \cdot 1 \cdots ,$$

where $x_n = |h^n(0)|$. Thus, **w** is a morphic sequence that satisfies the conditions of Theorem 5. This completes the proof of the theorem.

$\square$

Proof of Theorem 2 and of Corollary 3. For morphic numbers, let $\mu > 1$ be a Perron number, let **w** be as in the proof of Theorem 5, and let $\{n_j\}_{j\geq 1} = \{n : w_n = 1\}$. Set

$$\xi_{\mathbf{w}} = \sum_{j\geq 0} b^{-n_j},$$

where $b \geq 2$ is an integer. As $\limsup_{j\to\infty} n_{j+1}/n_j = \mu$, we get that $v_b(\xi_{\mathbf{w}}) = \mu - 1$. Also, if $\mu > 2$ then $\xi_{\mathbf{w}}$ belongs to class $\mathcal{C}$, and so $\mu(\xi_{\mathbf{w}}) = \mu$. This follows from the fact that $n_{j+1} \geq 2n_j$ for j sufficiently large, which enables us to use the Folding Lemma as in [19, 10] to construct the continued fraction expansion of a rational translate of $-\xi_{\mathbf{w}}$. For sake of completeness, we give the key argument. begin The Folding Lemma asserts that, for a positive integer t and a rational number $r/s = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h]$ with $h \geq 5$ odd and $a_h \geq 2$, then we have

$$\frac{r}{s} - \frac{1}{ts^2} = \frac{rst - 1}{ts^2} = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h, t - 1, 1, a_h - 1, a_{h-1}, \ldots, a_3, 2],$$

for $t \geq 2$, and

$$\frac{r}{s} - \frac{1}{s^2} = \frac{rs - 1}{s^2} = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h + 1, a_h - 1, a_{h-1}, \ldots, a_3, 2].$$

Let J be sufficiently large in order that $n_{j+1} \geq 2n_j$ for $j \geq J$ and there exists a positive integer p such that the continued fraction expansion of p/b^{n_J} has the form $[0; 1, 1, a_3, \ldots, a_{h-1}, a_h]$ with $h \geq 5$ odd and $a_h \geq 2$. The Folding Lemma applied first with $t = b^{n_{J+1}-2n_J}$, then with $t = b^{n_{J+2}-2n_{J+1}}$, and so on, gives us the continued fraction expansion of $\xi'_{\mathbf{w}} := p/b^{n_J} - 1/b^{n_{J+1}} - 1/b^{n_{J+2}} - \ldots$ This allows us to locate the largest partial quotients of $\xi'_{\mathbf{w}}$, thus, to find its best rational approximations. In turn, this yields the exact value of the irrationality exponent of $\xi_{\mathbf{w}}$. The fact that n_{j+1} is at least equal to $2n_j$ for every large j is crucial in this argument. end

Now consider automatic numbers. For $v = 1$, we can choose any morphic binary word that does not contain unbounded 0-blocks or 1-blocks (e.g., the Thue–Morse word) to be the base-b expansion of ξ, where $b \geq 2$ is any integer, and get that $v_b(\xi) = 0 = v - 1$. Suppose $v = p/q > 1$. We define a binary infinite word $\mathbf{u} = u_0 u_1 u_2 \cdots$ by letting $u_n = 1$ if and only if n belongs to the set

$$\bigcup_{h\geq 0} \{(p)p^h, (p+1)p^h, \ldots, (qp)p^h\}.$$

Then $\mathbf{u}$ is p-automatic, as its p-kernel contains only two sequences, namely, the sequence 0^ω and the sequence $0^p 1^{qp-p+1} 0^\omega$ [7, Theorem 6.6.2]. Let $\{n_j\}_{j\geq 0} = \{n : u_n = 1\}$. Since $(p+1)/p < p/q$, we have

$$\limsup_{j\to\infty} \frac{n_{j+1}}{n_j} = \lim_{h\to\infty} \frac{p \cdot p^{h+1}}{qp \cdot p^h} = \frac{p}{q} \ .$$

Consequently, the real number

$$\xi_{\mathbf{u}} = \sum_{j \geq 1} b^{-n_j}$$

satisfies

$$v_b(\xi_{\mathbf{u}}) = \frac{p}{q} - 1.$$

$\square$

Remark 1. We stress that, with the above construction for automatic numbers, we do not know the value of $\mu(\xi_{\mathbf{u}})$, because the condition $n_{j+1}/n_j \geq 2$ for all j sufficiently large is not satisfied. This is not the case with the slightly more complicated construction given in [10], which works under the assumption that p/q exceeds 2.

Remark 2. For morphic numbers, we can go a bit further. Keep the notation of the proof of Theorem 5. Let $\Gamma = \{\alpha, \beta\} \cup \Sigma$. Let $k, a_1, \ldots, a_k$ be positive integers. Define a morphism $g : \Gamma^* \to \Gamma^*$ by $g(\alpha) = \alpha\beta 0^{a_1}\beta 0^{a_2}\beta \ldots \beta 0^{a_k}$, $g(\beta) = \beta$, and $g(i) = h(i)$ for $i \in \Sigma$. Let $\mathbf{u} = g^\omega(\alpha)$. It is an easy induction to show that

$$\mathbf{u} = \alpha \cdot \beta \cdot 0^{a_1} \cdot \beta \cdot 0^{a_2} \cdot \beta \cdots \beta \cdot 0^{a_k} \cdot (h(0))^{a_1} \cdot \beta \cdot (h(0))^{a_2} \cdot \beta \cdots \beta \cdot (h(0))^{a_k} \cdot \beta \cdot (h^2(0))^{a_1} \cdot \beta \cdots.$$

Now let $\tau : \Gamma^* \to \{0, 1\}^*$ be the coding that maps β to 1 and all other letters to 0, and let $\mathbf{w} = \tau(\mathbf{u})$. For suitable choices for $a_1, \ldots a_k$, the value of the exponent v_b at the corresponding morphic number is neither rational, nor a Perron number minus 1. However, we do not know whether every positive algebraic number can be so attained.

References

[1] B. Adamczewski and J.-P. Allouche. Reversals and palindromes in continued fractions. *Theoret. Comput. Sci.* **380** (2007), 220–237.

[2] B. Adamczewski and Y. Bugeaud. Dynamics for β-shifts and Diophantine approximation. *Ergodic Theory Dynam. Systems* **27** (2007), 1695–1711.

[3] B. Adamczewski and J. Cassaigne. Diophantine properties of real numbers generated by finite automata. *Compos. Math.* **142** (2006), 1351–1372.

[4] B. Adamczewski and T. Rivoal. Irrationality measures for some automatic real numbers. *Math. Proc. Camb. Phil. Soc.* **147** (2009), 659–678.

[5] W. W. Adams and J. L. Davison. A remarkable class of continued fractions. *Proc. Amer. Math. Soc.* **65** (1977), 194–198.

[6] C. Allauzen. Une caractérisation simple des nombres de Sturm. *J. Théor. Nombres Bordeaux* **10** (1998), 237–241.

[7] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

[8] M. Amou and Y. Bugeaud. Expansions in integer bases and exponents of Diophantine approximation. *J. London Math. Soc.* **81** (2010), 297–316.

[9] P. E. Böhmer. Über die Transzendenz gewisser dyadischer Brüche. *Math. Ann.* **96** (1927), 367–377.

[10] Y. Bugeaud. Diophantine approximation and Cantor sets. *Math. Ann.* **341** (2008), 677–684.

[11] Y. Bugeaud. On the irrationality exponent of the Thue–Morse–Mahler numbers. *Ann. Inst. Fourier (Grenoble).* To appear.

[12] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.

[13] L. V. Danilov. Certain classes of transcendental numbers. *Mat. Zametki* **12** (1972), 149–154 (in Russian). English translation in *Math. Notes* **12** (1972), 524–527.

[14] J. L. Davison. A series and its associated continued fraction. *Proc. Amer. Math. Soc.* **63** (1977), 29–32.

[15] D. Krieger. On critical exponents in fixed points of non-erasing morphisms. *Theoret. Comput. Sci.* **376** (2007), 70–88.

[16] D. Krieger. *Critical exponents and stabilizers of infinite words.* PhD thesis, Waterloo, Ontario, Canada, April 2008.

[17] D. A. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press, 1995.

[18] M. Lothaire. *Algebraic combinatorics on words.* Encyclopedia of Mathematics and its Applications, 90. Cambridge University Press, Cambridge, 2002.

[19] J. O. Shallit. Simple continued fractions for some irrational numbers II, *J. Number Theory* **14** (1982), 228–231.