Zeitschrift: L'Enseignement Mathématique
Herausgeber: Commission Internationale de l'Enseignement Mathématique
Band: 56 (2010)

Artikel: Extremal properties of (Epi)Sturmian sequences and distribution modulo 1
Autor: Allouche, Jean-Paul / Glen, Amy
DOI: https://doi.org/10.5169/seals-283523

L'Enseignement Mathématique (2) 56 (2010), 365–401

EXTREMAL PROPERTIES OF (EPI)STURMIAN SEQUENCES AND DISTRIBUTION MODULO 1

by Jean-Paul ALLOUCHE and Amy GLEN

ABSTRACT. Starting from a study of Y. Bugeaud and A. Dubickas (2005) on a question about the distribution of real numbers modulo 1 *via* combinatorics on words, we survey some combinatorial properties of (epi)Sturmian sequences and distribution modulo 1 in connection to their work. In particular we focus on extremal properties of (epi)Sturmian sequences, some of which have been rediscovered several times.

1. INTRODUCTION

Not long ago, the first author came across a paper of Y. Bugeaud and A. Dubickas [22] where the authors describe all irrational numbers $\xi > 0$ such that the fractional parts $\{\xi b^n\}$, $n \geq 0$, all belong to an interval of length $1/b$, where $b \geq 2$ is a given integer. They also prove that $1/b$ is minimal, i.e., for any interval I of length $< 1/b$, there is no irrational number $\xi > 0$, such that the fractional parts $\{\xi b^n\}$, $n \geq 0$, all belong to I. An interesting and unexpected result in their paper is the following: *the irrational numbers $\xi > 0$ such that the fractional parts $\{\xi b^n\}$, $n \geq 0$, all belong to a closed interval of length $1/b$ are exactly the positive real numbers whose base b expansions are characteristic Sturmian sequences on $\{k, k+1\}$, where $k \in \{0, 1, \ldots, b-2\}$.* We recall that *Sturmian sequences* (resp. *characteristic Sturmian sequences*) are the codings of trajectories on a square billiard that start from a side (resp. from a corner) with an irrational slope; alternatively a Sturmian (resp.

characteristic Sturmian) sequence can be obtained by coding the sequence of cuts in an integer lattice over the positive quadrant of $\mathbf{R}^2$ made by a line of irrational slope (resp. a line of irrational slope through the origin), see Theorem 4 below for some other definitions.

We will see in particular that the combinatorial results underlying [22] were stated several times, in particular by P. Veerman who proved Bugeaud-Dubickas' number-theoretical statement in the case $b = 2$ as early as 1986–1987 (see [85, 86]).

The structure of the paper is as follows. Section 2 gives the combinatorial background of Bugeaud-Dubickas' result, Section 3 gathers results on Sturmian and episturmian sequences, Sections 4 and 5 address the relevant combinatorial extremal properties (including the description of the *lexicographic world*) and the history of their (re)discoveries, while Section 6 translates everything in terms of distribution modulo 1.

2. The combinatorial background of a result of Bugeaud and Dubickas

The main result of Bugeaud and Dubickas [22, Theorem 2.1] will be recalled in Section 6. Looking at the proof, we see that its core is a result in combinatorics on words that is encompassed by Theorems 1 and 2 below.

2.1 Sturmian sequences show up

In this section sequences take their values in $\{0,1\}$. We let T denote the *shift map* defined as follows: if $\boldsymbol{s} := (s_n)_{n\geq 0}$, then $T(\boldsymbol{s}) = T((s_n)_{n\geq 0}) := (s_{n+1})_{n\geq 0}$, and we let $\leq$ denote the lexicographical order on $\{0,1\}^{\mathbf{N}}$ induced by $0 < 1$.

Theorem 1. *An aperiodic sequence $\boldsymbol{s} := (s_n)_{n\geq 0}$ on $\{0,1\}$ is Sturmian if and only if there exists a sequence $\boldsymbol{u} := (u_n)_{n\geq 0}$ on $\{0,1\}$ such that $0\boldsymbol{u} \leq T^k(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for all $k \geq 0$. Moreover, $\boldsymbol{u}$ is the unique characteristic Sturmian sequence with the same slope as $\boldsymbol{s}$, and we have $0\boldsymbol{u} = \inf\{T^k(\boldsymbol{s}),\ k \geq 0\}$ and $1\boldsymbol{u} = \sup\{T^k(\boldsymbol{s}),\ k \geq 0\}$.*

THEOREM 2. *An aperiodic sequence* $\boldsymbol{u}$ *on* $\{0,1\}$ *is a characteristic Sturmian sequence if and only if, for all* $k \geq 0$,

$$0\boldsymbol{u} < T^k(\boldsymbol{u}) < 1\boldsymbol{u}\,.$$

Furthermore, we have $0\boldsymbol{u} = \inf\{T^k(\boldsymbol{u}),\ k \geq 0\}$ *and* $1\boldsymbol{u} = \sup\{T^k(\boldsymbol{u}),\ k \geq 0\}$.

[Theorem 2 is an easy consequence of Theorem 1. For a proof of Theorem 1, see Section 5.1.]

Actually Theorem 2 was known prior to [22]. G. Pirillo (who published it in [73]) indicated it to the first author who suggested that this could well be already in a paper by S. Gan [36] under a slightly disguised form (which is indeed the case). About eight years earlier J. Berstel and P. Séébold [19] and also J.-P. Borel and F. Laubie [20] proved one direction of Theorem 2, namely that characteristic Sturmian sequences satisfy the inequalities $0\boldsymbol{u} < T^k(\boldsymbol{u}) < 1\boldsymbol{u}$ for all $k \geq 0$. In fact, it seems that both theorems were proved for the first time (including the number-theoretical aspect for the case of base 2) by P. Veerman [85, 86]. For more on the history of that result (including other papers like [23]), see Section 5 (in particular Section 5.4).

2.2 GENERALIZATIONS

Two directions for generalizations are possible. One is purely combinatorial and looks at generalizations of Sturmian sequences; in particular *episturmian sequences*, which share many properties with Sturmian sequences and have similar extremal properties. In this direction, characterizations of finite and infinite (epi)Sturmian sequences via lexicographic orderings have recently been studied (see [38, 39, 41, 49, 52, 73, 74, 75]). The other type of generalization is number-theoretic and looks at distribution modulo 1 from a combinatorial point of view. Recent papers of Dubickas go in this direction; we cite two of them showing an unexpected occurrence of the Thue-Morse sequence [30, 31] (see Section 6).

3. MORE ON STURMIAN AND EPISTURMIAN SEQUENCES

Here we give some background on Sturmian and episturmian sequences.

3.1 TERMINOLOGY & NOTATION

In what follows, we shall use the following terminology and notation from combinatorics on words (see, e.g., [66]).

Let $\mathcal{A}$ denote a finite non-empty *alphabet*. If $w = x_1x_2\cdots x_m$ is a *finite word* over $\mathcal{A}$, where each $x_i \in \mathcal{A}$, then the *length* of w is $|w| := m$, and we let $|w|_a$ denote the number of occurrences of a letter a in w. The word of length 0 is called the *empty word*, denoted by ε. The *reversal* $\widetilde{w}$ of w is given by $\widetilde{w} := x_m x_{m-1}\cdots x_1$, and if $w = \widetilde{w}$, then w is called a *palindrome*.

An *infinite word* (or simply *sequence*) $\boldsymbol{x}$ over $\mathcal{A}$ is a sequence indexed by $\mathbf{N}$ with values in $\mathcal{A}$, i.e., $\boldsymbol{x} = x_0x_1x_2\cdots$, where each $x_i \in \mathcal{A}$. A finite word w is a *factor* of $\boldsymbol{x}$ if $w = \varepsilon$ or $w = x_i\cdots x_j$ for some i, j with $i \le j$. Furthermore, if w is not empty, w is said to be a *prefix* of $\boldsymbol{x}$ if $i = 0$, and we say that w is *right* (resp. *left*) *special* if wa, wb (resp. aw, bw) are factors of $\boldsymbol{x}$ for some letters a, $b \in \mathcal{A}$, $a \neq b$. The set of all factors of $\boldsymbol{x}$ is denoted by $F(\boldsymbol{x})$, and $F_n(\boldsymbol{x})$ denotes the set of factors of length n of $\boldsymbol{x}$, i.e., $F_n(\boldsymbol{x}) := F(\boldsymbol{x}) \cap \mathcal{A}^n$. Moreover, the *alphabet* of $\boldsymbol{x}$ is $\mathrm{Alph}(\boldsymbol{x}) := F(\boldsymbol{x}) \cap \mathcal{A}$. A factor of an infinite word $\boldsymbol{x}$ is *recurrent* in $\boldsymbol{x}$ if it occurs infinitely many times in $\boldsymbol{x}$. The sequence $\boldsymbol{x}$ itself is said to be *recurrent* if all of its factors are recurrent in it. Moreover $\boldsymbol{x}$ is said to be *uniformly recurrent* (or *minimal*) if it is recurrent and if, for any factor, the gaps between its consecutive occurrences are bounded.

If u, v are non-empty words over $\mathcal{A}$, then v^ω (resp. uv^ω) denotes the *periodic* (resp. *ultimately periodic*) infinite word $vvv\cdots$ (resp. $uvvv\cdots$) having $|v|$ as a *period*. An infinite word that is not ultimately periodic is said to be *aperiodic*.

For any infinite word $\boldsymbol{x} = x_0x_1x_2x_3\cdots$, recall that the *shift map* T is defined by $T(\boldsymbol{x}) = x_1x_2x_3\cdots$. This operator naturally extends to finite words as a *circular shift* by defining $T(xw) = wx$ for any letter x and finite word w.

The set of all finite (resp. infinite) words over $\mathcal{A}$ is denoted by $\mathcal{A}^*$ (resp. $\mathcal{A}^\omega$), and we define $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$, the set of all *non-empty* words over $\mathcal{A}$.

A *morphism* on $\mathcal{A}$ is a map $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ such that $\varphi(uv) = \varphi(u)\varphi(v)$ for all words u, v over $\mathcal{A}$. Clearly a morphism on $\mathcal{A}$ is uniquely determined by its restriction to $\mathcal{A}$ ($\varphi|_{\mathcal{A}}\colon \mathcal{A} \to \mathcal{A}^*$).

3.2 Sturmian sequences

Sturmian sequences were introduced in [71]. They are in some sense the "least complicated" aperiodic sequences on a binary alphabet, as is evident from Lemma 3 and Theorem 4 below. The following lemma can essentially be found in [71].

LEMMA 3 ([71]). *Let $\mathbf{s}$ be a sequence taking exactly $a \geq 2$ distinct values. Let $p(k)$ be the number of distinct factors of length k of $\mathbf{s}$ (the function $k \mapsto p(k)$ is called the* block-complexity *of the sequence $\mathbf{s}$). Then the following properties are equivalent.*

(i) *There exists $k_0 \geq 1$ such that $p(k_0+1) = p(k_0)$.*

(ii) *The sequence $(p(k))_{k\geq 1}$ is ultimately constant (i.e., constant from some index on).*

(iii) *There exists M such that $p(k) \leq M$ for all $k \geq 1$.*

(iv) *There exists $k_1 \geq 1$ such that $p(k_1) \leq k_1 + a - 2$.*

(v) *Let $g(k) = p(k) - k$. There exists $k_2 \geq 1$ such that $g(k_2+1) < g(k_2)$.*

(vi) *The sequence $\mathbf{s}$ is ultimately periodic.*

Proof. For any sequence, we clearly have $p(k+1) \geq p(k)$ for all $k \geq 0$. This implies on the one hand that properties (ii) and (iii) are equivalent. On the other hand, this implies the equivalence of properties (i) and (v). Namely letting $g(k) := p(k) - k$, we have $g(k+1) - g(k) = p(k+1) - p(k) - 1$.

The implications (vi) $\Rightarrow$ (ii) $\Rightarrow$ (iv) are straightforward. It thus suffices to prove that (iv) $\Rightarrow$ (i), and (i) $\Rightarrow$ (vi).

(iv) $\Rightarrow$ (i): if (i) is not true, then the sequence $(p(k))_{k\geq 0}$ is (strictly) increasing. Thus, for all $k \geq 1$, one has $p(k+1) \geq p(k) + 1$. Hence, by an easy induction, one has $p(k) \geq p(1) + k - 1 = a + k - 1$, i.e., $p(k) > a + k - 2$, for all $k \geq 1$.

(i) $\Rightarrow$ (vi): the equality $p(k_0+1) = p(k_0)$ shows that $\mathbf{s}$ has no right special factor of length k_0. But this implies in turn that $\mathbf{s}$ has no right special factor of length $k_0 + 1$ (such a factor would give a right special factor of length k_0 by removing its first letter). Iterating shows that $\mathbf{s}$ has no right special factor of length k, for any $k \geq k_0$. This implies that $\mathbf{s}$ is ultimately periodic ($\mathbf{s}$ can be written as a concatenation of words of length k_0 and each of these words must always be followed by the same word).

We see from Lemma 3 above that an aperiodic sequence taking exactly a distinct values must satisfy $p(k) \geq k + a - 1$. The "simplest" aperiodic sequences would thus be sequences with the smallest $p(k)$, i.e., sequences (if any) satisfying $p(k) = k + 1$ for all $k \geq 1$. Such sequences do exist; they are called *Sturmian sequences*. They are characterized in Theorem 4 below (see, e.g., [66]). Note that Sturmian sequences admit several equivalent definitions and have numerous characterizations; for instance, they can be characterized by their palindrome or return word structure [28, 54].

THEOREM 4. *For any infinite word* $\mathbf{s}$ *over* $\{a,b\}$, *the following properties are equivalent. If* $\mathbf{s}$ *satisfies these properties, then* $\mathbf{s}$ *is called* Sturmian.

- *The number of factors of* $\mathbf{s}$ *of length* n *is equal to* $n+1$, *for all* $n \geq 1$.
- *There exist an* irrational *real number* $\alpha > 0$ *and a real number* ρ, *respectively called the* slope *and the* intercept *of* $\mathbf{s}$, *such that* $\mathbf{s}$ *is equal to one of the following two infinite words:*

$$\mathbf{s}_{\alpha,\rho},\ \mathbf{s}'_{\alpha,\rho}\colon \mathbf{N} \to \{a,b\}$$

defined by

$$\mathbf{s}_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lfloor (n+1)\alpha+\rho \rfloor - \lfloor n\alpha+\rho \rfloor = \lfloor \alpha \rfloor \\ b & \text{if } \lfloor (n+1)\alpha+\rho \rfloor - \lfloor n\alpha+\rho \rfloor \neq \lfloor \alpha \rfloor \end{cases}$$

$$\mathbf{s}'_{\alpha,\rho}(n) = \begin{cases} a & \text{if } \lceil (n+1)\alpha+\rho \rceil - \lceil n\alpha+\rho \rceil = \lfloor \alpha \rfloor \\ b & \text{if } \lceil (n+1)\alpha+\rho \rceil - \lceil n\alpha+\rho \rceil \neq \lfloor \alpha \rfloor \end{cases}$$

for $n \geq 0$ *(where* $\lfloor x \rfloor$ *denotes the greatest integer* $\leq x$ *and* $\lceil x \rceil$ *denotes the least integer* $\geq x$*). Moreover,* $\mathbf{s}$ *is said to be* characteristic Sturmian *if* $\rho = \alpha$, *in which case* $\mathbf{s} = \mathbf{s}_{\alpha,\alpha} = \mathbf{s}'_{\alpha,\alpha}$.

EXAMPLE 5. Taking $a = 0$, $b = 1$, and $\alpha = \rho = (3-\sqrt{5})/2$, we get the characteristic Sturmian sequence 01001010..., which is called the *(binary) Fibonacci sequence.*

REMARK 6. By definition it is clear that any Sturmian sequence is over a 2-letter alphabet. It also follows from Lemma 3 that Sturmian sequences are aperiodic. Note that if we choose α to be rational in the above definition, we obtain (purely) periodic sequences, referred to as *periodic balanced* sequences – see below. (Some authors also use the name *periodic Sturmian sequences.*) We will call *characteristic periodic balanced sequences* those obtained with a *rational* slope $\alpha > 0$ and intercept $\rho = \alpha$ in Theorem 4. Also note that the names "slope" and "intercept" refer to the geometric realization of Sturmian words as approximations to the line $y = \alpha x + \rho$ (called *mechanical words*, see, e.g., [66, Chapter 2]). Finally, note that, given an irrational number $\alpha \in (0,1)$, the characteristic Sturmian sequence $\mathbf{s} = (s_n)_{n\geq 1}$ of slope α is given by $s_n = 0$ if $(n+1)(1-\alpha)$ modulo 1 is in the interval $[0, 1-\alpha)$ and $s_n = 1$ otherwise, for $n \geq 1$. For example, the infinite Fibonacci word $\mathbf{f} = (f_n)_{n\geq 1} = 0100101001001010010100100 1\cdots$ (which has slope $(3-\sqrt{5})/2$) is given by $f_n = 0$ if $(n+1)(\sqrt{5}-1)/2$ modulo 1 is in the

interval $[0,(\sqrt{5}-1)/2)$ and $f_n = 1$ otherwise, for $n \geq 1$. More generally, a Sturmian sequence of slope α and intercept ρ is given by a coding over a 2-letter alphabet of the orbit of ρ under the action of the irrational rotation R: $x \mapsto x+\alpha \pmod 1$. A good reference for this description is [76, Chapter 6].

All Sturmian sequences are "balanced" in the following sense.

DEFINITION 7. A finite or infinite word w over $\{a,b\}$ is said to be *balanced* if, for any factors u, v of w with $|u| = |v|$, we have $\big||u|_b - |v|_b\big| \leq 1$ (or equivalently $\big||u|_a - |v|_a\big| \leq 1$).

The term "balanced" for this property is relatively new; it appeared in [19, 18] (see also [66, Chapter 2]), and the notion itself dates back to [71, 25]. [Note that the French term is "équilibré".] In the pioneering paper [71], balanced infinite words over a 2-letter alphabet are called "Sturmian trajectories" and belong to three classes corresponding to: Sturmian; periodic balanced; and a class of non-recurrent infinite words that are ultimately periodic (but not periodic), called *skew words*. That is, the family of balanced infinite words over $\{a,b\}$ consists of all the Sturmian and periodic balanced infinite words over $\{a,b\}$ (which are recurrent), and the (non-recurrent) skew infinite words over $\{a,b\}$, the factors of which are balanced. In particular, we have the following result due to Morse and Hedlund [71], and Coven and Hedlund [25] (see also [66, Theorem 2.1.3]):

THEOREM 8. *A binary sequence is Sturmian if and only if it is balanced and aperiodic.*

NOTE. A description of skew words is given in part (ii) of Theorem 21. Simple examples are infinite words of the form $a^\ell b a^\omega$, where $\ell \in \mathbf{N}$.

It is important to note that a finite word is *finite Sturmian* (i.e., a factor of some Sturmian word) if and only if it is balanced [66, Chapter 2, Proposition 2.1.17]. Accordingly, balanced infinite words are precisely the infinite words whose factors are finite Sturmian. This concept is generalized in [41] by showing that the set of all infinite words whose factors are *finite episturmian* consists of the (recurrent) episturmian words and the (non-recurrent) *episkew* infinite words (i.e., non-recurrent infinite words, all of whose factors are finite episturmian), see Section 3.3.2.

For a comprehensive introduction to Sturmian words, see for instance [9, 66, 76] and references therein. Also see [23, 45, 75, 83, 84] for further work on skew words.

We end this section with a simple and useful proposition which deserves to be better known. Its two parts were suggested several years ago to the first author in the case of binary sequences by J. Cassaigne and J. Berstel respectively (private communications).

PROPOSITION 9. *Let $\mathbf{s}$ be a sequence taking exactly $a \geq 2$ distinct values and let $p(k)$ be the number of distinct factors of length k of $\mathbf{s}$.*

(i) *If $\mathbf{s}$ is aperiodic and admits at most one left special factor of each length, then one has $k+a-1 \leq p(k) \leq (a-1)k+1$ for all $k \geq 1$. In particular an aperiodic binary sequence which has at most one left special factor of each length is Sturmian.*

(ii) *If there exists $k_0 \geq 1$ such that $p(k) = k+a-1$ for all $k \geq k_0$, then $p(k) = k+a-1$ for all $k \geq 1$. In particular if a binary sequence satisfies $p(k) = k+1$ for all k larger than some k_0, then it is Sturmian.*

Proof. (i) Using part (iv) of Lemma 3, we have $p(k) \geq k+a-1$ for all $k \geq 1$, since $\mathbf{s}$ is aperiodic. On the other hand, erasing the first letter of all factors of $\mathbf{s}$ of length $k+1$ gives all factors of length k. There is at most one of these factors of length k which can be obtained from distinct factors of length $k+1$ (since $\mathbf{s}$ admits at most one left special factor of length k), and if so there can be at most a such distinct factors of length $k+1$ (since a left special factor can be extended on the left by at most a letters). Hence $p(k+1) - p(k) \leq a-1$ for all $k \geq 1$. By telescopic summation, this implies $p(k) \leq (a-1)(k-1) + p(1) = (a-1)(k-1) + a = ak - k + 1$.

(ii) Let k_1 be the least integer ≥ 1 such that for all $k \geq k_1$, one has $p(k) = k + a - 1$. Suppose that $k_1 > 1$, and let $\ell := k_1 - 1$. Then $p(\ell) \neq \ell + a - 1$. But $p(\ell) \leq p(k_1) = k_1 + a - 1 = \ell + a$. Hence either $p(\ell) = \ell + a$, or $p(\ell) \leq \ell + a - 2$. In either case $\mathbf{s}$ would be ultimately periodic (by Lemma 3 (i), resp. by part (iv) of Lemma 3), a contradiction. Hence $k_1 = 1$ and the claim about Sturmianicity follows from Theorem 4.

3.3 EPISTURMIAN SEQUENCES

It is well known that the set of factors of any Sturmian sequence is closed under reversal, i.e., if u is a factor of a Sturmian sequence $\mathbf{s}$, then its reversal $\widetilde{u}$ is also a factor of $\mathbf{s}$ (e.g., see [68] or [66, Proposition 2.1.19]). In fact:

THEOREM 10. *An aperiodic binary sequence* $\boldsymbol{s}$ *is Sturmian if and only if* $F(\boldsymbol{s})$ *is closed under reversal and* $\boldsymbol{s}$ *admits exactly one left special factor of each length.*

Proof. Let $\boldsymbol{s}$ be an aperiodic binary sequence. First suppose that $\boldsymbol{s}$ is Sturmian. For a proof of the fact that $F(\boldsymbol{s})$ is closed under reversal, see [68] or [66, Proposition 2.1.19]. Now we will show that $\boldsymbol{s}$ has *exactly* one left special factor of each length.

Let $p(n)$ denote the number of factors of length n of $\boldsymbol{s}$. Since $F(\boldsymbol{s})$ is closed under reversal, a factor of $\boldsymbol{s}$ is left special (resp. right special) if and only if its reversal is right special (resp. left special). Hence, for all $n \geq 1$, the difference $p(n+1) - p(n)$ is equal to the number of left special factors of $\boldsymbol{s}$ of length n. Therefore, since $p(n+1) - p(n) = 1$ for all $n \geq 1$ (by Theorem 4), $\boldsymbol{s}$ admits exactly one left special factor (or equivalently, right special factor) of each length.

The converse follows immediately from part (i) of Proposition 9.

Inspired by results of this flavour, Droubay, Justin, and Pirillo [27, 51] introduced the following natural generalization of Sturmian sequences on an arbitrary finite alphabet $\mathcal{A}$.

DEFINITION 11 ([27]). An infinite word $\boldsymbol{t} \in \mathcal{A}^{\omega}$ is said to be *episturmian* if its set of factors $F(\boldsymbol{t})$ is closed under reversal and $\boldsymbol{t}$ admits at most one left special factor (or equivalently, right special factor) of each length.

NOTE. When $\mathcal{A}$ is a 2-letter alphabet, this definition gives the Sturmian words as well as the periodic balanced words.

In the seminal paper [27], episturmian words were defined as an extension of *standard episturmian words*, which were first introduced as a generalization of characteristic Sturmian words using *iterated palindromic closure* (a construction due to de Luca [26]).

The *palindromic right-closure* $w^{(+)}$ of a finite word w is the (unique) shortest palindrome beginning with w (see [26]). More precisely, if $w = uv$ where v is the longest palindromic suffix of w, then $w^{(+)} = uv\tilde{u}$. For example, $\texttt{(tie)}^{(+)} = \texttt{tie it}$. The *iterated palindromic closure function* [50], denoted by *Pal*, is defined recursively as follows. Set $Pal(\varepsilon) = \varepsilon$ and, for any word w and letter x, define $Pal(wx) := (Pal(w)x)^{(+)}$. For instance, $Pal(abc) = (Pal(ab)c)^{(+)} = (abac)^{(+)} = abacaba$. Note that *Pal* is injective;

and moreover, it is clear from the definition that $Pal(w)$ is a prefix of $Pal(wx)$ for any word w and letter x. Hence, if v is a prefix of w, then $Pal(v)$ is a prefix of $Pal(w)$.

THEOREM 12 ([27]). *For an infinite word $\mathbf{s} \in \mathcal{A}^\omega$, the following properties are equivalent.*

(i) *There exists an infinite word $\Delta = x_1x_2x_3\ldots$ $(x_i \in \mathcal{A})$, called the* directive word *of $\mathbf{s}$, such that $\mathbf{s} = \lim_{n\to\infty} Pal(x_1x_2\cdots x_n)$.*

(ii) *$F(\mathbf{s})$ is closed under reversal and all of the left special factors of $\mathbf{s}$ are prefixes of it.*

An infinite word $\mathbf{s}$ satisfying the above properties is said to be standard episturmian *(or* epistandard *for short).*

The above characterization of epistandard words extends to the case of an arbitrary finite alphabet a construction given in [26] for all characteristic Sturmian words.

EXAMPLE 13. The epistandard word $\mathbf{r}$ directed by $\Delta = (abc)^\omega$ is known as the *Tribonacci word*; it begins in the following way:

$$\mathbf{r} = \underline{a}b\underline{a}c\underline{a}b\underline{a}a\underline{b}a\underline{c}a\underline{b}a\underline{b}\underline{a}c\underline{a}b\underline{a}a\underline{b}a\underline{c}a\underline{b}a\underline{a}b\underline{a}c\underline{a}\cdots,$$

where each palindromic prefix $Pal(x_1\cdots x_{n-1})$ is followed by an underlined letter x_n. More generally, for $k \geq 2$, the *k-bonacci word* is the epistandard word over $\{a_1, a_2, \ldots, a_k\}$ directed by $(a_1a_2\cdots a_k)^\omega$.

REMARK 14. In [27], Droubay et al. proved that an infinite word $\mathbf{t}$ is episturmian if and only if $F(\mathbf{t}) = F(\mathbf{s})$ for some epistandard word $\mathbf{s}$. They also proved that episturmian words are uniformly recurrent; hence any such infinite word is either (purely) periodic or aperiodic. The aperiodic episturmian words are precisely the episturmian words that admit exactly one left special factor of each length. In fact, an epistandard word $\mathbf{s}$ (and hence any episturmian word with the same set of factors $\mathbf{s}$) is periodic if and only if exactly one letter occurs infinitely often in the directive word of $\mathbf{s}$ (see [51, Proposition 2.9]).

The notion of a *directive word* (as defined for epistandard words in Theorem 12) extends to *all* episturmian words with respect to *episturmian morphisms*, which play a central role in the study of these words. Introduced first as a generalization of *Sturmian morphisms*, Justin and Pirillo [51] showed that

episturmian morphisms are exactly the morphisms that preserve the aperiodic episturmian words (i.e., the morphisms that map aperiodic episturmian words onto aperiodic episturmian words). Such morphisms naturally generalize to any finite alphabet the *Sturmian morphisms* on two letters. A morphism φ is said to be *Sturmian* if $\varphi(\boldsymbol{s})$ is Sturmian for any Sturmian word $\boldsymbol{s}$. The set of Sturmian morphisms over $\{a,b\}$ is closed under composition, and consequently it is a submonoid of the endomorphisms of $\{a,b\}^*$. Moreover, it is well known that the monoid of Sturmian morphisms is generated by the three morphisms: $(a \mapsto ab, b \mapsto a)$, $(a \mapsto ba, b \mapsto a)$, $(a \mapsto b, b \mapsto a)$ (see [19, 69]; see also Section 5.2 later).

By definition (see [27, 51]), the monoid of all *episturmian morphisms* is generated, under composition, by all the morphisms:

- ψ_a: $\psi_a(a) = a$, $\psi_a(x) = ax$ for any letter $x \neq a$;
- $\tilde{\psi}_a$: $\tilde{\psi}_a(a) = a$, $\tilde{\psi}_a(x) = xa$ for any letter $x \neq a$;
- θ_{ab}: exchange of letters a and b.

Moreover, the monoid of so-called *epistandard morphisms* is generated by all the ψ_a and the θ_{ab}, and the monoid of *pure episturmian morphisms* (resp. *pure epistandard morphisms*) is generated by the ψ_a and $\tilde{\psi}_a$ only (resp. the ψ_a only). The monoid of the *permutation morphisms* (i.e., the morphisms φ such that $\varphi(\mathcal{A}) = \mathcal{A}$) is generated by all the θ_{ab}.

As shown in [51], any episturmian word is the image of another episturmian word by some pure episturmian morphism and any episturmian word can be infinitely decomposed over the set of pure episturmian morphisms. This last property allows an episturmian word to be defined by one of its morphic decompositions or, equivalently, by a certain *spinned directive word*, which is an infinite sequence of rules for decomposing the given episturmian word by morphisms. See [42, 53] for recent work concerning directive words of episturmian words.

REMARK 15. The *shift-orbit* of an infinite word $\boldsymbol{x} \in \mathcal{A}^\omega$ is the set $\mathcal{O}(\boldsymbol{x}) = \{T^i(\boldsymbol{x}),\ i \geq 0\}$ and its *closure* is given by

$$\overline{\mathcal{O}}(\boldsymbol{x}) = \left\{ \boldsymbol{y} \in \mathcal{A}^\omega,\ \mathrm{Pref}(\boldsymbol{y}) \subseteq \bigcup_{i \geq 0} \mathrm{Pref}(T^i(\boldsymbol{x})) \right\},$$

where $\mathrm{Pref}(w)$ denotes the set of prefixes of a finite or infinite word w. Note that for any infinite word $\boldsymbol{t}$ and $\boldsymbol{x} \in \overline{\mathcal{O}}(\boldsymbol{t})$, $F(\boldsymbol{x}) \subseteq F(\boldsymbol{t})$. If, moreover, $\boldsymbol{t}$ is uniformly recurrent, then it follows that for each $n \geq 1$, $F_n(\boldsymbol{x}) = F_n(\boldsymbol{t})$, and hence $F(\boldsymbol{x}) = F(\boldsymbol{t})$ for any $\boldsymbol{x} \in \overline{\mathcal{O}}(\boldsymbol{t})$ (see for instance [76, Proposition 5.1.10] or [66, Proposition 1.5.9]). This implies that $\overline{\mathcal{O}}(\boldsymbol{x}) = \overline{\mathcal{O}}(\boldsymbol{t})$ for any $\boldsymbol{x} \in \overline{\mathcal{O}}(\boldsymbol{t})$;

in other words, $\overline{\mathcal{O}}(\boldsymbol{t})$ is a *minimal dynamical system* (see, e.g., [66, 76]). Accordingly, since episturmian words are uniformly recurrent, the closure of the shift-orbit of any episturmian $\boldsymbol{t}$ is a minimal dynamical system; in particular, $\overline{\mathcal{O}}(\boldsymbol{t})$ consists of all the episturmian words with the same set of factors as $\boldsymbol{t}$ (see, e.g., [79]).

Note that if $\boldsymbol{t}$ is aperiodic, then $\overline{\mathcal{O}}(\boldsymbol{t})$ contains a unique epistandard word with the same set of factors as $\boldsymbol{t}$, whereas if $\boldsymbol{t}$ is periodic, $\overline{\mathcal{O}}(\boldsymbol{t})$ contains two different epistandard words (see for instance [40, 42]).

3.3.1 Strict episturmian words

Definition 16. An epistandard word $\boldsymbol{s}$ (or any episturmian word with the same set of factors as $\boldsymbol{s}$) is said to be *strict* if every letter in the alphabet of $\boldsymbol{s}$ occurs infinitely often in its directive word.

Strict episturmian words on k letters are often said to be *k-strict*; these words have $(k-1)n+1$ distinct factors of length n for all $n \geq 1$ (as proven in [27, p. 549]) and they coincide with the k-letter *Arnoux-Rauzy sequences*, introduced in [15] for $k = 3$ and later defined for arbitrary $k \geq 3$ in [79]. In particular, the 2-strict episturmian words are exactly the Sturmian words since these words have $n+1$ distinct factors of length n for each $n \geq 1$ (recall Theorem 4).

Note that any episturmian word takes the form $\varphi(\boldsymbol{t})$ with φ an episturmian morphism and $\boldsymbol{t}$ an Arnoux-Rauzy sequence (or strict episturmian word). In this sense, episturmian words are only a slight generalization of Arnoux-Rauzy sequences. For example, the family of episturmian words on three letters $\{a, b, c\}$ consists of the Arnoux-Rauzy sequences over $\{a, b, c\}$, the Sturmian words over $\{a, b\}$, $\{b, c\}$, $\{a, c\}$ and their images under episturmian morphisms on $\{a, b, c\}$, and periodic infinite words of the form $\varphi(x)^\omega$, where φ is an episturmian morphism on $\{a, b, c\}$ and $x \in \{a, b, c\}$.

3.3.2 Episkew words. A finite word w is said to be *finite Sturmian* (resp. *finite episturmian*) if w is a factor of some infinite Sturmian (resp. episturmian) word.

Recall from Section 3.2 that skew words are ultimately periodic (but not periodic) infinite words, all of whose factors are finite Sturmian (or equivalently, balanced). Over a 2-letter alphabet, skew words constitute the family of non-recurrent balanced infinite words, whereas the recurrent balanced infinite words consist of the Sturmian words and the periodic balanced words.

Inspired by Morse and Hedlund's [71] skew words, *episkew words* were recently defined in [41] as non-recurrent infinite words, all of whose factors are finite episturmian. A number of equivalent definitions of such words were given in [41] (see also Theorem 21, to follow).

Episkew words were first alluded to (but not explicated) in [38]. Following that paper, these words showed up again in the study of inequalities characterizing finite and infinite episturmian words in relation to lexicographic orderings [41]; in fact, as detailed in Section 5.1, episturmian words have extremal properties similar to those of Sturmian words.

To learn more about episturmian and episkew words, see for instance the recent surveys [17, 40].

4. EXTREMAL WORDS

Suppose the alphabet $\mathcal{A}$ is totally ordered by the relation $\leq$. Then we can totally order $\mathcal{A}^+$ by the *lexicographic order* $\leq$, defined as follows. Given two non-empty finite words u, v on $\mathcal{A}$, we have $u < v$ if and only if either u is a prefix of v (with $u \neq v$) or $u = xau'$ and $v = xbv'$, for some finite words x, u', v' and letters a, b with $a < b$. This is the usual alphabetic ordering in a dictionary, and we say that u is *lexicographically less* than v. This notion naturally extends to infinite words, as follows. Let $\boldsymbol{u} = u_0u_1u_2\cdots$ and $\boldsymbol{v} = v_0v_1v_2\cdots$, where u_j, $v_j \in \mathcal{A}$. We define $\boldsymbol{u} < \boldsymbol{v}$ if there exists an index $i \geq 0$ such that $u_j = v_j$ for all $j = 0, \ldots, i-1$ and $u_i < v_i$.

Let w be a finite or infinite word on $\mathcal{A}$, and let k be a positive integer. We let $\min(w|k)$ (resp. $\max(w|k)$) denote the lexicographically smallest (resp. greatest) factor of w of length k for the given order (where $|w| \geq k$ if w is finite).

If w is infinite, then it is clear that $\min(w|k)$ and $\max(w|k)$ are prefixes of the respective words $\min(w|k+1)$ and $\max(w|k+1)$. So we can define, by taking limits, the following two infinite words (see [74]):

$$\min(w) = \lim_{k\to\infty} \min(w|k) \quad \text{and} \quad \max(w) = \lim_{k\to\infty} \max(w|k).$$

That is, we can associate with any infinite word $\boldsymbol{t}$ two infinite words $\min(\boldsymbol{t})$ and $\max(\boldsymbol{t})$ such that any prefix of $\min(\boldsymbol{t})$ (resp. $\max(\boldsymbol{t})$) is the *lexicographically* smallest (resp. greatest) amongst the factors of $\boldsymbol{t}$ of the same length.

For a finite word w on a totally ordered alphabet $\mathcal{A}$, $\min(w)$ denotes $\min(w|k)$ where k is maximal such that all $\min(w|j)$, $j = 1, 2, \ldots, k$, are prefixes of $\min(w|k)$. The word $\max(w)$ is defined similarly (see [41]).

The following definition, given in [41], will be useful in the next section, where we survey recent work concerning extremal properties of (epi)Sturmian sequences, particularly inequalities characterizing such words (finite and infinite).

DEFINITION 17. An *acceptable pair* for an alphabet $\mathcal{A}$ is a pair $(a, <)$, where a is a letter in $\mathcal{A}$ and $<$ is a total order on $\mathcal{A}$ such that $a = \min(\mathcal{A})$.

5. EXTREMAL PROPERTIES

In 2003, Pirillo [73] (see also [74]) proved that, for infinite words $\boldsymbol{s}$ on a 2-letter alphabet $\{a, b\}$ with $a < b$, the inequality

$$a\boldsymbol{s} \leq \min(\boldsymbol{s}) \leq \max(\boldsymbol{s}) \leq b\boldsymbol{s} \tag{1}$$

characterizes the *characteristic Sturmian words* and *characteristic periodic balanced words*.

REMARK 18. Characteristic periodic balanced sequences, which correspond to the "Sturmian" sequences with *rational* slope $\alpha > 0$ and intercept $\rho = \alpha$ (see Theorem 4 and Remark 6), are precisely the sequences of the form $(Pal(v)xy)^\omega$, where $v \in \{a, b\}^*$ and $\{x, y\} = \{a, b\}$ (see for instance [8, 17, 27]). Also note that if $\boldsymbol{s}$ is a characteristic Sturmian sequence, then $a\boldsymbol{s} = \min(\boldsymbol{s})$ and $b\boldsymbol{s} = \max(\boldsymbol{s})$. On the other hand, if $\boldsymbol{s}$ is a characteristic periodic balanced sequence, then either:

- $a\boldsymbol{s} < \min(\boldsymbol{s})$ and $b\boldsymbol{s} = \max(\boldsymbol{s})$ when $\boldsymbol{s}$ takes the form $(Pal(v)ab)^\omega$, or
- $a\boldsymbol{s} = \min(\boldsymbol{s})$ and $\max(\boldsymbol{s}) < b\boldsymbol{s}$ when $\boldsymbol{s}$ takes the form $(Pal(v)ba)^\omega$.

For example, the characteristic periodic balanced sequence $\boldsymbol{s} := (Pal(ab)ab)^\omega = (abaab)^\omega$ satisfies

$$a\boldsymbol{s} = a(abaab)^\omega < \min(\boldsymbol{s}) = (aabab)^\omega \quad \text{and} \quad b\boldsymbol{s} = b(abaab)^\omega = \max(\boldsymbol{s}),$$

whereas $\boldsymbol{s}' := (Pal(ab)ba)^\omega = (ababa)^\omega$ satisfies

$$a\boldsymbol{s}' = a(ababa)^\omega = \min(\boldsymbol{s}') \quad \text{and} \quad \max(\boldsymbol{s}') = (babaa)^\omega < b\boldsymbol{s}' = b(ababa)^\omega.$$

More generally, given two characteristic periodic balanced sequences $\boldsymbol{s}$, $\boldsymbol{s}'$ of the form $\boldsymbol{s} = (Pal(v)ab)^\omega$ and $\boldsymbol{s}' = (Pal(v)ba)^\omega$ for some $v \in \{a, b\}^*$, we have

$$\min(\boldsymbol{s}) = \min(\boldsymbol{s}') = (aPal(v)b)^\omega \quad \text{and} \quad \max(\boldsymbol{s}) = \max(\boldsymbol{s}') = (bPal(v)a)^\omega.$$

See [8, 74] for more details.

The preceding result of Pirillo concerning characteristic Sturmian words and characteristic periodic balanced words (property (1)) encompasses Theorem 2 – one of the key properties underlying the main theorem in Bugeaud and Dubickas' paper [22]. In fact, as mentioned previously, Theorem 2 was known much earlier – in 1993, Berstel and Séébold [19] (as well as Borel and Laubie [20]) proved one direction of the theorem, namely that characteristic Sturmian words satisfy (1). This Sturmian extremal property also resurfaced in 2001, under a different guise, in a paper of S. Gan [36]. However, it seems that P. Veerman [86] was actually the first to prove (1) for Sturmian sequences in 1987, albeit from a symbolic dynamical perspective and in an implicit way. A year prior, Veerman had already proved in [85, Theorem 2] that characteristic Sturmian sequences have the above extremal property; it was not until [86, Theorem 2.1] that he proved the equivalence. Motivated by the combinatorics of the Mandelbrot set, Bullett and Sentenac [23] reproved these results of Veerman, in the language of ordered sets.

In this section, we shall first discuss the combinatorial work of Pirillo and others in relation to the inequalities (1) and their generalizations. Following this, we will consider in more detail the earlier work by Berstel and Séébold [19], Gan [36], and Veerman [85, 86].

5.1 PIRILLO'S WORK CONTINUED

Continuing his work in relation to the inequalities (1), Pirillo [74] proved further that, in the case of an arbitrary finite alphabet $\mathcal{A}$, an infinite word $\boldsymbol{s}$ on $\mathcal{A}$ is *epistandard* if and only if, for any acceptable pair $(a, <)$, we have

$$a\boldsymbol{s} \leq \min(\boldsymbol{s})\,. \tag{2}$$

Moreover, $\boldsymbol{s}$ is a strict epistandard word if and only if (2) holds with strict equality for any order [52].

In a similar spirit, Pirillo [75] defined *fine words* over two letters; that is, an infinite word $\boldsymbol{t}$ over a 2-letter alphabet $\{a, b\}$ $(a < b)$ is said to be *fine* if $(\min(\boldsymbol{t}), \max(\boldsymbol{t})) = (a\boldsymbol{s}, b\boldsymbol{s})$ for some infinite word $\boldsymbol{s}$. These infinite words were characterized in [75] by showing that fine words on $\{a, b\}$ are exactly the *Sturmian* and *skew* infinite words (see Section 3.2). Specifically:

THEOREM 19. *Let* $\boldsymbol{t}$ *be an infinite word over* $\{a,b\}$. *The following properties are equivalent:*

(i) $\boldsymbol{t}$ *is fine,*

(ii) *either* $\boldsymbol{t}$ *is Sturmian, or* $\boldsymbol{t}$ *is an ultimately periodic (but not periodic) shift of an infinite word of the form* $\mu(x^\ell y x^\omega)$ *for some* $\ell \in \mathbf{N}$, *where* μ *is a pure standard morphism on* $\{a,b\}$ *and* $\{x,y\} = \{a,b\}$ *(these words are the skew words).*

In other words, a fine word over two letters is either a Sturmian word or an ultimately periodic (but not periodic) infinite word, all of whose factors are Sturmian.

Pirillo [75] remarked that perhaps his characterization of fine words could be generalized to an arbitrary finite alphabet; indeed, Glen [38] soon generalized this result by extending Pirillo's definition of fine words to more than two letters. That is:

DEFINITION 20 ([38]). An infinite word $\boldsymbol{t}$ on $\mathcal{A}$ is said to be *fine* if there exists an infinite word $\boldsymbol{s}$ such that $\min(\boldsymbol{t}) = a\boldsymbol{s}$ for any acceptable pair $(a, <)$.

NOTE. It is easy to see that Pirillo's original 2-letter definition of a fine word is a special instance of the above definition. Certainly, as there are only two lexicographic orders on words over a 2-letter alphabet, it follows from Definition 20 that a fine word $\boldsymbol{t}$ over $\{a,b\}$ $(a < b)$ satisfies $(\min(\boldsymbol{t}), \max(\boldsymbol{t})) = (a\boldsymbol{s}, b\boldsymbol{s})$ for some infinite word $\boldsymbol{s}$.

Glen [38] characterized these generalized fine words (given in Definition 20) by showing that such an infinite word is either a *strict* episturmian word or a *strict episkew word*. More precisely:

THEOREM 21 ([38]). *Let* $\boldsymbol{t}$ *be an infinite word with* $\mathrm{Alph}(\boldsymbol{t}) = \mathcal{A}$. *Then,* $\boldsymbol{t}$ *is fine if and only if one of the following holds:*

(i) $\boldsymbol{t}$ *is an* $\mathcal{A}$*-strict episturmian word;*

(ii) $\boldsymbol{t}$ *is non-recurrent and takes the form* $\mu(x\boldsymbol{s})$, *where* x *is a letter,* $\boldsymbol{s}$ *is a strict epistandard word on* $\mathcal{A}\setminus\{x\}$, *and* μ *is a pure episturmian morphism on* $\mathcal{A}$.

REMARK 22. Note that part (ii) of Theorem 21 gives the form of so-called *strict episkew words*; it is slightly simpler to what was originally given in [41],

thanks to Richomme (private communication). Also note that strict episkew words on a 2-letter alphabet are precisely the skew words (see [40]). One can also compare Theorem 21 with Theorem 19. A simple example of an episkew word is $c\boldsymbol{f} := cabaababaaba\ldots$, where $\boldsymbol{f}$ is the Fibonacci sequence on $\{a, b\}$.

EXAMPLE 23 ([38]). Let $\mathcal{A} = \{a, b, c\}$ with $a < b < c$. Let $\boldsymbol{f}$ denote the infinite Fibonacci word over $\{a, b\}$, i.e., the epistandard word directed by $(ab)^\omega$. Then, the following infinite words are fine.

- $\boldsymbol{f} = abaababaabaaba\cdots$
- $c\boldsymbol{f} = \underline{c}abaababaabaaba\cdots$
- $\widetilde{f}_4 c\boldsymbol{f} = aaba\underline{c}abaababaabaaba\cdots$
- $\psi_a(\boldsymbol{f}) = aabaaabaabaaabaaaba\cdots$
- $\psi_c(c\boldsymbol{f}) = \underline{c}cacbcacacbcacbcacacbcacacbca\cdots$
- $\psi_c(\widetilde{f}_4 c\boldsymbol{f}) = cacacbca\underline{c}cacbcacacbcacbcacacbcaca\cdots$

Let us note, for example, that $\psi_c(\boldsymbol{f})$ is *not* fine since it is a *non-strict* epistandard word. That is, $\psi_c(\boldsymbol{f})$ is an epistandard word with directive word $c(ab)^\omega$, so it is not strict, nor does it take the second form given in Theorem 21.

Continuing this work, Glen, Justin, and Pirillo [41] recently proved new characterizations of *finite* Sturmian and episturmian words via lexicographic orderings. As a consequence, they were able to characterize by lexicographic order all episturmian words in a *wide sense* (episturmian and episkew infinite words). Similarly, they characterized by lexicographic order all balanced infinite words on a 2-letter alphabet; in other words, all Sturmian, periodic balanced, and skew infinite words, the factors of which are (finite) Sturmian.

In the finite case:

THEOREM 24 ([41]). *A finite word w on $\mathcal{A}$ is episturmian if and only if there exists a finite word u on $\mathcal{A}$ such that, for any acceptable pair $(a, <)$, we have*

$$au_{|m|-1} \leq m\,, \tag{3}$$

where $m = \min(w)$ for the considered order.

A corollary of Theorem 24 is the following new characterization of finite Sturmian words (i.e., finite balanced words).

COROLLARY 25 ([41]). *A finite word w on $\mathcal{A} = \{a, b\}$, $a < b$, is not Sturmian (in other words, not balanced) if and only if there exists a finite word $u \in \{a, b\}^*$ such that aua is a prefix of $\min(w)$ and bub is a prefix of $\max(w)$.*

In the infinite case, a characterization of episturmian words in the *wide sense* follows almost immediately from Theorem 24. That is:

COROLLARY 26 ([41]). *An infinite word $\boldsymbol{t}$ on $\mathcal{A}$ is episturmian in the wide sense (i.e., episturmian or episkew) if and only if there exists an infinite word $\boldsymbol{u}$ on $\mathcal{A}$ such that*

$$a\boldsymbol{u} \leq \min(\boldsymbol{t})$$

for any acceptable pair $(a, <)$.

Consequently, an infinite word $\boldsymbol{s}$ on $\{a, b\}$ $(a < b)$ is balanced (i.e., Sturmian, periodic balanced, or skew) if and only if there exists an infinite word $\boldsymbol{u}$ on $\{a, b\}$ such that

$$a\boldsymbol{u} \leq \min(\boldsymbol{s}) \leq \max(\boldsymbol{s}) \leq b\boldsymbol{u}\,. \tag{4}$$

For any sequence $\boldsymbol{s}$, $\max(\boldsymbol{s})$ is the same as $\sup\{T^k(\boldsymbol{s}), k \geq 0\}$, and similarly $\min(\boldsymbol{s}) = \inf\{T^k(\boldsymbol{s}), k \geq 0\}$, where the infimum and supremum are taken with respect to the lexicographic order. The preceding result therefore shows that a sequence $\boldsymbol{s}$ in $\{0,1\}^\omega$ is balanced if and only if there exists a sequence $\boldsymbol{u} \in \{0,1\}^\omega$ such that $0\boldsymbol{u} \leq T^k(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for all $k \geq 0$. In particular, a sequence $\boldsymbol{s}$ on $\{0,1\}$ being Sturmian is equivalent to $\boldsymbol{s}$ being aperiodic and the existence of a sequence $\boldsymbol{u}$ on $\{0,1\}$ such that $0\boldsymbol{u} \leq T^k(\boldsymbol{s}) \leq 1\boldsymbol{u}$. Moreover, it follows from the proof of Theorem 19 (or Theorem 21) that $\boldsymbol{u}$ is the unique characteristic Sturmian sequence having the same slope as $\boldsymbol{s}$. This is exactly Theorem 1. For the sake of completeness, we give a direct proof here.

Direct proof of Theorem 1. Let $\boldsymbol{s}$ be an aperiodic sequence on $\{0,1\}$. First suppose that $\boldsymbol{s}$ is a Sturmian sequence. Since it contains both 0's and 1's, there exist two binary sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $0\boldsymbol{x} := \inf\{T^k(\boldsymbol{s}),\ k \geq 0\}$ and $1\boldsymbol{y} := \sup\{T^k(\boldsymbol{s}),\ k \geq 0\}$. We claim that $\boldsymbol{x} \geq \boldsymbol{y}$. Namely, if $\boldsymbol{x} < \boldsymbol{y}$, there exist a (possibly empty) word w and two infinite sequences $\boldsymbol{x}'$ and $\boldsymbol{y}'$ such that $\boldsymbol{x} = w0\boldsymbol{x}'$ and $\boldsymbol{y} = w1\boldsymbol{y}'$. Hence $0\boldsymbol{x} = 0w0\boldsymbol{x}'$ and $1\boldsymbol{y} = 1w1\boldsymbol{y}'$. Since

any factor of $\inf\{T^k(\boldsymbol{s}),\ k \geq 0\}$ and of $\sup\{T^k(\boldsymbol{s}),\ k \geq 0\}$ is a factor of $\boldsymbol{s}$, we have that both $0w0$ and $1w1$ are factors of $\boldsymbol{s}$. Hence $\boldsymbol{s}$ is unbalanced (see Definition 7 and the comments following it), but it was supposed Sturmian, a contradiction (Theorem 8). Thus $\boldsymbol{x} \geq \boldsymbol{y}$, and hence

$$\forall k \geq 0\,, \quad 0\boldsymbol{x} \leq T^k(\boldsymbol{s}) \leq 1\boldsymbol{y} \leq 1\boldsymbol{x}\,.$$

Now suppose that $\boldsymbol{s}$ has the property that there exists a binary sequence $\boldsymbol{u}$ such that

$$\forall k \geq 0\,, \quad 0\boldsymbol{u} \leq T^k(\boldsymbol{s}) \leq 1\boldsymbol{u}\,. \tag{5}$$

Let z be a left special factor (if any) of $\boldsymbol{s}$, and let z' be the prefix of $\boldsymbol{u}$ that has the same length as z. Since $0z$ and $1z$ are both factors of $\boldsymbol{s}$, there exist two integers ℓ_1 and ℓ_2 such that $T^{\ell_1}(\boldsymbol{s})$ begins with $0z$ and $T^{\ell_2}(\boldsymbol{s})$ begins with $1z$. We deduce from the inequalities (5) with $k = \ell_1$ (resp. ℓ_2) that

$$0z' \leq 0z \quad \text{and} \quad 1z \leq 1z'.$$

This implies

$$z' \leq z \quad \text{and} \quad z \leq z'$$

hence $z = z'$. Thus $\boldsymbol{s}$ has at most one left special factor of each length. Hence $\boldsymbol{s}$ is Sturmian (Proposition 9), and its left special factors are exactly the prefixes of $\boldsymbol{u}$.

This implies furthermore that $\boldsymbol{u}$ belongs to the closure of the shift-orbit of $\boldsymbol{s}$, hence it is Sturmian. But the prefixes of $0\boldsymbol{u}$ and $1\boldsymbol{u}$ are also factors of $\boldsymbol{s}$. Hence $0\boldsymbol{u}$ and $1\boldsymbol{u}$ are also in the closure of the shift-orbit of $\boldsymbol{s}$, thus Sturmian. This implies that $\boldsymbol{u}$ is Sturmian characteristic (see, e.g., [66, Proposition 2.1.22]). Thus $\boldsymbol{u}$ is the (unique) characteristic Sturmian sequence having the same slope as $\boldsymbol{s}$.

REMARK 27. We noted in the Introduction that Theorem 2 can be easily deduced from Theorem 1. Actually Theorem 1 can also be deduced from Theorem 2: it suffices to remember that the closure of the shift-orbit of a characteristic Sturmian sequence $\boldsymbol{u}$ is exactly the set of all Sturmian sequences having the same slope as $\boldsymbol{u}$ (see for instance [66, Proposition 2.1.25]), and all of these Sturmian sequences have the same set of factors ([66, Proposition 2.1.18], or [68]). See also Remark 33 later.

Recently, Richomme [78] proved that episturmian words can be characterized via a nice "local balance property". That is:

THEOREM 28 ([78]). *For a recurrent infinite word $\mathbf{t} \in \mathcal{A}^\omega$, the following assertions are equivalent:*

(i) *$\mathbf{t}$ is episturmian;*

(ii) *for each factor u of $\mathbf{t}$, there exists a letter a such that $\mathcal{A}u\mathcal{A} \cap F(\mathbf{t}) \subseteq au\mathcal{A} \cup \mathcal{A}ua$;*

(iii) *for each palindromic factor u of $\mathbf{t}$, there exists a letter a such that $\mathcal{A}u\mathcal{A} \cap F(\mathbf{t}) \subseteq au\mathcal{A} \cup \mathcal{A}ua$.*

Roughly speaking, the above theorem says that for any factor u of a given episturmian word $\mathbf{t}$, there exists a unique letter a such that every occurrence of u in $\mathbf{t}$ is immediately preceded or followed by a in $\mathbf{t}$. When $|\mathcal{A}| = 2$, property (ii) of Theorem 28 is equivalent to the definition of balance. Indeed, Coven and Hedlund [25] stated that an infinite word $\mathbf{s}$ over $\{a, b\}$ is not balanced if and only if there exists a palindrome u such that aua and bub are both factors of $\mathbf{s}$. As pointed out in [78], this property can be rephrased as follows: an infinite word $\mathbf{s}$ is Sturmian if and only if $\mathbf{s}$ is aperiodic and, for any factor u of $\mathbf{s}$, the set of factors belonging to $\mathcal{A}u\mathcal{A}$ is a subset of $au\mathcal{A} \cup \mathcal{A}ua$ or a subset of $bu\mathcal{A} \cup \mathcal{A}ub$.

REMARK 29. Recall that the set of all infinite words in $\mathcal{A}^\omega$ having episturmian factors consists of the (recurrent) episturmian words and the (non-recurrent) episkew words in $\mathcal{A}^\omega$. Therefore, since properties (ii) and (iii) in Theorem 28 concern only factors, one readily deduces that these properties in fact characterize the episturmian and episkew words in $\mathcal{A}^\omega$. So the recurrence hypothesis in the statement of the theorem restricts attention to episturmian words only.

We will now use Theorem 28 to give an alternative (simpler) proof of the following analogue of Theorem 1 for episturmian sequences, which was originally proved in [39] (also see [41]). This result, in particular, gives a more precise version of Corollary 26 under the recurrence hypothesis.

THEOREM 30. *A recurrent infinite word $\mathbf{t}$ on $\mathcal{A}$ is episturmian if and only if there exists an infinite word $\mathbf{u}$ on $\mathcal{A}$ such that, for any acceptable pair $(a, <)$,*

$$a\mathbf{u} \leq T^i(\mathbf{t}) \quad \text{for all } i \geq 0 .$$

Moreover, if $\mathbf{t}$ is aperiodic, then $\mathbf{u}$ is the unique epistandard word with the

same set of factors as $\boldsymbol{t}$ *(i.e., the unique epistandard word in the closure of the shift-orbit of* $\boldsymbol{t}$*), and for any acceptable pair* $(a, <)$, $a\boldsymbol{u} = \inf\{T^k(\boldsymbol{t}),\ k \geq 0\}$ *if and only if the letter* a *occurs infinitely often in the directive word of* $\boldsymbol{u}$.

Proof. Let $\boldsymbol{t}$ be a recurrent infinite word on $\mathcal{A}$.

First suppose that $\boldsymbol{t}$ is episturmian. Let x be a letter in $\mathcal{A}$ and consider two different total orders $<_1$ and $<_2$ on $\mathcal{A}$ such that $(x, <_1)$ and $(x, <_2)$ are acceptable pairs. Then there exist infinite words $\boldsymbol{u}$ and $\boldsymbol{v}$ on $\mathcal{A}$ such that

$$x\boldsymbol{u} = \inf_1\{T^k(\boldsymbol{t}), k \geq 0\} \quad \text{for the total order } <_1 \text{ on } \mathcal{A}\,, \tag{6}$$

and

$$x\boldsymbol{v} = \inf_2\{T^k(\boldsymbol{t}), k \geq 0\} \quad \text{for the total order } <_2 \text{ on } \mathcal{A}\,. \tag{7}$$

(Here, $\inf_i$ denotes the infimum with respect to the order $<_i$ for $i = 1, 2$.) We will show that $\boldsymbol{u} = \boldsymbol{v}$. By equations (6) and (7), we have

$$x\boldsymbol{u} \leq_1 x\boldsymbol{v} \quad \text{and} \quad x\boldsymbol{v} \leq_2 x\boldsymbol{u}\,.$$

Hence, if u and v are prefixes of the respective words $\boldsymbol{u}$ and $\boldsymbol{v}$ with $|u| = |v|$, then we have $u \leq_1 v$ and $v \leq_2 u$. This implies that $u = v$, and therefore $\boldsymbol{u} = \boldsymbol{v}$. Hence, for a given letter x in $\mathcal{A}$, there exists a unique infinite word $\boldsymbol{u}$ on $\mathcal{A}$ such that

$$x\boldsymbol{u} = \inf_x\{T^k(\boldsymbol{t}), k \geq 0\} \quad \text{for any acceptable pair } (x, <_x)\,. \tag{8}$$

Now consider another letter y in $\mathcal{A} \setminus \{x\}$. By what precedes, we know there exists a unique infinite word $\boldsymbol{v}$ on $\mathcal{A}$ such that

$$y\boldsymbol{v} = \inf_y\{T^k(\boldsymbol{t}), k \geq 0\} \quad \text{for any acceptable pair } (y, <_y)\,. \tag{9}$$

Again, we will show that $\boldsymbol{u} = \boldsymbol{v}$. Suppose not. Then there exist a (possibly empty) word w and two infinite words $\boldsymbol{u}'$ and $\boldsymbol{v}'$ over $\mathcal{A}$ such that $\boldsymbol{u} = wz_1\boldsymbol{u}'$ and $\boldsymbol{v} = wz_2\boldsymbol{v}'$ for some letters z_1 and z_2 with $z_1 \neq z_2$. Hence $x\boldsymbol{u} = xwz_1\boldsymbol{u}'$ and $y\boldsymbol{v} = ywz_2\boldsymbol{v}'$, and therefore the words xwz_1 and ywz_2 are both factors of $\boldsymbol{t}$, since any factor of $x\boldsymbol{u}$ and of $y\boldsymbol{v}$ is also a factor of $\boldsymbol{t}$ (by (8) and (9)). But then, by Richomme's local balance property (Theorem 28), $z_2 = x$ or $z_1 = y$.

If $z_2 = x$, then for any acceptable pair $(x, <_x)$, we have $x <_x z_1$ (since $z_1 \neq z_2$), and hence $x\boldsymbol{v}$ $(= xwx\boldsymbol{v}')$ $<_x$ $x\boldsymbol{u}$ $(= xwz_1\boldsymbol{u}')$, contradicting the (lexicographical) minimality of $\boldsymbol{u}$ with respect to the total order $<_x$. Likewise, if $z_1 = y$, then for any acceptable pair $(y, <_y)$, we have $y <_y z_2$ (since $z_1 \neq z_2$), and hence $y\boldsymbol{u}$ $(= ywz_1\boldsymbol{u}')$ $<_y$ $y\boldsymbol{v}$ $(= ywz_2\boldsymbol{v}')$, a contradiction. Thus $\boldsymbol{u} = \boldsymbol{v}$.

Hence, there exists a (unique) infinite word $\boldsymbol{u}$ on $\mathcal{A}$ such that, for any acceptable pair $(a, <)$, $a\boldsymbol{u} \leq T^i(\boldsymbol{t})$ for all $i \geq 0$.

Conversely, suppose there exists an infinite word $\boldsymbol{u}$ on $\mathcal{A}$ such that, for any acceptable pair $(a, <)$, we have

$$a\boldsymbol{u} \leq T^i(\boldsymbol{t}) \quad \text{for all } i \geq 0. \tag{10}$$

Let z be a left special factor (if any) of $\boldsymbol{t}$, and let z' denote the prefix of $\boldsymbol{u}$ with $|z'| = |z|$. Since z is left special in $\boldsymbol{t}$, there exist at least two distinct letters x, y such that xz and yz are both factors of $\boldsymbol{t}$. In particular, there exist non-negative integers ℓ_1 and ℓ_2 such that $T^{\ell_1}(\boldsymbol{t})$ begins with xz and $T^{\ell_2}(\boldsymbol{t})$ begins with yz. Thus, by inequality (10), we have

$$xz' \leq_x xz \quad \text{for any acceptable pair } (x, <_x),$$

and

$$yz' \leq_y yz \quad \text{for any acceptable pair } (y, <_y).$$

Hence $z' \leq_x z$ and $z' \leq_y z$, and this implies that $z = z'$. Therefore $\boldsymbol{t}$ has at most one left special factor of each length and the left special factors of $\boldsymbol{t}$ are exactly the prefixes of $\boldsymbol{u}$. Thus $F(\boldsymbol{u}) \subseteq F(\boldsymbol{t})$; in particular, $\boldsymbol{u}$ is in the closure of the shift-orbit of $\boldsymbol{t}$.

Now suppose that $\boldsymbol{t}$ is not episturmian. Then, by Theorem 28, there exists a word w (possibly empty) and letters a, b, c, and d with $\{a, b\} \cap \{c, d\} = \varnothing$ such that awb and cwd are both factors of $\boldsymbol{t}$. Since $a \neq c$, the word w is a left special factor of $\boldsymbol{t}$, and therefore w is a prefix of $\boldsymbol{u}$.

Let ℓ_1 and ℓ_2 be non-negative integers such that $T^{\ell_1}(\boldsymbol{t})$ begins with awb and $T^{\ell_2}(\boldsymbol{t})$ begins with cwd. Then, for any two acceptable pairs $(a, <_a)$ and $(c, <_c)$, we have

$$a\boldsymbol{u}\ (= awz\cdots) \leq_a T^{\ell_1}(\boldsymbol{t})\ (= awb\cdots), \tag{11}$$

and

$$c\boldsymbol{u}\ (= cwz\cdots) \leq_c T^{\ell_2}(\boldsymbol{t})\ (= cwd\cdots). \tag{12}$$

Inequality (11) implies that $z \leq_a b$, whereas inequality (12) implies that $z \leq_c d$, and moreover $z \leq_c b$ and $z \leq_a d$. These inequalities imply that $z = b = d$, a contradiction.

Hence $\boldsymbol{t}$ is episturmian, and therefore $\boldsymbol{u}$ is episturmian too (since $\boldsymbol{u}$ is in the closure of the shift-orbit of $\boldsymbol{t}$, which consists of all episturmian words with the same set of factors as $\boldsymbol{t}$ – see Remark 15 or [40]). Moreover, $\boldsymbol{u}$ is epistandard since all of its left special factors are prefixes of it. Therefore, for

any letter x in $\mathcal{A}$, $x\boldsymbol{u}$ is episturmian if and only if x occurs infinitely often in the directive word of $\boldsymbol{u}$ (see [51, Theorem 3.17], [39, Theorem 2.6], or [78, Theorem 6]). Hence, for any acceptable pair $(a, <)$, $a\boldsymbol{u} = \inf\{T^k(\boldsymbol{t}), k \geq 0\}$ if and only if the letter a occurs infinitely often in the directive word of $\boldsymbol{u}$.

REMARK 31. An unrelated connection between finite balanced words (i.e., finite Sturmian words) and lexicographic ordering was recently studied by Jenkinson and Zamboni [49], who presented three new characterizations of "cyclically" balanced finite words via orderings. Their characterizations are based on the ordering of shift-orbits, either lexicographically or with respect to the 1-*norm* $|\cdot|_1$, which counts the number of occurrences of the symbol 1 in a given finite word over $\{0,1\}$.

5.2 STURMIAN MORPHISMS

Prior to the recent work of Pirillo and others, the extremal property (1) was shown to hold for characteristic Sturmian sequences in a paper by Berstel and Séébold [19]. Here is a reformulation of their result (recalling the definition of $s_{\alpha,\rho}$ from Section 3.2, and letting $\boldsymbol{c}_\alpha := s_{\alpha,\alpha} = s'_{\alpha,\alpha}$ denote the unique characteristic Sturmian sequence of slope α):

PROPOSITION 32 ([19, Property 7]). *Let $\alpha > 0$ be an irrational number. Then, for all $i \geq 1$, we have*

$$a\boldsymbol{c}_\alpha < T^i(a\boldsymbol{c}_\alpha) \quad \textit{and} \quad b\boldsymbol{c}_\alpha > T^i(b\boldsymbol{c}_\alpha).$$

In particular, for all $i \geq 0$, we have

$$a\boldsymbol{c}_\alpha < T^i(\boldsymbol{c}_\alpha) < b\boldsymbol{c}_\alpha.$$

REMARK 33. Recall from Remark 15 that the closure of the shift-orbit of any Sturmian word $\boldsymbol{s}$ is a minimal dynamical system consisting of all the Sturmian words with the same set of factors as $\boldsymbol{s}$ (see also [66, Proposition 2.1.25]). In particular, if $\boldsymbol{s}$ is a Sturmian word with (irrational) slope α, then $\overline{\mathcal{O}}(\boldsymbol{s})$ consists of all Sturmian words of slope α (e.g., see [66, Propositions 2.1.18] or [68]). Accordingly, the second part of Proposition 32 (see also Theorems 1 and 2) tells us that $a\boldsymbol{c}_\alpha$ and $b\boldsymbol{c}_\alpha$ are the lexicographically least and greatest Sturmian words of slope α, respectively.

Proposition 32 was also proved by Borel and Laubie [20] in the same year (1993). In [19], Berstel and Séébold showed that it is an easy consequence of the following more general result.

PROPOSITION 34. *Let $\alpha > 0$ be an irrational number and let ρ, ρ' be real numbers such that $0 \leq \rho,\ \rho' < 1$. Then*

$$s_{\alpha,\rho} < s_{\alpha,\rho'} \iff \rho < \rho'.$$

The above proposition was one of numerous results in [19] leading to the proof of a now well-known characterization of *Sturmian morphisms*, i.e., morphisms that preserve Sturmian words. Specifically, a morphism on $\{a,b\}$ is *Sturmian* if and only if it can be expressed as a finite composition of the following morphisms, in any number and order:

$$E\colon \begin{array}{ccc} a & \mapsto & b \\ b & \mapsto & a \end{array}, \qquad \varphi\colon \begin{array}{ccc} a & \mapsto & ab \\ b & \mapsto & a \end{array}, \qquad \widetilde{\varphi}\colon \begin{array}{ccc} a & \mapsto & ba \\ b & \mapsto & a \end{array}.$$

(Note that $\varphi = \psi_a \theta_{ab}$ and $\widetilde{\varphi} = \bar{\psi}_a \theta_{ab}$; see Section 3.3.)

This result played a particularly important role in Berstel and Séébold's characterization of morphisms that preserve characteristic Sturmian words – the so-called *characteristic* or *standard (Sturmian) morphisms*. That is, a morphism on $\{a,b\}$ is *standard* if and only if it is expressible as a finite composition of the morphisms E and φ in any number and order [19]. The fact that there is no occurrence of the morphism $\widetilde{\varphi}$ in such a composition is due to Proposition 32.

5.3 THE LEXICOGRAPHIC WORLD

As mentioned previously, a disguised form of Theorem 2 (see also (1)) appeared in S. Gan's paper [36]; in fact, as we shall see, Theorem 1 can be deduced from the main results in [36]. Gan came across this property of Sturmian sequences whilst endeavouring to obtain a complete description of the *lexicographic world*, defined as follows.

For any two infinite words $\boldsymbol{x},\ \boldsymbol{y} \in \{0,1\}^\omega$, define the set

$$\Sigma_{\boldsymbol{xy}} := \{\boldsymbol{s} \in \{0,1\}^\omega,\ \forall i \geq 0,\ \boldsymbol{x} \leq T^i(\boldsymbol{s}) \leq \boldsymbol{y}\}.$$

The *lexicographic world* $\mathcal{L}$ is defined by

$$\mathcal{L} := \{(\mathbf{x},\mathbf{y}) \in \{0,1\}^\omega \times \{0,1\}^\omega,\ \Sigma_{\mathbf{xy}} \neq \varnothing\}.$$

Gan proved in [36, Lemma 2.1] that

$$\mathcal{L} = \{(\boldsymbol{u},\boldsymbol{v}) \in \{0,1\}^\omega \times \{0,1\}^\omega,\ \boldsymbol{v} \geq \phi(\boldsymbol{u})\},$$

where $\phi\colon \{0,1\}^\omega \to \{0,1\}^\omega$ is the map defined by

$$\phi(\boldsymbol{x}) := \inf\{\boldsymbol{y} \in \{0,1\}^\omega,\ \Sigma_{\boldsymbol{xy}} \neq \varnothing\}.$$

As Gan points out in that paper, the set $\mathcal{L}$ is closely related to the bifurcation of a Lorenz-like map (see [64] for example).

The following theorem combines Corollary 5.6 and Theorem 5.7 from Gan's paper [36] (see also Theorem 1.1 in the same paper). It shows in particular that any element in the image of ϕ is a Sturmian or periodic balanced sequence in $\{0,1\}^\omega$ (and such sequences are the lexicographically greatest amongst their shifts).

THEOREM 35. *For any sequence $\boldsymbol{s} \in \{0,1\}^\omega$, the following conditions are equivalent.*

(i) $\boldsymbol{s} = \phi(\boldsymbol{x})$ *for some sequence* $\boldsymbol{x} \in \{0,1\}^\omega$.

(ii) $\boldsymbol{s}$ *is a Sturmian or periodic balanced sequence satisfying* $T^i(\boldsymbol{s}) \leq \boldsymbol{s}$ *for all* $i \geq 0$. *Moreover, if* $\boldsymbol{x}$ *begins with* 1, *then* $\phi(\boldsymbol{x}) = 1^\omega$, *and if* $\boldsymbol{x} = 0\boldsymbol{u}$ *for some* $\boldsymbol{u} \in \{0,1\}^\omega$, *then* $\phi(\boldsymbol{x})$ *is the unique Sturmian or periodic balanced sequence* $\boldsymbol{s}$ *in* $\{0,1\}^\omega$ *satisfying* $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ *and* $T^i(\boldsymbol{s}) \leq \boldsymbol{s}$ *for all* $i \geq 0$.

In the process of establishing Theorem 35, Gan also proved the following description of *Sturmian minimal sets* (see [44] for a definition; also note that minimal sets correspond to minimal dynamical systems).

THEOREM 36 ([36]). *A minimal set M is a Sturmian minimal set if and only if $M \subseteq [0\boldsymbol{x}, 1\boldsymbol{x}] := \{\boldsymbol{y} \in \{0,1\}^\omega, 0\boldsymbol{x} \leq \boldsymbol{y} \leq 1\boldsymbol{x}\}$ for some $\boldsymbol{x} \in \{0,1\}^\omega$. Moreover, for any $\boldsymbol{x} \in \{0,1\}^\omega$, there exists a unique Sturmian minimal set in $[0\boldsymbol{x}, 1\boldsymbol{x}]$.*

Theorem 36 actually encompasses the first part of Theorem 1; indeed, it can be interpreted as follows: a uniformly recurrent sequence $\boldsymbol{y} \in \{0,1\}^\omega$ satisfies $0\boldsymbol{x} \leq T^i(\boldsymbol{y}) \leq 1\boldsymbol{x}$ for all $i \geq 0$ and some binary sequence $\boldsymbol{x}$ if and only if $\boldsymbol{y}$ is a Sturmian or periodic balanced sequence. As discussed in Section 5.1, this result was recently rediscovered by Glen, Justin, and Pirillo [41] (see (4)), but in a slightly stronger form without the uniform recurrence condition, giving that $\boldsymbol{y}$ is either a Sturmian sequence, a periodic balanced sequence, or a skew sequence (i.e., $\boldsymbol{y}$ is a balanced sequence).

The second part of Theorem 1 can also be deduced from Gan's work, as follows. Let $\boldsymbol{u}$ be any characteristic Sturmian sequence on $\{0,1\}$. Then, by Theorem 35, the sequence $\boldsymbol{s} := \phi(0\boldsymbol{u})$ is the unique Sturmian sequence satisfying $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ and $T^i(\boldsymbol{s}) \leq \boldsymbol{s}$ for all $i \geq 0$. Suppose $\boldsymbol{x}$ is the

unique characteristic Sturmian sequence in $\overline{\mathcal{O}}(\boldsymbol{s})$, the closure of the shift-orbit of $\boldsymbol{s}$. Then $0\boldsymbol{x}$ and $1\boldsymbol{x}$ are Sturmian sequences, by [66, Proposition 2.1.22]. Moreover, $0\boldsymbol{x}$ and $1\boldsymbol{x}$ have the same set of factors as $\boldsymbol{x}$ since the prefixes of $\boldsymbol{x}$ are exactly its left special factors. Hence, both $0\boldsymbol{x}$ and $1\boldsymbol{x}$ are in $\overline{\mathcal{O}}(\boldsymbol{s})$, and therefore, since $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for all $i \geq 0$, we have $0\boldsymbol{u} \leq 0\boldsymbol{x}$ and $1\boldsymbol{x} \leq 1\boldsymbol{u}$. These inequalities imply that $\boldsymbol{u} = \boldsymbol{x}$. Thus, for any characteristic Sturmian sequence $\boldsymbol{x}$, we have $0\boldsymbol{x} < T^i(\boldsymbol{x}) < 1\boldsymbol{x}$ for all $i \geq 0$. This establishes the forward direction of Theorem 2, and it follows that for any Sturmian sequence $\boldsymbol{s}$ on $\{0,1\}$, we have $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for all $i \geq 0$, where $\boldsymbol{u}$ is the unique characteristic Sturmian sequence with the same slope as $\boldsymbol{s}$ (recall Remark 33). This proves the second part of Theorem 1 and from this theorem one can easily deduce both directions of Theorem 2 (see Remark 27).

REMARK 37. By Remark 33, the lexicographically greatest and least Sturmian sequences in the closure of the shift-orbit of a Sturmian sequence $\boldsymbol{s}$ on $\{0,1\}$ are $0\boldsymbol{u}$ and $1\boldsymbol{u}$, where $\boldsymbol{u}$ is the unique characteristic Sturmian sequence with the same slope as $\boldsymbol{s}$. We thus deduce from Theorems 1 and 35 that, for any sequence $\boldsymbol{x}$ on $\{0,1\}$ beginning with 0, the sequence $\phi(\boldsymbol{x})$ is a Sturmian or periodic balanced sequence of the form $1\boldsymbol{u}$. Moreover, if $\phi(\boldsymbol{x})$ is Sturmian, then $\boldsymbol{u}$ is the unique characteristic Sturmian sequence with the same slope as $\phi(\boldsymbol{x})$.

The following lemma was a key step in Gan's proof of Theorem 36. It involves the *block condition* (BC): a sequence $\boldsymbol{s} \in \{0,1\}^\omega$ *satisfies the BC* if, for any finite word w on $\{0,1\}$, at least one of the words $0w0$ and $1w1$ is not a factor of $\boldsymbol{s}$.

LEMMA 38 ([36, Lemma 4.4]). *A sequence $\boldsymbol{s} \in \{0,1\}^\omega$ satisfies the BC if and only if there exists a sequence $\boldsymbol{u}$ such that $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for all $i \geq 0$.*

This result is essentially the characterization of balanced infinite words given in [41] (see (4)). Indeed, the BC is equivalent to the balance property, as defined in Definition 7. See Section 3 in [25], in which the balance property is called the *Sturmian block condition* (see also [78]). Note that the BC of Coven and Hedlund [25, Lemma 3.06, p. 143] is (seemingly) stronger than Gan's in that "for any finite word w" is replaced by "for any palindrome w"; actually both BC conditions are equivalent to the balance property.

REMARK 39. As explained by Labarca and Moreira in [61], the terminology "lexicographical world" was coined in 2000, in a preprint version of [63] (which appeared only in 2006) in which the authors extended the work of Hubbard and Sparrow [46]. For more on the lexicographic(al) world, the reader can look at, e.g., [62, 63] and the references therein. See also the recent paper [8], in which the present two authors give a complete description of the lexicographic world in the process of describing the minimal intervals containing all fractional parts $\{\xi 2^n\}$, for some positive real number ξ, and for all $n \geq 0$.

5.4 THE EARLY WORK OF VEERMAN: 1986 & 1987

Let $\mathcal{S}^\alpha$ denote the set of all Sturmian sequences of (irrational) slope $\alpha > 0$ over the alphabet $\{0,1\}$ (i.e., $a \mapsto 0$, $b \mapsto 1$ in Theorem 4). As noted, e.g., in [16], each Sturmian sequence $\boldsymbol{s} \in \mathcal{S}^\alpha$ can be viewed as the binary expansion of some real number $r(\boldsymbol{s})$ modulo 1. Moreover, it is easily verified that, for any $\boldsymbol{s}$, $\boldsymbol{s}' \in \mathcal{S}^\alpha$, we have $\boldsymbol{s} < \boldsymbol{s}'$ if and only if $r(\boldsymbol{s}) < r(\boldsymbol{s}')$. Furthermore, by Remark 33, we know that the lexicographically least and greatest sequences in $\mathcal{S}^\alpha$ are $0\boldsymbol{c}_\alpha$ and $1\boldsymbol{c}_\alpha$, respectively. In terms of binary expansions, as $r(1\boldsymbol{c}_\alpha) = 1/2 + r(0\boldsymbol{c}_\alpha)$, it follows that the set $r(\mathcal{S}^\alpha) := \{r(\boldsymbol{s}) \in [0,1),\ \boldsymbol{s} \in \mathcal{S}^\alpha\}$ is completely contained within the closed interval $[r(0\boldsymbol{c}_\alpha), r(1\boldsymbol{c}_\alpha)]$ of length $1/2$ and not in any smaller interval.

This latter result (to compare with Bugeaud-Dubickas' result where base 2 is replaced with base b [22]) is essentially a reformulation of Theorem 2, p. 558 in Veerman's paper [85], which also states that $r(\mathcal{S}^\alpha)$ is a Cantor set of Lebesgue measure zero. The converse of this theorem was proved one year later by Veerman in [86, Theorem 2.1, p. 193–194]. As such, it seems that Veerman was the first to (implicitly) prove the Sturmian extremal property given in Theorem 1, under the framework of symbolic dynamics.

Actually, Veerman's main result in [86] shows that a sequence $\boldsymbol{s}$ in $\{0,1\}^\omega$ satisfies the inequalities $0\boldsymbol{u} \leq T^i(\boldsymbol{s}) \leq 1\boldsymbol{u}$ for some sequence $\boldsymbol{u} \in \{0,1\}^\omega$ and for all $i \geq 0$ if and only if $\boldsymbol{s}$ is a Sturmian sequence or a periodic balanced sequence (cf. (4)). A few years earlier (in 1984), Gambaudo et al. [35] had already proved the periodic case (i.e., the case when α is rational); Veerman considered his Theorem 2.1 in [86] to be a generalization of their main result.

REMARK 40. Note that the set $r(\mathcal{S}^\alpha)$ is a dynamical system under the operation of the *doubling map* $\sigma: x \mapsto 2x \pmod 1$ on the one-dimensional torus $\mathbf{T} = \mathbf{R}/\mathbf{Z}$. This was the point of view of Veerman and also that of Bullet and Sentenac [23], who gave reformulations and self-contained combinatorial proofs of some of Veerman's results in [85, 86]. In particular, Bullett and Sentenac gave another proof of the following result (which can be deduced from Veerman's work): for each closed interval $C_\mu = [\mu, 1/2 + \mu]$ of length $1/2$ (where $\mu \in \mathbf{T}$), there exists a unique α such that $r(\mathcal{S}^\alpha)$ is contained in C_μ and there is no other dynamical system for the doubling map that is a strict subset of C_μ. This fact was recently used by Jenkinson [47] to prove new characterizations of *Sturmian measures*, which have applications to ergodic optimization of convex functions. Another important application is in the combinatorial description of the Mandelbrot set (e.g., see [23, 57]).

REMARK 41. In the study of kneading sequences of *Lorenz maps* (i.e., a certain class of piece-wise monotonic maps on $[0,1]$ with a single discontinuity), Glendinning, Hubbard, and Sparrow [43, 46] have investigated so-called *allowed pairs* $(\boldsymbol{r}, \boldsymbol{s})$ of distinct binary sequences in $\{0,1\}^\omega$ satisfying

$$\boldsymbol{r} \le T^i(\boldsymbol{r}) < \boldsymbol{s} \quad \text{and} \quad \boldsymbol{r} < T^i(\boldsymbol{s}) \le \boldsymbol{s} \quad \text{for all } i \ge 0\,.$$

In particular, it was shown in [46] that these allowed pairs are exactly the pairs of (distinct) binary sequences in $\{0,1\}^\omega$ that are realizable as kneading invariants of a topologically expansive Lorenz map. (Note that the case $\boldsymbol{s} = 1^\omega$ was studied by Acquier, Cosnard, and Masse in [1].) Moreover it can be deduced from property (1) that the allowed pairs of the form $(0\boldsymbol{u}, 1\boldsymbol{u})$ are those where $\boldsymbol{u}$ is a characteristic Sturmian sequence.

6. BACK TO DISTRIBUTION MODULO 1 : THE THUE-MORSE SEQUENCE SHOWS UP

As indicated in the Introduction, we began writing this survey after the publication of the paper of Bugeaud and Dubickas [22], whose starting point goes back to a paper of Mahler [67]. In that paper Mahler defines the set of Z-numbers

$$\left\{\xi \in \mathbf{R}\,,\ \xi > 0\,,\ \forall n \ge 0\,,\ 0 \le \left\{\xi \left(\frac{3}{2}\right)^n\right\} < \frac{1}{2}\right\},$$

where $\{x\}$ is the fractional part of the real number x. Mahler proved that this set is at most countable. It is still an open problem to prove that this set is actually empty. More generally, given a real number $\alpha > 1$ and an interval $(s,t) \subset (0,1)$ one can ask whether there exists $\xi > 0$ such that, for all $n \geq 0$, we have $s \leq \{\xi\alpha^n\} < t$. Flatto, Lagarias, and Pollington [34, Theorem 1.4] proved that, if $\alpha = p/q$ with p,q coprime integers and $p > q \geq 2$, then any interval (s,t) such that for some $\xi > 0$, one has that $\{\xi(p/q)^n\} \in (s,t)$ for all $n \geq 0$, must satisfy $t - s \geq 1/p$. The main result in [22] reads as follows.

THEOREM 42 (Bugeaud-Dubickas). *Let $b \geq 2$ be an integer and let ξ be an irrational number. Then the numbers $\{\xi b^n\}$ cannot all lie in an interval of length $< 1/b$. Furthermore there exists a closed interval I of length $1/b$ containing the numbers $\{\xi b^n\}$ for all $n \geq 0$ if and only if the sequence of base b-digits of the fractional part of ξ is a Sturmian sequence $\mathbf{s}$ on the alphabet $\{k, k+1\}$ for some $k \in \{0,1,\ldots,b-2\}$. If this is the case, then ξ is transcendental, and the interval I is semi-open. It is open unless there exists an integer $j \geq 1$ such that $T^j(\mathbf{s})$ is a characteristic Sturmian sequence on the alphabet $\{k, k+1\}$.*

The reader will easily see the relation between Theorem 42 and Theorems 1 and 2. Note that the first assertion in Theorem 42 is generalized to algebraic real numbers > 1 by Dubickas in [29]. Also note that two other papers by Dubickas [30, 31] deal with links between distribution of $\{\xi\alpha^n\}$ modulo 1 and combinatorics on words. Furthermore the *Thue-Morse sequence*, defined as the fixed point beginning with 0 of the morphism $0 \to 01$, $1 \to 10$, shows up in these two papers: in [30] for the study of "small" and "large" limit points of $\|\xi(p/q)^n\|$, the distance to the nearest integer of the product of any non-zero real number ξ by the powers of a rational; in [31] for the study of the "small" and "large" limit points of the sequence of fractional parts $\{\xi b^n\}$, where $b < -1$ is a negative rational number and ξ is a real number. For work in a similar vein and with an avatar of the Thue-Morse sequence, see [55].

Interestingly enough, the Thue-Morse sequence also appeared in 1983 in another question of distribution, as a by-product of the combinatorial study of a set of sequences related to iterating continuous maps of the unit interval (see [4, 6]).

THEOREM 43. *Define the set* $\widetilde{\Gamma}$ *by*

$$\widetilde{\Gamma} := \{x \in [0,1]\,,\ \forall k \geq 0\,,\ 1-x \leq \{2^k x\} \leq x\}\,.$$

Then the smallest limit point of $\widetilde{\Gamma}$ *is the number* $\alpha := \sum a_n/2^n$, *where* $(a_n)_{n\geq 0}$ *is the Thue-Morse sequence. The set* $\widetilde{\Gamma}$ *contains only countably many elements less than* α *and they are all rational. Furthermore any segment on the right of* α *contains uncountably many elements of* $\widetilde{\Gamma}$. *This structure around* α *repeats at infinitely many scales:* $\widetilde{\Gamma}$ *is a fractal set.*

The reader will have guessed that Theorem 43 above is a by-product of the combinatorial study of the set

$$\Gamma := \{\boldsymbol{u} \in \{0,1\}^{\mathbf{N}},\ \forall k \geq 0,\ \overline{\boldsymbol{u}} \leq T^k(\boldsymbol{u}) \leq \boldsymbol{u}\}\,, \tag{13}$$

where $\overline{\boldsymbol{u}}$ is the sequence obtained by switching 0's and 1's in $\boldsymbol{u}$ (see [4]).

An avatar of the set Γ (where large inequalities are replaced by strict inequalities) was studied in [33] in the description of *univoque numbers*, i.e., real numbers β in $(1,2)$ such that there exists a unique base β-expansion of 1 as $1 = \sum_{j\geq 1} u_j \beta^{-j}$, with $u_j \in \{0,1\}$. See [7] for more details.

In [5] the first author uses Theorem 1 to prove that a Sturmian sequence $\boldsymbol{s}$ on $\{0,1\}$ belongs to the set Γ (see (13)) if and only if there exists a characteristic Sturmian sequence $\boldsymbol{u}$ beginning with 1 such that $\boldsymbol{s} = 1\boldsymbol{u}$. (In particular, a Sturmian sequence belonging to Γ must begin with 11.) As an immediate corollary we have that a real number $\beta \in (1,2)$ is univoque and *self-Sturmian* (i.e., the greedy β-expansion of 1 is a Sturmian sequence) if and only if the β-expansion of 1 is of the form $1\boldsymbol{u}$, where $\boldsymbol{u}$ is a characteristic Sturmian sequence beginning with 1. Self-Sturmian numbers were introduced in [24], where it was proved that such numbers are transcendental (see also [60] for more on related questions). Theorem 2 was used in [24] and a proof of Theorem 1 was also given in a preprint version of that paper (see http://arxiv.org/abs/math/0308140); it was deleted from the final version, as D. Y. Kwon explained to J.-P. Allouche: first because a referee suggested it was "folklore", and second because actually only one direction of Theorem 2 was needed. Self-sturmian numbers have since been generalized to *self-episturmian numbers* in [39], where an analogue of Theorem 1 for episturmian sequences can also be found (see Theorem 30).

Also note that sets related to the set Γ and to the lexicographic world occur in the study of badly approximable numbers in [72].

We end this section with a last remark which, while pointing to a new statement, might lead number-theorists to a yet-to-be explored field.

REMARK 44. It is tempting to try to convert the extremal property for episturmian sequences given in Corollary 26 (see [41]) to a result in distribution modulo 1. From now on, $<$ will denote the "usual" order on $D := \{0, 1, \ldots, d-1\}$; other orders will be denoted by $\prec$. As we have seen, an infinite word $\boldsymbol{t}$ on D is episturmian in the wide sense (i.e., episturmian or episkew) if and only if there exists an infinite word $\boldsymbol{u}$ such that

$$(*) \qquad a\boldsymbol{u} \preceq \min(\boldsymbol{t})$$

for any acceptable pair $(a, \prec)$. Actually, replacing the "usual" order on D by another total order is the same as keeping the order but replacing each j in this set by $\sigma(j)$, where σ is a permutation of D. More precisely, $(a, \prec)$ is an acceptable pair if and only if there exists a permutation $\sigma = \sigma_{\preceq}$ of D such that $\sigma(a) = 0$ and $i \preceq j \Leftrightarrow \sigma(i) \leq \sigma(j)$. Hence, another way of formulating $(*)$ above is as follows: there exists an infinite word $\boldsymbol{u}$ such that for all permutations σ of D one has

$$0\sigma(\boldsymbol{u}) \leq \min(\sigma(\boldsymbol{t}))\,,$$

where $\sigma(u_0u_1u_2\ldots) := \sigma(u_0)\sigma(u_1)\sigma(u_2)\ldots$ (for finite or infinite words on D). Hence translating extremal properties of episturmian sequences to properties of distribution modulo 1 for real numbers consists of looking at reals x in $(0,1)$ such that there exists a real y in $(0,1)$ with $\frac{1}{d}y_\sigma \leq \{d^k x_\sigma\}$ for all integers k and for all permutations σ (where x_σ is the real number obtained from x by applying the permutation σ digitwise). If $d = 2$, permuting 0's and 1's in a real number x written in base 2 is the same as replacing x by $1-x$. Hence, in that case, the inequalities $\frac{1}{2}y_\sigma \leq \{2^k x_\sigma\}$ boil down to the two families of inequalities $\frac{1}{2}y \leq \{2^k x\}$ and $\frac{1}{2}(1-y) \leq \{2^k(1-x)\} = 1 - \{2^k x\}$, i.e., $\frac{1}{2}y \leq \{2^k x\} \leq \frac{1}{2} + \frac{1}{2}y$ for all k. This is precisely the question from which we started our paper, but for general d it does not seem that number-theorists have been interested in distribution modulo 1 combined with permuting digits.

7. ADDENDUM

While writing this survey we came across several extra relevant references; other extra references were suggested by the referees. We give them here. About combinatorics of words and Lorenz maps [10, 11, 12, 13, 56, 81], about extremal properties of Sturmian sequences or measures [21, 48, 58, 59], about the distribution of $\{\xi\alpha^n\}$ [2, 3, 32, 87, 88], and last but not least

the historical paper of Lorenz [65] (see also [82]). Finally note that relations between Sturmian sequences and Markoff numbers would need a separate survey, since many results were found since the nice survey [80].

ACKNOWLEDGEMENTS. The authors would like to thank J. Berstel, J. Cassaigne, P. de la Harpe, J. Justin, D. Y. Kwon, G. Pirillo, G. Richomme, P. Séébold, and L. Q. Zamboni, as well as the referees, for discussions, comments, or suggestions.

REFERENCES

[1] ACQUIER, M.-H., M. COSNARD et C. MASSE. Structure de bifurcations des familles à un paramètre de fonctions croissantes par morceaux possédant une seule discontinuité. *C. R. Acad. Sci. Paris Sér. I Math. 300* (1985), 17–22.

[2] AKIYAMA, S. Mahler's Z-number and $3/2$ number systems. *Unif. Distrib. Theory 3* (2008), 91–99.

[3] AKIYAMA, S., C. FROUGNY and J. SAKAROVITCH. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math. 168* (2008), 53–91.

[4] ALLOUCHE, J.-P. Théorie des nombres et automates. Thèse d'État, 1983, Université Bordeaux I. Available at: http://tel.archives-ouvertes.fr/tel-00343206/fr/.

[5] —— A note on univoque self-Sturmian numbers. *Theor. Inform. Appl. 42* (2008), 659–662.

[6] ALLOUCHE, J.-P. et M. COSNARD. Itérations de fonctions unimodales et suites engendrées par automates. *C. R. Acad. Sci. Paris Sér. I Math. 296* (1983), 159–162.

[7] ALLOUCHE, J.-P. and M. COSNARD. Non-integer bases, iteration of continuous real maps, and an arithmetic self-similar set. *Acta Math. Hungar. 91* (2001), 325–332.

[8] ALLOUCHE, J.-P. and A. GLEN. Distribution modulo 1 and the lexicographic world. *Ann. Sci. Math. Québec 33* (2009), 125–143.

[9] ALLOUCHE, J.-P. and J. SHALLIT. *Automatic Sequences. Theory, Applications, Generalizations*. Cambridge University Press, Cambridge, 2003.

[10] ALSEDÀ, LL. and A. FALCÓ. A characterization of the kneading pair for bimodal degree one circle maps. *Ann. Inst. Fourier (Grenoble) 47* (1997), 273–304.

[11] ALSEDÀ, LL. and A. FALCÓ. On the topological dynamics and phase-locking renormalization of Lorenz-like maps. *Ann. Inst. Fourier (Grenoble) 53* (2003), 859–883.

[12] ALSEDÀ, LL. and F. MAÑOSAS. Kneading theory and rotation intervals for a class of circle maps of degree one. *Nonlinearity 3* (1990), 413–452.

[13] ALSEDÀ, LL. and F. MAÑOSAS. Kneading theory for a family of circle maps with one discontinuity. *Acta Math. Univ. Comenian. (N.S.) 65* (1996), 11–22.

[14] ANAGNOSTOPOULOU, V. and O. JENKINSON. Which beta-shifts have a largest invariant measure? *J. London Math. Soc. (2) 79* (2009), 445–464.

[15] ARNOUX, P. et G. RAUZY. Représentation géométrique de suites de complexité $2n + 1$. *Bull. Soc. Math. France 119* (1991), 199–215.

[16] BERSTEL, J. Recent results on extensions of Sturmian words. *Internat. J. Algebra Comput. 12* (2002), 371–385.

[17] —— Sturmian and episturmian words (a survey of some recent results). In: *Algebraic Informatics. Second International Conference, CAI 2007*, Thessaloniki, Greece, 2007, 23–47. Lecture Notes in Computer Science *4728*. Springer, Berlin, 2007.

[18] BERSTEL, J. and P. SÉÉBOLD. A remark on morphic Sturmian words. *Theor. Inform. Appl. 28* (1994), 255–263.

[19] BERSTEL, J. and P. SÉÉBOLD. A characterization of Sturmian morphisms. In: *Mathematical Foundations of Computer Science 1993 (Gdańsk, 1993)*, 281–290. Lecture Notes in Computer Science *711*. Springer, Berlin, 1993.

[20] BOREL, J.-P. et F. LAUBIE. Quelques mots sur la droite projective réelle. *J. Théor. Nombres Bordeaux 5* (1993), 23–51.

[21] BOUSCH, T. Le poisson n'a pas d'arêtes. *Ann. Inst. H. Poincaré Probab. Statist. 36* (2000), 489–508.

[22] BUGEAUD, Y. and A. DUBICKAS. Fractional parts of powers and Sturmian words. *C. R. Math. Acad. Sci. Paris 341* (2005), 69–74.

[23] BULLETT, S. and P. SENTENAC. Ordered orbits of the shift, square roots, and the devil's staircase. *Math. Proc. Cambridge Philos. Soc. 115* (1994), 451–481.

[24] CHI, D. P. and D. Y. KWON. Sturmian words, β-shifts, and transcendence. *Theoret. Comput. Sci. 321* (2004), 395–404.

[25] COVEN, E. M. and G. A. HEDLUND. Sequences with minimal block growth. *Math. Systems Theory 7* (1973), 138–153.

[26] DE LUCA, A. Sturmian words: structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci. 183* (1997), 45–82.

[27] DROUBAY, X., J. JUSTIN and G. PIRILLO. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci. 255* (2001), 539–553.

[28] DROUBAY, X. and G. PIRILLO. Palindromes and Sturmian words. *Theoret. Comput. Sci. 223* (1999), 73–85.

[29] DUBICKAS, A. Arithmetical properties of powers of algebraic numbers. *Bull. London Math. Soc. 38* (2006), 70–80.

[30] —— On the distance from a rational power to the nearest integer. *J. Number Theory 117* (2006), 222–239.

[31] —— On a sequence related to that of Thue-Morse and its applications. *Discrete Math. 307* (2007), 1082–1093.

[32] —— Powers of a rational number modulo 1 cannot lie in a small interval. *Acta Arith. 137* (2009), 233–239.

[33] ERDŐS, P., I. JOÓ and V. KOMORNIK. Characterization of the unique expansions $1 = \sum_{i=1}^{\infty} q^{-n_i}$ and related problems. *Bull. Soc. Math. France 118* (1990), 377–390.

[34] FLATTO, L., J. C. LAGARIAS and A. D. POLLINGTON. On the range of fractional parts $\{\xi(p/q)^n\}$. *Acta Arith. 70* (1995), 125–147.

[35] GAMBAUDO, J.-M., O. LANFORD et C. TRESSER. Dynamique symbolique des rotations. *C. R. Acad. Sci. Paris Sér. I Math. 299* (1984), 823–826.

[36] GAN, S. Sturmian sequences and the lexicographic world. *Proc. Amer. Math. Soc. 129* (2001), 1445–1451.

[37] GLEN, A. Powers in a class of $\mathcal{A}$-strict standard episturmian words. In: *5th International Conference on Words, Université du Québec à Montréal*, Publications du LaCIM *36* (2005), 249–263; and *Theoret. Comput. Sci. 380* (2007), 330–354.

[38] —— A characterization of fine words over a finite alphabet. *Theoret. Comput. Sci. 391* (2008), 51–60.

[39] —— Order and quasiperiodicity in episturmian words. In: *Proceedings of the 6th International Conference on Words*, 144–158. Marseille, France, September 17-21, 2007.

[40] GLEN, A. and J. JUSTIN. Episturmian words: A survey. *Theor. Inform. Appl. 43* (2009), 403–442.

[41] GLEN, A., J. JUSTIN and G. PIRILLO. Characterizations of finite and infinite episturmian words via lexicographic orderings. *European J. Combin. 29* (2008), 45–58.

[42] GLEN, A., F. LEVÉ and G. RICHOMME. Directive words of episturmian words: equivalences and normalization. *Theor. Inform. Appl. 43* (2009), 299–319.

[43] GLENDINNING, P. and C. SPARROW. Prime and renormalisable kneading invariants and the dynamics of expanding Lorenz maps. *(Homoclinic chaos (Brussels, 1991)). Physica D 62* (1993), 22–50.

[44] HEDLUND, G. A. Sturmian minimal sets. *Amer. J. Math. 66* (1944), 605–620.

[45] HEINIS, A. and R. TIJDEMAN. Characterisation of asymptotically Sturmian sequences. *Publ. Math. Debrecen 56* (2000), 415–430.

[46] HUBBARD, J. H. and C. T. SPARROW. The classification of topologically expansive Lorenz maps. *Comm. Pure Appl. Math. 43* (1990), 431–443.

[47] JENKINSON, O. Optimization and majorization of invariant measures. *Electron. Res. Announc. Amer. Math. Soc. 13* (2007), 1–12.

[48] —— A partial order on $\times 2$-invariant measures. *Math. Res. Lett. 15* (2008), 893–900.

[49] JENKINSON, O. and L. Q. ZAMBONI. Characterisations of balanced words via orderings. *Theoret. Comput. Sci. 310* (2004), 247–271.

[50] JUSTIN, J. Episturmian morphisms and a Galois theorem on continued fractions. *Theor. Inform. Appl. 39* (2005), 207–215.

[51] JUSTIN, J. and G. PIRILLO. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci. 276* (2002), 281–313.

[52] JUSTIN, J. and G. PIRILLO. On a characteristic property of Arnoux-Rauzy sequences. *Theor. Inform. Appl. 36* (2002), 385–388.

[53] JUSTIN, J. and G. PIRILLO. Episturmian words: shifts, morphisms and numeration systems. *Internat. J. Found. Comput. Sci. 15* (2004), 329–348.

[54] JUSTIN, J. and L. VUILLON. Return words in Sturmian and episturmian words. *Theor. Inform. Appl. 34* (2000), 343–356.

[55] KANEKO, H. Distribution of geometric sequences modulo 1. *Result. Math. 52* (2008), 91–109.

[56] KELLER, G. and M. ST. PIERRE. Topological and measurable dynamics of Lorenz maps. In: *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, 333–361. Springer-Verlag, Berlin, 2001.

[57] KELLER, K. *Invariant Factors, Julia Equivalences and the (Abstract) Mandelbrot Set.* Lecture Notes in Mathematics *1732*. Springer-Verlag, Berlin, 2000.

[58] KIEFFER, J. C. Sturmian minimal systems associated with the iterates of certain functions on an interval. In: *Dynamical Systems (College Park, 1986–87)*, 354–360. Lecture Notes in Mathematics *1342*. Springer, Berlin, 1988.

[59] KRÜGER, T., S. SCHMELING, R. WINKLER and L. Q. ZAMBONI. Dynamics of kneading sequences. Unpublished preprint (1999).

[60] KWON, D. Y. A devil's staircase from rotations and irrationality measures for Liouville numbers. *Math. Proc. Camb. Philos. Soc. 145* (2008), 739–756.

[61] LABARCA, R. and C. MOREIRA. Bifurcation of the essential dynamics of Lorenz maps and applications to Lorenz-like flows: contributions to the study of the expanding case. *Bol. Soc. Brasil. Mat. (N.S.) 32* (2001), 107–144.

[62] LABARCA, R. and C. MOREIRA. Bifurcations of the essential dynamics of Lorenz maps on the real line and the bifurcation scenario for the linear family. *Sci. Ser. A Math. Sci. (N.S.) 7* (2001), 13–29.

[63] LABARCA, R. and C. MOREIRA. Essential dynamics for Lorenz maps on the real line and the lexicographical world. *Ann. Inst. H. Poincaré Anal. Non Linéaire 23* (2006), 683–694.

[64] LABARCA, R. and S. PLAZA. Bifurcation of discontinuous maps of the interval and palindromic numbers. *Bol. Soc. Mat. Mexicana (3) 7* (2001), 99–116.

[65] LORENZ, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci. 20* (1963), 130–141.

[66] LOTHAIRE, M. *Algebraic Combinatorics on Words*. Encyclopedia of Mathematics and its Applications *90*. Cambridge University Press, Cambridge, 2002.

[67] MAHLER, K. An unsolved problem on the powers of $3/2$. *J. Austral. Math. Soc. 8* (1968), 313–321.

[68] MIGNOSI, F. Infinite words with linear subword complexity. *Theoret. Comput. Sci. 65* (1989), 221–242.

[69] MIGNOSI, F. et P. SÉÉBOLD. Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux 5* (1993), 221–233.

[70] MIGNOSI, F. and L. Q. ZAMBONI. On the number of Arnoux-Rauzy words. *Acta Arith. 101* (2002), 121–129.

[71] MORSE, M. and G. A. HEDLUND. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math. 62* (1940), 1–42.

[72] NILSSON, J. *Sur les nombres mal approximables par les nombres q-adiques*. Doctoral Thesis, Lund University, LTH, 2007. Available at: http://tel.archives-ouvertes.fr/tel-00273870/.

[73] PIRILLO, G. Inequalities characterizing standard Sturmian words. *Pure Math. Appl. 14* (2003), 141–144.

[74] —— Inequalities characterizing standard Sturmian and episturmian words. *Theoret. Comput. Sci. 341* (2005), 276–292.

[75] —— Morse and Hedlund's skew Sturmian words revisited. *Ann. Comb. 12* (2008), 115–121.

[76] PYTHEAS FOGG, N. *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lecture Notes in Mathematics *1794*. Springer-Verlag, Berlin, 2002.

[77] RAUZY, G. Mots infinis en arithmétique. In: *Automata on Infinite Words (Le Mont-Dore, 1984)*, 165–171. Lecture Notes in Computer Science *192*. Springer, Berlin, 1985.

[78] RICHOMME, G. A local balance property of episturmian words. In: *Developments in Language Theory*, 371–381. Lecture Notes in Computer Science *4588*. Springer, Berlin, 2007.

[79] RISLEY, R. N. and L. Q. ZAMBONI. A generalization of Sturmian sequences: combinatorial structure and transcendence. *Acta Arith. 95* (2000), 167–184.

[80] SERIES, C. The geometry of Markoff numbers. *Math. Intelligencer 7* (1985), 20–29.

[81] SILVA, L. and J. SOUSA RAMOS. Topological invariants and renormalization of Lorenz maps. *Physica D 162* (2002), 233–243.

[82] SPARROW, C. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*. Applied Mathematical Sciences *41*. Springer-Verlag, New York-Berlin, 1982.

[83] TIJDEMAN, R. On complementary triples of Sturmian bisequences. *Indag. Math. (N.S.) 7* (1996), 419–424.

[84] —— Intertwinings of periodic sequences. *Indag. Math. (N.S.) 9* (1998), 113–122.

[85] VEERMAN, P. Symbolic dynamics and rotation numbers. *Physica A 134* (1986), 543–576.

[86] —— Symbolic dynamics of order-preserving orbits. *Physica D 29* (1987), 191–201.

[87] ZAÏMI, T. An arithmetical property of powers of Salem numbers. *J. Number Theory 120* (2006), 179–191.

[88] —— On integer and fractional parts of powers of Salem numbers. *Arch. Math. (Basel) 87* (2006), 124–128.

(Reçu le 31 juillet 2009)

Jean-Paul Allouche

CNRS, Math., Équipe Combinatoire et Optimisation
Université Pierre et Marie Curie
Case 189
4, place Jussieu
F-75252 Paris Cedex 05
France
e-mail : allouche@math.jussieu.fr

Amy Glen

Department of Mathematics and Statistics
School of Chemical and Mathematical Sciences
Murdoch University
Perth, WA 6150
Australia
e-mail : amy.glen@gmail.com