CHARACTERISATIONS OF BALANCED WORDS VIA ORDERINGS

Oliver Jenkinson and Luca Q. Zamboni

Queen Mary, University of London and University of North Texas

Abstract. Three new characterisations of balanced words are presented. Each of these characterisations is based on the ordering of a shift orbit, either lexicographically or with respect to the norm $|\cdot|_1$ (which counts the number of occurrences of the symbol 1).

1. Statement of Results

Sturmian[1] sequences were first studied by Morse & Hedlund [MoHe] in the second of their fundamental papers on the symbolic coding of geodesic flows. Since then there have been numerous works dedicated to the study of Sturmian sequences and their generalisations from various points of view including combinatorics, number theory, ergodic theory and dynamical systems. Some of these are listed as references, while for a recent survey on the theory of Sturmian sequences we refer the reader to the chapter by J. Berstel and P. Séébold in [BeSe].

In this paper we study certain finite factors of Sturmian sequences which we call *balanced words* (see below for precise definitions). We obtain three new characterisations of balanced words in terms of different orderings on words. An unrelated connection between balanced words and the lexicographic ordering was recently studied by S. Gan [Ga], and later generalised by J. Justin and G. Pirillo [JuPi] to Arnoux-Rauzy words.

Throughout this article, $\mathbb{N}$ will denote the set of non-negative integers. We begin by recalling some basic definitions concerning infinite words (or sequences) $\omega = \omega_0\omega_1\omega_2\ldots \in A^{\mathbb{N}}$ on a finite alphabet A. We say ω is *periodic* if there exists $n \geq 1$ such that $\omega_k = \omega_{k+n}$ for every $k \geq 0$. In this case we say n is a *period* of ω. The *least period* of ω is the smallest such period. We say ω is *eventually periodic* if there exist $n \geq 1$ and $K \geq 0$ such that $\omega_k = \omega_{k+n}$ for all $k \geq K$. We say ω is *aperiodic* if it is not eventually periodic.

A length-m *factor* of ω is a finite subword $\omega_j\omega_{j+1}\ldots\omega_{j+m-1}$, for some $j \in \mathbb{N}$. We say ω is *recurrent* if each factor of ω occurs in ω an infinite number of times. It is easy to see that every recurrent sequence is either aperiodic or periodic. Finally, ω is said to be *balanced* if for each symbol $a \in A$ and all pairs of factors u and v of ω of equal length, we have $||u|_a - |v|_a| \leq 1$, where $|u|_a$ and $|v|_a$ denote the number of occurrences of a in u

Keywords: balanced, Sturmian, lexicographic order

[1]The nomenclature was originally motivated by a connection with differential equations.

and v respectively. Thus a recurrent balanced sequence is either periodic or aperiodic. A recurrent aperiodic balanced sequence in $\{0,1\}^{\mathbb{N}}$ is called a *Sturmian sequence.*

Henceforth we will restrict to the binary alphabet $\{0,1\}$. Define $\{0,1\}^{**}$ to be $\{0,1\}^{\mathbb{N}} \cup \cup_{q=0}^{\infty}\{0,1\}^q$. We now introduce some terminology pertaining to finite words $w \in \{0,1\}^q$. We say w is a *Sturmian word* if w is a factor of a Sturmian sequence. We say w is *balanced* if the periodic sequence $w^\infty = wwww\cdots$ is balanced. Every balanced word is a Sturmian word but not conversely.

Example 1. The word 00101 is balanced, and hence Sturmian.

The word $w = 001010$ is Sturmian but is not balanced; the periodic sequence $w^\infty = www\ldots$ has factors $z = 101$, $z' = 000$, for which $||z|_1 - |z'|_1| = 2$.

The word 0011 is not Sturmian.

Define the *shift* $\sigma : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ by $\sigma(\omega)_i = \omega_{i+1}$. Similarly define $\sigma : \{0,1\}^q \to \{0,1\}^q$ by $\sigma(w_0 \ldots w_{q-1}) = w_1 \ldots w_{q-1}w_0$.

For any real number $\beta \geq 1$, and any length-m word $z = z_0 \ldots z_{m-1} \in \{0,1\}^m$, define

$$|z|_\beta = \sum_{k=0}^{m-1} z_k \beta^{m-k}.$$

We will refer to $|\cdot|_\beta$ as the β-*norm*, though strictly speaking it is not a norm.

Define the order $<_\beta$ on length-m words by

$$z <_\beta z' \quad \text{if and only if} \quad |z|_\beta < |z'|_\beta.$$

Similarly

$$z =_\beta z' \quad \text{if and only if} \quad |z|_\beta = |z'|_\beta.$$

We say $z \leq_\beta z'$ if and only if either $z <_\beta z'$ or $z =_\beta z'$.

The choices $\beta = 1$ and $\beta = 2$ will be particularly important.

Another ordering is the *lexicographic* ordering, denoted by $<_L$. We say $z <_L z'$ if and only if there exists $j = j(z,z') \in \{0,\ldots,m-1\}$ such that $z_k = z'_k$ for all $k = 0,\ldots,j-1$, and $z_j < z'_j$. We say $z =_L z'$ if and only if $z = z'$. We say $z \leq_L z'$ if and only if either $z <_L z'$ or $z =_L z'$.

The proof of the following result is easy, and is left as an exercise.

Lemma 1. *Let $m \in \mathbb{N}$. The orderings $<_\beta$ and $<_L$ on $\{0,1\}^m$ are related by the following properties.*

1. *If $\beta \geq 2$ then the order $<_\beta$ (resp. $\leq_\beta$) is the same as $<_L$ (resp. $\leq_L$).*
2. *If $\beta \geq 2$ then $z =_\beta z'$ if and only if $z = z'$. In particular, this is the case for $\beta = 2$.*
3. *If $1 \leq \beta < 2$ then there exist words $z \neq z'$ such that $|z|_\beta = |z'|_\beta$. In particular, this is the case for $\beta = 1$.*

Henceforth suppose p and q are positive integers, with $p < q$. For simplicity we will suppose p and q to be coprime, although suitable modifications of all our results hold without this assumption.

Definition. Suppose $1 \leq p < q$ are positive integers such that $\gcd(p,q) = 1$. Let $\mathcal{W}_{p,q}$ denote the set of all words $w \in \{0,1\}^q$ with $|w|_1 = p$. If $w \in \mathcal{W}_{p,q}$ then the symbol 1 occurs with frequency $\frac{p}{q}$ in w, so we say that $\frac{p}{q}$ is the *frequency* of the word w.

Since $\gcd(p,q) = 1$ then any element of $\mathcal{W}_{p,q}$ has least period q under the shift map σ. We will write $w \sim w'$ if there exists $0 \leq k \leq q-1$ such that $w' = \sigma^k w$. In this case we say that w, w' are *cyclically conjugate*, or that w, w' are *cyclic shifts* of one another. The equivalence class $\{\sigma^i w : 0 \leq i \leq q-1\}$ of each $w \in \mathcal{W}_{p,q}$ contains exactly q elements. Let

$$\mathbb{W}_{p,q} = \mathcal{W}_{p,q}/\sim$$

denote the corresponding quotient. Elements of $\mathbb{W}_{p,q}$ are called *orbits*. It will usually be convenient to denote an equivalence class in $\mathbb{W}_{p,q}$ by one of its elements w.

Remark. It is a well known fact (cf. [BuSe]; see also Proposition 1 and the Remark following it) that there are precisely q balanced words in $\mathcal{W}_{p,q}$, all of which are in the same orbit. That is, there is a *unique* balanced orbit in each $\mathbb{W}_{p,q}$.

Definition. Suppose $v \in \mathcal{W}_{p,q}$. For $0 \leq j \leq q-1$, a length-$(j+1)$ *factor* of v is a factor of $\omega = v^\infty$ of the form $\omega_i \omega_{i+1} \dots \omega_{i+j}$, where $0 \leq i \leq q-1$. So the collection of length-$(j+1)$ factors of v has cardinality q (i.e. some of the factors might be the same but we count them with multiplicity). We will write

$$v^{(0)}(j) \leq_1 v^{(1)}(j) \leq_1 \dots \leq_1 v^{(q-1)}(j)$$

to denote a $\leq_1$-ordering of this collection. Although such a $\leq_1$-ordering is not unique (see Example 2), the vector

$$N(v,j) = \left(|v^{(0)}(j)|_1, \dots, |v^{(q-1)}(j)|_1\right)$$

and the partial sums

$$S(v,i,j) = \sum_{k=0}^{i} |v^{(k)}(j)|_1$$

are both well-defined (i.e. independent of the choice of $\leq_1$-ordering of the orbit of v).

Example 2. Let $v = 0001011 \in \mathcal{W}_{3,7}$. If $j = 4$ then a 1-norm ordering of the length-$(j+1)$ factors of v is

$$00010 \leq_1 00101 \leq_1 01100 \leq_1 10001 \leq_1 11000 \leq_1 01011 \leq_1 10110\ .$$

This $\leq_1$-ordering is not unique. For example an alternative is

$$00010 \leq_1 11000 \leq_1 01100 \leq_1 00101 \leq_1 10001 \leq_1 10110 \leq_1 01011\ .$$

We have $N(v,4) = (1,2,2,2,2,3,3)$, and the partial sums $S(v,i,4)$ are given by

i :	0	1	2	3	4	5	6
$S(v,i,4)$:	1	3	5	7	9	12	15

Given $1 \le p < q$ with $\gcd(p,q) = 1$, an elementary observation (see Lemma 7) is that for any $0 \le j \le q-1$, the final sum $S(v, q-1, j)$ (i.e. the sum of all entries in $N(v,j)$) is the same for every $v \in \mathcal{W}_{p,q}$. The value of this sum is $(j+1)p$. A natural problem, then, is to consider the possible values of the partial sums $S(v,i,j)$ for $0 \le j < q-1$, as v varies over the set $\mathbb{W}_{p,q}$. This leads us to our first characterisation of balanced words: the unique balanced orbit in $\mathbb{W}_{p,q}$ (we shall see that such an orbit always exists) is precisely the one which maximises every partial sum.

Theorem A. *(Dominance of* 1*-norm partial sums)*
Let w be a balanced word in $\mathcal{W}_{p,q}$ and v a non-balanced word in $\mathcal{W}_{p,q}$. Then

$$S(w,i,j) \ge S(v,i,j)$$

for all $0 \le i, j \le q-1$.
Moreover, there exist $0 \le i_1, j_1 \le q-1$ such that

$$S(w,i_1,j_1) > S(v,i_1,j_1).$$

Example 3. The word $v = 0001011 \in \mathcal{W}_{3,7}$ is not balanced. The vector $N(v,4) = (1,2,2,2,2,3,3)$ and the partial sums $S(v,i,4)$ were computed in Example 2.

The word $w = 0010101 \in \mathcal{W}_{3,7}$ is balanced. A $\le_1$-ordering of the length-5 factors of w is

$$00101 \le_1 01001 \le_1 01010 \le_1 01010 \le_1 10010 \le_1 10100 \le_1 10101\,.$$

Therefore $N(w,4) = (2,2,2,2,2,2,3)$, and we can compare the partial sums of v and w:

i :	0	1	2	3	4	5	6
$S(v,i,4)$:	1	3	5	7	9	12	15
$S(w,i,4)$:	2	4	6	8	10	12	15

We observe that $S(w,i,4) \ge S(v,i,4)$ for all i, and that $S(w,i_1,4) > S(v,i_1,4)$ for various i_1, consistent with Theorem A.

Our second characterisation of balanced words involves a connection between the $\le_1$-ordering and the lexicographic ordering.

Notation. Given an orbit $[w] \in \mathbb{W}_{p,q}$, let

$$w_{(0)} <_L w_{(1)} <_L \ldots <_L w_{(q-1)} \tag{1.1}$$

denote the lexicographic ordering of its elements. Define the *lexicographic array* $A[w]$ of the orbit $[w]$ to be the $q \times q$ matrix whose i^{th} row is $w_{(i)}$. We will index this array by $0 \le i, j \le q-1$, so that $A[w] = (A[w]_{ij})_{i,j=0}^{q-1}$.

For $0 \le i, j \le q-1$, let $w_{(i)}[j]$ denote the length-$(j+1)$ prefix of $w_{(i)}$; so the $w_{(i)}[j]$ are the length-$(j+1)$ factors of w, counted with multiplicity. For each j this induces the following lexicographic ordering:

$$w_{(0)}[j] \le_L w_{(1)}[j] \le_L \ldots \le_L w_{(q-1)}[j]\ . \tag{1.2}$$

Although the inequalities in (1.1) are all strict, since $w_{(0)}, w_{(1)}, \ldots, w_{(q-1)}$ are all distinct, in general the same is not true of (1.2).

Example 4. Consider again the balanced word $w = 0100101 \in \mathcal{W}_{3,7}$. The lexicographic ordering of $[w]$ is

$$0010101 <_L 0100101 <_L 0101001 <_L 0101010 <_L 1001010 <_L 1010010 <_L 1010100,$$

so that the corresponding lexicographic array is

$$A[w] = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

The lexicographic ordering $w_{(0)}[4] \leq_L \ldots \leq_L w_{(6)}[4]$ of the length-5 prefixes in the orbit $[w]$ is

$$00101 \leq_L 01001 \leq_L 01010 \leq_L 01010 \leq_L 10010 \leq_L 10100 \leq_L 10101\,.$$

We observe that this lexicographic ordering is also a $\leq_1$-ordering (coinciding with that of Example 3).

By contrast, let us next consider the non-balanced word $v = 0001011$, whose lexicographic array is

$$A[v] = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The lexicographic ordering $v_{(0)}[4] \leq_L \ldots \leq_L v_{(6)}[4]$ of the length-5 prefixes in the orbit $[v]$ is

$$00010 \leq_L 00101 \leq_L 01011 \leq_L 01100 \leq_L 10001 \leq_L 10110 \leq_L 11000\,.$$

This time the lexicographic ordering of the length-5 prefixes is *not* a $\leq_1$-ordering. For example $01011 <_L 01100$, yet $01100 <_1 01011$.

These observations, namely the coincidence (resp. non-coincidence) of lexicographic orderings and $\leq_1$-orderings for the balanced (resp. non-balanced) word, are manifestations of the following theorem.

Theorem B. *(Compatibility of lexicographic and* 1*-norm orderings)*
Suppose $w \in \{0,1\}^q$. The following are equivalent:

(1) *w is a balanced word,*
(2) *$w_{(i)}[j] \le_1 w_{(i+1)}[j]$ for all $0 \le i \le q-2$ and $0 \le j \le q-1$.*

Our third characterisation of balanced words connects the lexicographic ordering to the *dynamic* ordering of the corresponding orbit. For general words $w \in \mathcal{W}_{p,q}$, the relation between the dynamic ordering $w, \sigma w, \ldots \sigma^{q-1} w$ and the lexicographic ordering $w_{(0)} <_L w_{(1)} <_L \ldots <_L w_{(q-1)}$ is not well understood; indeed we are not aware of any literature on this subject. The general problem is to understand the permutation π_w on the set $\{0, \ldots, q-1\}$ defined by

$$w_{(i)} = \sigma^{\pi_w(i)}(w).$$

We call π_w the *lexidynamic permutation* for the word w.

Unlike lexicographic ordering, the notion of dynamic ordering is not quite well-defined on the orbit space $\mathbb{W}_{p,q}$, due to the ambiguity in choosing equivalence class representatives. However, if w, w' are in the same shift orbit, with $w' = \sigma^k(w)$, say, then

$$w'_{(i)} = w_{(i)} = \sigma^{\pi_w(i)}(w) = \sigma^{\pi_w(i)-k}(w'),$$

where $\pi_w(i) - k$ is understood modulo k, so that the permutations π_w and $\pi_{w'}$ are easily related. For consistency we shall therefore, unless stated otherwise, always choose the word $w_{(0)}$, which we call the *lexicographically minimal representation*, as the representative of its orbit.

Of course for any given word, the determination of the lexidynamic permutation is a finite problem, the obvious algorithm being to generate both the lexicographic and dynamic orderings, then simply compare them. The defect of this approach is twofold: on the one hand it is algorithmically inefficient, while on the other it is unlikely to yield any conceptual insight into the relation between dynamical and lexicographic orderings.

For balanced words, however, we can do much better; in this case the relation between dynamic and lexicographic orderings *is* completely understood. First we need a definition.

Definition. We say a word $w \in \{0,1\}^q$ has the *lexicographic constant shift property* if its lexidynamic permutation π_w is a power of the cyclic permutation $(0, 1, \ldots, q-1)$ on $\{0, \ldots, q-1\}$; that is, if there exists an integer $\mathfrak{m} = \mathfrak{m}(w) \in \{1, \ldots, q-1\}$ such that

$$w_{(i)} = \sigma^{i\mathfrak{m}} w_{(0)} \quad \text{for all } 0 \le i \le q-1.$$

We call $\mathfrak{m} = \mathfrak{m}(w)$ the *lexicographic shift constant* associated to w; if it exists then it is clearly unique, and is relatively prime to q.

Example 5. Consider the balanced word $w = 0010101$. This word is the lexicographically minimal representation of its orbit. From the lexicographic ordering in Example 4 we see that w has the lexicographic constant shift property, with lexicographic shift constant $\mathfrak{m}(w) = 5$.

The non-balanced word $v = 0001011$ is also the lexicographically minimal representation of its orbit. Its lexidynamic permutation π_v is given by

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 4 & 6 & 3 & 5 \end{pmatrix},$$

which is not a power of a cyclic permutation. Hence v does not have the lexicographic constant shift property.

It turns out that the words with the lexicographic constant shift property are precisely the balanced words, and that moreover there is a simple formula for the associated lexicographic shift constant. This is the content of the following theorem.

Theorem C. *(The lexicographic constant shift property)*

A word $w \in \mathcal{W}_{p,q}$ is balanced if and only if it has the lexicographic constant shift property. Moreover the lexicographic shift constant $\mathfrak{m}(w)$ is defined by the congruence $\mathfrak{m}(w)\, p \equiv 1 \pmod{q}$; that is, $\mathfrak{m}(w)$ is the multiplicative inverse of p modulo q.

The organisation of this article is as follows. Section 2 consists of some preparatory material on balanced words. Theorem A is proved in Section 3, while Theorem B is proved in Section 4. Section 5 is devoted to the proof of Theorem C, and to a study of the fine structure of the lexicographic array of a balanced orbit. This structure is described in Theorem D, and readily leads to the proof of Theorem C.

2. Balanced words

This section is preparatory. We begin by introducing notation, then go on to show (Proposition 1 and the Remark following it) that for each pair of positive integers $1 \leq p < q$ with $\gcd(p, q) = 1$, there exists a unique balanced orbit in $\mathbb{W}_{p,q}$. In other words, if $w, w' \in \mathcal{W}_{p,q}$ are balanced, then they are necessarily cyclic permutations of one another. This result is well known (see [BuSe] for example), though our proof appears to be new. We then examine more closely the unique balanced orbit in $\mathbb{W}_{p,q}$. We organise our ideas around a particular word in this orbit, the *post-minimal balanced word* $w_{p,q}$, which plays a central rôle in our later proof of Theorem C.

Notation. Throughout this section, whenever we write a positive rational number in the form $\frac{p}{q}$, we understand that the integers p and q are both positive and $\gcd(p, q) = 1$. For instance, if $\frac{p}{q} = 0 \cdot 2$ then $p = 1$ and $q = 5$.

A rational number $\frac{p}{q} < 1$ has exactly two possible continued fraction expansions. The number of digits in these expansions are j and $j+1$, for some $j \in \mathbb{N}$. To remove this ambiguity we will always choose the continued fraction expansion with an *even* number of digits. More precisely, there exists a unique $k \in \mathbb{N}$ and unique positive integers $n_0, n_1, \ldots, n_{2k+1}$ such that

$$\frac{p}{q} = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{\cdots \cfrac{1}{n_{2k+1}.}}}}}$$

We call this the *even length continued fraction expansion* of $\frac{p}{q}$, and henceforth write it as $\frac{p}{q} = [n_0, n_1, \ldots, n_{2k+1}]$

For any $\frac{p}{q} = [n_0, n_1, \ldots, n_{2k+1}] \neq \frac{1}{2}$ we define its *predecessor* $\frac{p'}{q'}$ by:

$$\frac{p'}{q'} = \begin{cases} [n_0 - 1, n_1, n_2, \ldots, n_{2k+1}] & \text{if } 0 < \frac{p}{q} < \frac{1}{2} \\ [n_0, n_1 - 1, n_2, \ldots, n_{2k+1}] & \text{if } \frac{1}{2} < \frac{p}{q} < 1 \text{ with } n_1 \geq 2 \\ [n_2 + 1, n_3, \ldots, n_{2k+1}] & \text{if } \frac{1}{2} < \frac{p}{q} < 1 \text{ with } n_1 = 1. \end{cases}$$

Note that if $\frac{1}{2} < \frac{p}{q} < 1$ with $n_1 = 1$ then necessarily $2k + 1 \geq 3$.

A straightforward computation gives:

Lemma 2. *If $\frac{p}{q} \neq \frac{1}{2}$ has predecessor $\frac{p'}{q'}$ then*

$$\frac{p}{q} = \begin{cases} \frac{p'}{p'+q'} & \text{if } 0 < \frac{p}{q} < \frac{1}{2} \\ \frac{q'}{2q'-p'} & \text{if } \frac{1}{2} < \frac{p}{q} < 1. \end{cases}$$

Alternatively, we have $\frac{p'}{q'} = f(\frac{p}{q})$ where

$$f(x) = \begin{cases} \frac{x}{1-x} & \text{if } x \in (0, \frac{1}{2}) \\ \frac{2x-1}{x} & \text{if } x \in (\frac{1}{2}, 1) \end{cases}$$

Since by assumption $\gcd(p', q') = 1$, it follows that $\gcd(p', p'+q') = \gcd(q', 2q'-p') = 1$. In other words the above expressions give both p and q as functions of p' and q'. It is easy to see that in each case $q' < q$. Some of the results that follow will be proved by induction on the integer q, and hence the inductive hypothesis will be applied to the integer q'.

Definition. Define the morphisms $\tau_0, \tau_1 : \{0,1\}^{**} \to \{0,1\}^{**}$ by

$$\tau_0 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 01 \end{cases} \qquad \tau_1 : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 1 \end{cases} .$$

That is (cf. [BeSe]), $\tau_0 = \varphi \circ E$ and $\tau_1 = E \circ \varphi$ where E is the exchange morphism $0 \mapsto 1$, $1 \mapsto 0$, and φ is the Fibonacci morphism $0 \mapsto 01$, $1 \mapsto 0$.

Then for any positive integers p, q with $1 \leq p < q$ and $\gcd(p, q) = 1$, we define the *post-minimal balanced word* $w_{p,q} \in \{0,1\}^{**}$ by

$$w_{p,q} = \tau_0^{n_0 - 1} \circ \tau_1^{n_1} \circ \tau_0^{n_2} \circ \cdots \circ \tau_1^{n_{2k+1}}(0),$$

where $\frac{p}{q} = [n_0, n_1, n_2, \ldots, n_{2k+1}]$ is the even length continued fraction expansion.

The reason for calling $w_{p,q}$ *post-minimal* and *balanced* will become apparent later. In Proposition 1 we will show that $w_{p,q}$ belongs to $\mathcal{W}_{p,q}$ and is indeed balanced. In Lemma 6 we will show that $w_{p,q}$ is the shift image of the lexicographically minimal word in the unique balanced orbit in $\mathbb{W}_{p,q}$.

This definition of $w_{p,q}$ immediately gives:

Lemma 3. *If $\frac{p'}{q'}$ is the predecessor of $\frac{p}{q} \neq \frac{1}{2}$ then*

$$w_{p,q} = \begin{cases} \tau_0(w_{p',q'}) & \text{if } 0 < \frac{p}{q} < \frac{1}{2} \\ \tau_1(w_{p',q'}) & \text{if } \frac{1}{2} < \frac{p}{q} < 1 \end{cases} .$$

The first important property of the post-minimal balanced word $w_{p,q}$ is the following.

Proposition 1. *Let p, q be positive integers with $1 \leq p < q$ and $\gcd(p,q) = 1$. Then $w_{p,q}$ is a balanced word in $\mathcal{W}_{p,q}$.*

Proof. It is a basic fact that the morphisms τ_i preserve balance. That is, a $\{0,1\}$-word u is balanced if and only if $\tau_i(u)$ is balanced (see [BeSe], Proposition 2.3.1). Since $w_{p,q}$ is defined by applying a certain composition of these morphisms to the balanced word 0, it follows that $w_{p,q}$ is itself balanced.

It remains to show that $w_{p,q}$ belongs to $\mathcal{W}_{p,q}$. Our proof of this will be by induction on q.

First consider the case $\frac{p}{q} = \frac{1}{2} = [1,1]$. Here we have $w_{1,2} = \tau_1(0) = 10$, which is clearly an element of $\mathcal{W}_{1,2}$, as required.

Next suppose $\frac{p}{q} \neq \frac{1}{2}$, and let $\frac{p'}{q'}$ be the predecessor of $\frac{p}{q}$, where $1 \leq p' < q'$ and $\gcd(p', q') = 1$. Since $q' < q$, we will take our inductive hypothesis to be that $w_{p',q'} \in \mathcal{W}_{p',q'}$, and show that this implies $w_{p,q} \in \mathcal{W}_{p,q}$.

If $0 < \frac{p}{q} < \frac{1}{2}$ then Lemma 3 gives

$$w_{p,q} = \tau_0(w_{p',q'}),$$

so by definition of τ_0, and from Lemma 2, we have

$$|w_{p,q}|_1 = |w_{p',q'}|_1 = p' = p$$

and

$$|w_{p,q}| = |w_{p,q}|_0 + |w_{p,q}|_1 = |w_{p',q'}| + |w_{p',q'}|_1 = q' + p' = q,$$

so indeed $w_{p,q} \in \mathcal{W}_{p,q}$.

If $\frac{1}{2} < \frac{p}{q} < 1$ then Lemma 3 gives

$$w_{p,q} = \tau_1(w_{p',q'}),$$

so by definition of τ_1 we have

$$|w_{p,q}|_1 = |w_{p',q'}| = q' = p$$

and

$$|w_{p,q}| = |w_{p,q}|_0 + |w_{p,q}|_1 = |w_{p',q'}|_0 + |w_{p',q'}| = (q' - p') + q' = 2q' - p' = q,$$

and as before we see that $w_{p,q} \in \mathcal{W}_{p,q}$. $\square$

Remark. Using the morphisms τ_i and induction on q, it is easy to see that if w is a balanced word in $\mathcal{W}_{p,q}$ then w is a cyclic permutation of $w_{p,q}$. Consequently the orbit set $\mathbb{W}_{p,q}$ contains one and only one balanced orbit. This orbit consists of precisely q points, since $\gcd(p,q)=1$.

Example 6. For $\frac{p}{q}=\frac{3}{7}$, the even length continued fraction expansion is $\frac{p}{q}=\frac{3}{7}=[2,3]$. Therefore its predecessor is $\frac{p'}{q'}=[1,3]=\frac{3}{4}$. Hence $w_{3,7}=\tau_0\circ\tau_1^3(0)=\tau_0(1110)=0101010$.

The following lemma gives more information on the structure of the post-minimal balanced word $w_{p,q}$. We first need some definitions.

Definitions. A $\{0,1\}$-word x is called *Sturmian left special* (resp. *Sturmian right special*) if there exists a Sturmian sequence ω such that $0x$ and $1x$ (resp. $x0$ and $x1$) are both factors of ω. We say x is *Sturmian bispecial* if there exists a Sturmian sequence ω such that $0x$, $1x$, $x0$, and $x1$ are each factors of ω.

A Sturmian sequence $\omega\in\{0,1\}^{\mathbb{N}}$ is called *characteristic* if each prefix $\omega_0\omega_1\dots\omega_n$ is left special. For any $\alpha\in[0,1]\setminus\mathbb{Q}$, it is well known that there is a unique characteristic Sturmian sequence of frequency α (where the *frequency* of a Sturmian sequence ω is defined to be $\lim_{n\to\infty}\frac{1}{n}|\omega_0\dots\omega_{n-1}|_1$).

The *mirror image* of a $\{0,1\}$-word $x=x_0\dots x_n$ is the word $\tilde{x}$ defined by $\tilde{x}_i=x_{n-i}$. We say that x is a *palindrome* if $\tilde{x}=x$.

Lemma 4. *Let $1\le p<q$ be positive integers with $\gcd(p,q)=1$. Then*

$$w_{p,q}=x10$$

where x is a Sturmian bispecial word.

In particular the word x is a palindrome.

Proof. It is a basic fact that all Sturmian bispecial words are palindromes, and that if x is a bispecial Sturmian word (or the empty word), then both $\tau_0(x)0$ and $\tau_1(x)1$ are Sturmian bispecial (see [BeSe], Propositions 2.1.19 and 2.1.23). The result will now follow by induction on q.

If $\frac{p}{q}=\frac{1}{2}$, then $w_{1,2}=10$ and hence the result holds with x the empty word.

If $\frac{p}{q}\ne\frac{1}{2}$ then it has a predecessor $\frac{p'}{q'}$. As inductive hypothesis let us suppose we can write $w_{p',q'}=x'10$ with x' either empty or Sturmian bispecial. If $0<\frac{p}{q}<\frac{1}{2}$ then Lemma 3 gives

$$w_{p,q}=\tau_0(w_{p',q'})=\tau_0(x'10)=\tau_0(x')010,$$

and as noted above $\tau_0(x')0$ is Sturmian bispecial. Similarly if $\frac{1}{2}<\frac{p}{q}<1$ then

$$w_{p,q}=\tau_1(w_{p',q'})=\tau_1(x'10)=\tau_1(x')110,$$

and $\tau_1(x')1$ is Sturmian bispecial, as required. $\square$

Example 7. For $\frac{p}{q}=\frac{3}{7}$ we have $w_{3,7}=0101010$ from Example 6, and in this case $x=01010$.

Notation. Given a $\{0,1\}$-word $x=x_0\dots x_n$, where $x_n\in\{0,1\}$, we will write $x_nxx_n^{-1}$ to denote the word $x_nx_0\dots x_{n-1}$.

We will need the following lemma of Pirillo [Pir2] (see also the earlier [Pir1]).

Lemma 5. [Pir2] *Let $x \in \{0,1\}^n$ for some $n \in \mathbb{N}$. Then $0x1$ is cyclically conjugate to $1x0$ if and only if x is Sturmian bispecial.*

We are now ready to justify calling $w_{p,q}$ the post-minimal balanced word: the following result characterises it as the shift image of the lexicographically minimal point in the unique balanced orbit of frequency $\frac{p}{q}$.

Proposition 6. *Let $1 \leq p < q$ be positive integers with $\gcd(p,q) = 1$. Let*

$$w_{(0)} <_L w_{(1)} <_L \cdots <_L w_{(q-1)}$$

be the lexicographic ordering of the unique balanced orbit in $\mathbb{W}_{p,q}$.

If we write $w_{p,q} = x10$, where x is Sturmian bispecial then

$$w_{(0)} = 0x1 \qquad \text{and} \qquad w_{(q-1)} = 1x0.$$

In particular,

$$w_{p,q} = \sigma(w_{(0)}).$$

Moreover, $w_{(0)}$ and $w_{(q-1)}$ are mirror images of one another, i.e., $\widetilde{w_{(0)}} = w_{(q-1)}$.

Proof. First observe that $0x1 = 0w_{p,q}0^{-1} = \sigma^{-1}(w_{p,q})$ is a cyclic shift of $w_{p,q}$ since $w_{p,q}$ ends in 10, by Lemma 4. From Lemma 5 we deduce that $1x0$ is a cyclic shift of $0x1$, and hence of $w_{p,q}$.

We will next show that $0x1 = 0w_{p,q}0^{-1}$ is the lexicographically smallest cyclic shift of $w_{p,q}$. The proof that $1x0$ is the largest cyclic shift of $w_{p,q}$ is almost identical, and is left as an exercise for the reader.

The periodic sequence generated by the word $w_{p,q}$ can be written as the infinite composition

$$w_{p,q}^{\infty} = \lim_{n\to\infty} \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \tau_0^{n_2} \circ \cdots \circ \tau_1^{n_{2k+1}} \circ \tau_0^{n}(1).$$

To prove that $0w_{p,q}0^{-1}$ is the lexicographic smallest cyclic shift of $w_{p,q}$ it suffices to show that

$$0w_{p,q}^{\infty} = (0w_{p,q}0^{-1})^{\infty} \leq_L \sigma^i(0w_{p,q}^{\infty})$$

for each $i \geq 1$. Suppose to the contrary that for some $i \geq 1$ we had $0w_{p,q}^{\infty} >_L \sigma^i(0w_{p,q}^{\infty})$. Then there exists a word u (possibly empty) such that $0u1$ is a prefix of $0w_{p,q}^{\infty}$ and $0u0$ a prefix of $\sigma^i(0w_{p,q}^{\infty})$. It follows that there exists a prefix U of $w_{p,q}^{\infty}$ which begins in $u1$ and contains $0u0$ as a subfactor. It is well known (see Proposition III.7 in [RiZa] for example) that for each n the word $\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \tau_0^{n_2} \circ \cdots \circ \tau_1^{n_{2k+1}} \circ \tau_0^{n}(1)$ is a prefix of a characteristic Sturmian sequence, and therefore each prefix of $w_{p,q}^{\infty}$ is a prefix of some characteristic Sturmian sequence. Hence there exists a characteristic Sturmian sequence ω beginning in U and hence in $u1$. Since each prefix of ω is left special, we deduce that both $1u1$ and $0u0$ are factors of ω, contradicting the fact that ω is balanced. Hence $0w_{p,q}^{\infty}$ is less than or equal to all of its shifts, as required. □

Example 8. For $\frac{p}{q} = \frac{3}{7}$ we have seen that $w_{3,7} = 0101010 = x10$, where $x = 01010$. Now $w_{(0)} = 0010101$, $w_{(q-1)} = 1010100$ (cf. Example 4). So indeed $w_{(0)} = 0x1$, $w_{(q-1)} = 1x0$, and $w_{3,7} = \sigma(w_{(0)})$, consistent with Proposition 6.

3. Proof of Theorem A

We first collect together some obvious facts, which the reader will easily verify.

Lemma 7. *Suppose* $w \in \mathcal{W}_{p,q}$*. Let* $A[w] = (A[w]_{ij})_{i,j=0}^{q-1}$ *be the corresponding lexicographic array. Then*

(1) *The sum of the entries in any row of* $A[w]$ *is* p*,*
(2) *The sum of the entries in any column of* $A[w]$ *is* p*,*
(3) $S(w, q-1, j) = (j+1)p$ *for every* $0 \le j \le q-1$*,*
(4) $\sum_{k=0}^{q-1} |w_{(k)}[j]|_1 = (j+1)p$ *for every* $0 \le j \le q-1$*.*

We are now ready to prove Theorem A. For convenience we first recall its statement.

Theorem A. *(Dominance of* 1*-norm partial sums)*
Let w *be a balanced word in* $\mathcal{W}_{p,q}$ *and* v *a non-balanced word in* $\mathcal{W}_{p,q}$*. Then*

$$S(w, i, j) \ge S(v, i, j)$$

for all $0 \le i, j \le q-1$*.*
Moreover, there exist $0 \le i_1, j_1 \le q-1$ *such that*

$$S(w, i_1, j_1) > S(v, i_1, j_1).$$

Proof of Theorem A. Suppose $w \in \mathcal{W}_{p,q}$ is balanced, and $v \in \mathcal{W}_{p,q}$ is non-balanced. We will first show that $S(w,i,j) \ge S(v,i,j)$ for all $0 \le i, j \le q-1$. If the result is false then there exist $0 \le i_0, j_0 \le q-1$ such that $S(w, i_0, j_0) < S(v, i_0, j_0)$. If $w^{(0)}(j_0) \le_1 \dots \le_1 w^{(q-1)}(j_0)$ is any $\le_1$-ordering of the length-(j_0+1) factors of w, and $v^{(0)}(j_0) \le_1 \dots \le_1 v^{(q-1)}(j_0)$ is any $\le_1$-ordering of the length-(j_0+1) factors of v, this means that

$$\sum_{k=0}^{i_0} |w^{(k)}(j_0)|_1 < \sum_{k=0}^{i_0} |v^{(k)}(j_0)|_1. \tag{3.1}$$

We may suppose that i_0 is chosen as small as possible, in the sense that

$$\sum_{k=0}^{i} |w^{(k)}(j_0)|_1 \ge \sum_{k=0}^{i} |v^{(k)}(j_0)|_1 \quad \text{for all } i < i_0. \tag{3.2}$$

Taking $i = i_0 - 1$ in (3.2), and subtracting from (3.1), we obtain

$$|w^{(i_0)}(j_0)|_1 \le |v^{(i_0)}(j_0)|_1 - 1. \tag{3.3}$$

Now w is balanced, so there exists N_{j_0} such that $|w^{(k)}(j_0)|_1$ equals either N_{j_0} or $N_{j_0}+1$ for all $0 \le k \le q-1$. Thus

$$|w^{(k)}(j_0)|_1 \le |w^{(i_0)}(j_0)|_1 + 1 \quad \text{for all } k > i_0. \tag{3.4}$$

By definition of $\leq_1$ we know that

$$|v^{(i_0)}(j_0)|_1 \leq |v^{(k)}(j_0)|_1 \quad \text{for all } k > i_0. \tag{3.5}$$

Combining (3.3), (3.4) and (3.5) gives

$$|w^{(k)}(j_0)|_1 \leq |v^{(k)}(j_0)|_1 \quad \text{for all } k > i_0.$$

Hence

$$\sum_{k=i_0+1}^{q-1} |w^{(k)}(j_0)|_1 \leq \sum_{k=i_0+1}^{q-1} |v^{(k)}(j_0)|_1. \tag{3.6}$$

Combining (3.1) and (3.6) gives

$$\sum_{k=0}^{q-1} |w^{(k)}(j_0)|_1 < \sum_{k=0}^{q-1} |v^{(k)}(j_0)|_1.$$

This is a contradiction, since by part (3) of Lemma 7 we know that

$$\sum_{k=0}^{q-1} |w^{(k)}(j_0)|_1 = S(w, q-1, j_0) = (j_0+1)p = S(v, q-1, j_0) = \sum_{k=0}^{q-1} |v^{(k)}(j_0)|_1.$$

This contradiction completes the proof of the first part of Theorem A.

The proof of the second part will also be by contradiction. Let us suppose that there do not exist $0 \leq i_1, j_1 \leq q-1$ for which $S(w, i_1, j_1) > S(v, i_1, j_1)$. In view of the first part of the theorem this means that $S(w, i, j) = S(v, i, j)$ for all $0 \leq i, j \leq q-1$. Consequently

$$|w^{(i)}(j)|_1 = |v^{(i)}(j)|_1 \quad \text{for all } 0 \leq i, j \leq q-1 \tag{3.7}$$

where $w^{(0)}(j) \leq_1 \ldots \leq_1 w^{(q-1)}(j)$ and $v^{(0)}(j) \leq_1 \ldots \leq_1 v^{(q-1)}(j)$ are any choices of $\leq_1$-orderings.

However v is not balanced, so we can find some $0 \leq j_0 \leq q-1$, and some $0 \leq i_0 < i_0' \leq q-1$ such that $|v^{(i_0')}(j_0)|_1 \geq v^{(i_0)}(j_0)|_1 + 2$. Now w is balanced, so there exists N_{j_0} such that $|w^{(i)}(j_0)|_1$ equals either N_{j_0} or $N_{j_0}+1$ for every $0 \leq i \leq q-1$. In particular, $|w^{(i_0)}(j_0)|_1$ equals either N_{j_0} or $N_{j_0}+1$, and $|w^{(i_0')}(j_0)|_1$ equals either N_{j_0} or $N_{j_0}+1$. Therefore either $v^{(i_0)}(j_0)|_1 < |w^{(i_0)}(j_0)|_1$ or $|v^{(i_0')}(j_0)|_1 > |w^{(i_0')}(j_0)|_1$. This contradicts (3.7), so we are done. □

4. Proof of Theorem B

The following result gives a very practical way of writing down the lexicographic array associated to a balanced word.

Proposition 2. *Let $[w]$ be the unique balanced orbit in $\mathbb{W}_{p,q}$. Define $u = u_{p,q} \in \mathcal{W}_{p,q}$ by*

$$u = 0 \dots 0 \underbrace{1 \dots 1}_{p}.$$

Then, for $0 \leq i, j \leq q-1$,

(1) $A[w]_{ij} = (\sigma^{jp} u)_i$,
(2) *The j^{th} column of $A[w]$ is (the vector transpose of) the word $\sigma^{jp} u$*
(3) $w_{(i)} = u_i (\sigma^p u)_i (\sigma^{2p} u)_i \dots (\sigma^{(q-1)p} u)_i$.

Example 9. Consider the balanced word $w = 0010101 \in \mathcal{W}_{3,7}$. By definition $u_{3,7} = 0000111$. The lexicographic array $A[w]$ is given in Example 4, and we see that successive columns of $A[w]$ are given by rotating upwards the block of 1's.

Proof of Proposition 2. The conditions (1), (2), (3) are clearly equivalent, so it will suffice to check condition (3).

A well known characterisation of the balanced orbit in $\mathbb{W}_{p,q}$, due to Morse & Hedlund [MoHe], is as the symbolic coding of the rotation by angle p/q on the circle (see also Lemmas 2.1.14 and 2.1.15 of [BeSe]). More precisely, let $T : [0, 1) \to [0, 1)$ be the rational rotation $T(x) = x + p/q \pmod 1$, and let $I = [(q-p)/q, 1)$. For each $0 \leq i \leq q-1$, associate to the rational number i/q the length-q word $c = c(i, q)$ defined by $c_k = \chi_I(T^k(i/q))$ (where χ_I is the characteristic function for the subinterval I). Each $c(i, q)$ is balanced, and the words $c(0, q), \dots, c(q-1, q)$ are all distinct. Moreover, each $c(i, q)$ is lexicographically smaller than $c(i+1, q)$, in fact if $\chi_I(T^k(i/q)) = 1$, then either $\chi_I(T^k((i+1)/q)) = 1$ or $\chi_I(T^{k-1}((i+1)/q)) = 1$. So in fact $c(i, q) = w_{(i)}$ for $0 \leq i \leq q-1$.

The definition of $u = u_{p,q}$ means that $u_i = 1$ if and only if $i/q \in I$, and more generally, $(\sigma^{kp} u)_i = 1$ if and only if $T^k(i/q) \in I$. But we also know that $c(i, q)_k = 1$ if and only if $T^k(i/q) \in I$. Combining these observations gives $c(i, q)_k = (\sigma^{kp} u)_i$ for all $0 \leq k \leq q-1$.

But we know that $c(i, q) = w_{(i)}$, hence $(w_{(i)})_k = (\sigma^{kp} u)_i$, concluding the proof of the proposition. □

We are now ready to prove Theorem B. For convenience we first recall its statement.

Theorem B. *(Compatibility of lexicographic and 1-norm orderings)*
Suppose $w \in \{0, 1\}^q$. The following are equivalent:

(1) *w is a balanced word,*
(2) *$w_{(i)}[j] \leq_1 w_{(i+1)}[j]$ for all $0 \leq i \leq q-2$ and $0 \leq j \leq q-1$.*

Proof of Theorem B. We first prove that (1) $\Rightarrow$ (2). Let $w \in \mathcal{W}_{p,q}$ be balanced. From the definition of $u = u_{p,q}$ we see that, for each $0 \leq i \leq q-2$,

$$\sum_{k=0}^{j} (\sigma^{kp} u)_i \leq \sum_{k=0}^{j} (\sigma^{kp} u)_{i+1} \quad \text{for any } 0 \leq j \leq q-1. \tag{4.1}$$

In fact, if $(\sigma^{kp} u)_i = 1$, then either $(\sigma^{kp} u)_{i+1} = 1$ or $(\sigma^{(k-1)p} u)_{i+1} = 1$. But by part (3) of Proposition 2, we know that $w_{(i)}[j] = u_i (\sigma^p u)_i (\sigma^{2p} u)_i \dots (\sigma^{jp} u)_i$ and $w_{(i+1)}[j] = u_{i+1} (\sigma^p u)_{i+1} (\sigma^{2p} u)_{i+1} \dots (\sigma^{jp} u)_{i+1}$.

Thus the lefthand side of (4.1) is precisely $|w_{(i)}[j]|_1$, and the righthand side of (4.1) is precisely $|w_{(i+1)}[j]|_1$. So (4.1) asserts that $w_{(i)}[j] \leq_1 w_{(i+1)}[j]$, as required.

Now we prove that (2) $\Rightarrow$ (1). Suppose w is not balanced. We would like to show there exist equal length subwords v, v' of w such that $v <_L v'$ yet $v' <_1 v$. For convenience we will say that a word is *allowed* (resp. *disallowed*) if it is a subword (resp. not a subword) of the period-q sequence w^∞.

A well known result (see Proposition 2.1.3 in [BeSe]) guarantees the existence of a subword $a =: a^{[1]}$ of w such that both $0a0$ and $1a1$ are allowed.

Now *if* there exists some allowed word c such that both $0c1a1$ and $1c0a0$ are allowed, then we can set $v = 0c1a1$ and $v' = 1c0a0$ and we are done, since $v <_L v'$ yet $v' <_1 v$.

So let us suppose, for a contradiction, that there does *not* exist c such that both $0c1a1$ and $1c0a0$ are allowed.

Let b be the longest subword of w such that both $b0a0$ and $b1a1$ are allowed. Our assumption above means, however, that either $1b0a0$ or $0b1a1$ (or possibly both) are disallowed.

We claim that indeed $1b0a0$ or $0b1a1$ are *both* disallowed, and hence that $0b0a0$ and $1b1a1$ are both allowed.

To prove this claim, first suppose that $0b1a1$ is disallowed. Then clearly $1b1a1$ must be allowed, since $b1a1$ must have a prefix symbol, and our alphabet is binary.

But now this implies that $1b0a0$ is also disallowed, since if it were allowed then $1b$ would be a common prefix to both $1a1$ and $0a0$, contradicting the choice of b as the *longest* such common prefix. Hence $0b0a0$ must be allowed.

An analogous argument shows that if $1b0a0$ is disallowed then necessarily $0b1a1$ is also disallowed, and hence that $0b0a0$ and $1b1a1$ are both allowed. Therefore the claim is proved.

But now we have proved the existence of a word $b =: a^{[2]}$ such that both $0b0$ and $1b1$ are allowed, and moreover such that $0b0a0$ and $1b1a1$ are both allowed. That is, both $0a^{[2]}0a^{[1]}0$ and $1a^{[2]}1a^{[1]}1$ are allowed.

We can now repeat our argument. That is, we suppose, for a contradiction, that there does *not* exist c such that both $0c1b1a1$ and $1c0b0a0$ are allowed.

In this way, we show that if the (2) $\Rightarrow$ (1) part of the theorem is *false* then for any $r \in \mathbb{N}$ we can find subwords $a^{[1]}, \ldots, a^{[r]}$ of w such that both $v_0^{[r]} := 0a^{[r]}0a^{[r-1]}0 \ldots 0a^{[2]}0a^{[1]}0$ and $v_1^{[r]} := 1a^{[r]}1a^{[r-1]}1 \ldots 1a^{[2]}1a^{[1]}1$ are subwords of w (i.e. are subwords of the period-q sequence $w^\infty \in \{0,1\}^{\mathbb{N}}$). That is, for arbitrarily large r we can find equal length subwords $v_0^{[r]}, v_1^{[r]}$ of w^∞ such that $|v_1^{[r]}|_1 - |v_0^{[r]}|_1 = r+1$. This contradicts the periodicity of w^∞, since clearly for any periodic sequence the difference in 1-norm of any two length-n subwords is bounded independently of n. Therefore Theorem B is proved. $\square$

5. Proof of Theorem C

To prove Theorem C we will need several preliminary results. The proof of the following lemma is left as an exercise.

Lemma 8. *Let p, q be positive integers with $1 \leq p \leq q$ and $\gcd(p, q) = 1$. Given $w \in \mathcal{W}_{p,q}$,* *and the corresponding lexicographic array* $A[w] = (A[w]_{i,j})_{i,j=0}^{q-1}$. *Then*

$$\begin{pmatrix} A[w]_{i,0} \\ A[w]_{i+1,0} \end{pmatrix} = \begin{pmatrix} (w_{(i)})_0 \\ (w_{(i+1)})_0 \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{if } 0 \le i \le q-p-2 \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{if } i = q-p-1 \\ \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \text{if } q-p \le i \le q-2 \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{if } i = q-1. \end{cases}$$

Moreover,

$$\begin{pmatrix} A[w]_{q-p-1,0} & A[w]_{q-p-1,1} \\ A[w]_{q-p,0} & A[w]_{q-p,1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 9. *Suppose* $w \in \mathcal{W}_{p,q}$ *has the lexicographic constant shift property. Let*

$$w_{(0)} <_L w_{(1)} <_L \ldots <_L w_{(q-1)}$$

be the lexicographic ordering of w.

For any $0 \le i \le q-2$, *there exists a unique* $k_0 = k_0(i)$ *such that*

$$\begin{pmatrix} (w_{(i)})_{k_0} & (w_{(i)})_{k_0+1} \\ (w_{(i+1)})_{k_0} & (w_{(i+1)})_{k_0+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

If $k \notin \{k_0, k_0+1\}$ *then* $(w_{(i)})_k = (w_{(i+1)})_k$.

Example 10. Let $w = 0010101$. We noted in Example 5 that w has the constant lexicographic shift property. From the lexicographic array in Example 4 it is readily verified that w satisfies the conclusion of Lemma 9.

Proof of Lemma 9. The definition of the lexicographic shift constant $\mathfrak{m} = \mathfrak{m}(w)$ means that the k^{th} digit of $w_{(i)}$ equals the $(k - i\mathfrak{m} \pmod q))^{th}$ digit of $w_{(0)}$. Now there is some unique $0 \le r \le q-1$ such that $k - i\mathfrak{m} = r\mathfrak{m} \pmod q$. Thus the k^{th} digit of $w_{(i)}$ equals the $(r\mathfrak{m} \pmod q))^{th}$ digit of $w_{(0)}$, which in turn equals the 0^{th} digit of $w_{(r)}$.

Analogously, the k^{th} digit of $w_{(i+1)}$ equals the $(r\mathfrak{m} \pmod q))^{th}$ digit of $w_{(1)}$, which in turn equals the 0^{th} digit of $w_{(r+1)}$.

That is,

$$\begin{pmatrix} (w_{(i)})_k \\ (w_{(i+1)})_k \end{pmatrix} = \begin{pmatrix} (w_{(r)})_0 \\ (w_{(r+1)})_0 \end{pmatrix}.$$

By Lemma 8 we deduce that $(w_{(i)})_k = (w_{(r)})_0 = (w_{(r+1)})_0 = (w_{(i+1)})_k$ unless $r = q-p-1$ or $r = q-1$.

If $r = q-p-1$ then again by Lemma 8 we see that

$$\begin{pmatrix} (w_{(i)})_{k_0} & (w_{(i)})_{k_0+1} \\ (w_{(i+1)})_{k_0} & (w_{(i+1)})_{k_0+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

as required. $\square$

We are now able to prove one half of Theorem C, namely that if a word $w \in \mathcal{W}_{p,q}$ has the lexicographic constant shift property then it must be balanced.

Proof of the "if" part of Theorem C.

Suppose $w \in \mathcal{W}_{p,q}$ has the lexicographic constant shift property. Let $i = 0$ and apply Lemma 9. This means there exists k_0 such that $w_{(0)}$ and $w_{(1)}$ (the first and second words in the lexicographic ordering of w) satisfy $(w_{(0)})_k = (w_{(1)})_k$ for all $k \notin \{k_0, k_0+1\}$, and that

$$\begin{pmatrix} (w_{(0)})_{k_0} & (w_{(0)})_{k_0+1} \\ (w_{(1)})_{k_0} & (w_{(1)})_{k_0+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore there exist words u, v such that

$$w_{(0)} = u01v \quad , \quad w_{(1)} = u10v.$$

But $w_{(0)}, w_{(1)}$ are cyclic shifts of each other, so that $1vu0$ is a cyclic shift of $0vu1$.

Applying Lemma 5 we see that vu is Sturmian bispecial, so that w is balanced as required. $\square$

The rest of the article is devoted to the proof of the "only if" part of Theorem C: showing that every balanced orbit has the lexicographic constant shift property, and verifying the formula for the lexicographic shift constant.

Let $[w_{p,q}]$ denote the orbit of the post-minimal balanced word $w_{p,q}$. Clearly $[w_{p,q}]$ has cardinality q. For $i \in \{0,1\}$ let $[w_{p,q}]_i$ denote all cyclic shifts of $w_{p,q}$ ending in an i. The next lemma shows how the morphisms τ_i may be used to generate the orbit of $w_{p,q}$ from the orbit of $w_{p',q'}$, where $\frac{p'}{q'}$ is the predecessor of $\frac{p}{q}$.

Lemma 10. *Let p, q be positive integers with $1 \le p < q \neq 2$ and $\gcd(p,q) = 1$. If $\frac{p'}{q'}$ is the predecessor of $\frac{p}{q}$, where $1 \le p' < q'$ and $\gcd(p',q') = 1$, then*

(1) *If $0 < \frac{p}{q} < \frac{1}{2}$ then each $w \in [w_{p,q}]$ is either of the form $w = \tau_0(w')$ for some $w' \in [w_{p',q'}]$ (there are q' such w) or of the form $w = 1\tau_0(w')1^{-1}$ for some $w' \in [w_{p',q'}]_1$ (there are p' such w).*

(2) *If $\frac{1}{2} < \frac{p}{q} < 1$ then each $w \in [w_{p,q}]$ is either of the form $w = \tau_1(w')$ for some $w' \in [w_{p',q'}]$ (there are q' such w) or of the form $w = 0\tau_1(w')0^{-1}$ for some $w' \in [w_{p',q'}]_0$ (there are $q'-p'$ such w).*

Proof. We prove (1) and leave the proof of (2), which is essentially identical, as an exercise for the reader. If $0 < \frac{p}{q} < \frac{1}{2}$ then $w_{p,q} = \tau_0(w_{p',q'})$, by Lemma 3. Since $\tau_0(0) = 0$ and $\tau_0(1) = 01$, it follows that if $w \in [w_{p,q}]$ begins in a 0, then it is the image under τ_0 of a cyclic permutation of $w_{p',q'}$, while if w begins in a 1 then $1^{-1}w1$ is the image under τ_0 of a cyclic permutation of $w_{p',q'}$.

To see that there are q' such w beginning in a 0 we will show that if $w_1, w_2 \in [w_{p',q'}]$ with $w_1 <_L w_2$, then $\tau_0(w_1) <_L \tau_0(w_2)$. But, if $w_1 <_L w_2$, then there exist words u, u_1, u_2 (with u possibly empty) such that $w_1 = u0u_1$ and $w_2 = u1u_2$. Since w_1, w_2 have equal numbers of 0s and 1s, it follows that $|u_1| = |u_2| > 0$. Applying τ_0 to both w_1 and w_2, we see that $\tau_0(u)00$ is a prefix of $\tau_0(w_1)$ while $\tau_0(u)01$ is a prefix of $\tau_0(w_2)$, whence $\tau_0(w_1) <_L \tau_0(w_2)$.

Similarly, to see that there are p' words w of the form $w = 1\tau_0(w')1^{-1}$ with $w' \in [w_{p',q'}]_1$, it suffices to show that if $w_1 <_L w_2$ with $w_1, w_2 \in [w_{p',q'}]_1$, then $1\tau_0(w_1)1^{-1} <_L 1\tau_0(w_2)1^{-1}$. But if $w_1 <_L w_2$ then there exist words u, u_1, u_2 (u possibly empty) with $w_1 = u0u_11$ and $w_2 = u1u_21$, and $|u_1| = |u_2| > 0$. It follows that $\tau_0(u)00$ is a prefix of $\tau_0(w_1)$ while $\tau_0(u)01$ is a prefix of $\tau_0(w_2)$. Thus $1\tau_0(u)00$ is a prefix of $1\tau_0(w_1)1^{-1}$ while $1\tau_0(u)01$ is a prefix of $1\tau_0(w_2)1^{-1}$, whence $1\tau_0(w_1)1^{-1} <_L 1\tau_0(w_2)1^{-1}$. $\square$

Example 11. If $\frac{p}{q} = \frac{3}{7} = [2, 3]$ then we are in case (1) of Lemma 10. Now $w_{3,7} = 0101010$, so

$$[w_{3,7}] = \{0010101,\ 0100101,\ 0101001,\ 0101010,\ 1001010,\ 1010010,\ 1010100\}.$$

The predecessor is $\frac{p'}{q'} = [1, 3] = \frac{3}{4}$, and we have $w_{3,4} = \tau_1^3(0) = 1110$. Therefore

$$[w_{3,4}] = [w_{3,4}]_0 \cup [w_{3,4}]_1 = \{1110\} \cup \{0111,\ 1011,\ 1101\}.$$

We now see that

$$\tau_0 : \begin{cases} 0111 \mapsto 0010101 \\ 1011 \mapsto 0100101 \\ 1101 \mapsto 0101001 \\ 1110 \mapsto 0101010 \end{cases},$$

hence also

$$1\tau_0(0111)1^{-1} = 1001010 \quad , \quad 1\tau_0(1011)1^{-1} = 1010010 \quad , \quad 1\tau_0(1101)1^{-1} = 1010100\ ,$$

as predicted by Lemma 10.

Definition. Let $\frac{p}{q} = [n_0, n_1, \ldots, n_{2k+1}]$ be the even length continued fraction expansion for $\frac{p}{q}$, where p, q are positive integers such that $1 \le p < q$ and $\gcd(p, q) = 1$.

If $p \neq 1$, then define a word $z_{p,q} \in \{0, 1\}^{**}$ by

$$z_{p,q} = \begin{cases} \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_0^{n_{2k}} \circ \tau_1^{n_{2k+1}-1}(0) & \text{if } n_{2k+1} \ge 2 \\ \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_0^{n_{2k-2}} \circ \tau_1^{n_{2k-1}}(0) & \text{if } n_{2k+1} = 1 \end{cases}.$$

We then define the positive integer $\mathfrak{m}_{p,q}$ by:

$$\mathfrak{m}_{p,q} = \begin{cases} 1 & \text{if } p = 1 \\ |z_{p,q}| & \text{otherwise} \end{cases}.$$

It follows then that $\mathfrak{m}_{p,q}$ is simply the penultimate convergent of $\frac{p}{q}$ where we express $\frac{p}{q}$ in terms of its odd length continued fraction expansion. Using the well known formula (see for example [HaWr, Theorem 150]) $p_nq_{n-1} - p_{n-1}q_n = (-1)^{n+1}$ relating consecutive convergents, we deduce that $p\,\mathfrak{m}_{p,q} - p_{n-1}q = 1$ (since $n = 2k + 1$ is odd). Therefore the quantity $\mathfrak{m}_{p,q}$ defined above is simply the multiplicative inverse of p modulo q.

The "only if" part of Theorem C will follow from part (4) of the following Theorem D, which gives a very precise description of the lexicographic array for balanced periodic orbits.

Theorem D. [2] *Let $\frac{p}{q} = [n_0, n_1, \dots, n_{2k+1}]$ be the even length continued fraction expansion for $\frac{p}{q}$, where p, q are positive integers with $1 \le p < q$ and $\gcd(p,q) = 1$. Let*

$$w_{(0)} <_L w_{(1)} <_L \cdots <_L w_{(q-1)}$$

denote the lexicographic ordering of the unique balanced orbit in $\mathbb{W}_{p,q}$.

For each $0 \le i \le q-1$, let $u_{(i)}$ denote the length-$\mathfrak{m}_{p,q}$ prefix of $w_{(i)}$, and define $v_{(i)}$ by $w_{(i)} = u_{(i)} v_{(i)}$. Then

1. *$|u_{(i)}|_1 = |u_{(0)}|_1$ for all $0 \le i < q-1$, and $|u_{(q-1)}|_1 = |u_{(0)}|_1 + 1$.*
2. *$u_{(q-1)}$ begins and ends in 1, and $v_{(0)}, \dots, v_{(p-1)}$ all end in 1, while $v_{(p)}, \dots, v_{(q-1)}$ all end in 0.*
3. *If $p \ne 1$, then $u_{(0)}, \dots, u_{(p-2)}$ all end in 1, $u_{(p-1)}$ ends in 10, and $u_{(0)} = 0 z_{p,q} 0^{-1}$. If $p = 1$, then $u_{(i)} = 0$ for $0 \le i < q-1$, and $u_{(q-1)} = 1$.*
4. *$w_{(i+1)} = v_{(i)} u_{(i)}$ for $0 \le i < q-1$, and $w_{(0)} = v_{(q-1)} u_{(q-1)}$.*
5. *$u_{(q-1)}$ and $v_{(q-1)}$ are both palindromes.*

Example 12. Suppose $\frac{p}{q} = \frac{3}{7} = [2, 3]$. Then $z_{3,7} = \tau_0 \circ \tau_1^2(0) = \tau_0(110) = 01010$, so that $\mathfrak{m}_{p,q} = |z_{p,q}| = 5$. The balanced orbit in $\mathbb{W}_{3,7}$ is generated by $w = 0010101$, and in the notation of Theorem D we have

i	$w_{(i)}$	$u_{(i)}$	$v_{(i)}$
0	0010101	00101	01
1	0100101	01001	01
2	0101001	01010	01
3	0101010	01010	10
4	1001010	10010	10
5	1010010	10100	10
6	1010100	10101	00

We can easily check the conclusions of Theorem D for this example:

1. $|u_{(i)}|_1 = 2 = |u_{(0)}|_1$ for $0 \le i \le 5$, and $|u_{(6)}|_1 = 3 = |u_{(0)}|_1 + 1$.
2. $u_{(6)} = 10101$ begins and ends in 1, and $v_{(0)}, v_{(1)}, v_{(2)}$ all equal 01, hence all end in 1. Also $v_{(0)}, v_{(1)}, v_{(2)}, v_{(3)}$ all end in 0.
3. $u_{(0)} = 00101 = 0 z_{3,7} 0^{-1}$ and $u_{(1)} = 01001$ both end in 1, while $u_{(2)}$ ends in 10.
4. The key properties $w_{(i+1)} = v_{(i)} u_{(i)}$ for $0 \le i < q-1$, and $w_{(0)} = v_{(q-1)} u_{(q-1)}$ are readily verified.
5. $u_{(6)} = 10101$ and $v_{(6)} = 00$ are both palindromes.

Proof of Theorem D. We first verify the result in case $p = 1$, or equivalently when the even length continued fraction expansion of $\frac{p}{q}$ is of the form $\frac{p}{q} = [n_0, 1]$. In this case $w_{(i)} = 0^{q-(1+i)} 1 0^i$ and hence $u_{(i)} = 0$ (for $0 \le i < q-1$), $u_{(q-1)} = 1$, and $v_{(i)} = 0^{q-(1+i+1)} 1 0^i$,

[2] cf. work of deLuca and Mignosi [dLMi] on related properties of Sturmian words.

establishing (1), (2), and (3). Moreover for $0 \le i < q-1$ we have

$$\begin{aligned} w_{(i+1)} &= 0^{q-(1+i+1)}10^{i+1} \\ &= 0^{q-(1+i+1)}10^{i}0 \\ &= v_{(i)}u_{(i)}, \end{aligned}$$

thereby establishing (4).

Finally we see that $u_{(q-1)} = 1$ while $v_{(q-1)} = 0^{q-1}$, thus establishing (5).

Next suppose $p \neq 1$. We will proceed by induction on q. The base case here is $\frac{p}{q} = \frac{2}{3} = [1,2]$. In this case $\mathfrak{m}_{p,q} = |z_{p,q}| = |\tau_1(0)| = |10| = 2$, and $w_{(0)} = 011$, $w_{(1)} = 101$, and $w_{(2)} = 110$, from which it follows that $u_{(0)} = 01$, $u_{(1)} = 10$, $u_{(2)} = 11$, $v_{(0)} = v_{(1)} = 1$ and $v_{(2)} = 0$, thereby establishing (1)-(5), as required.

Suppose $\frac{p'}{q'}$ is the predecessor of $\frac{p}{q}$, where $1 \le p' < q'$ and $\gcd(p',q') = 1$. Our inductive hypothesis will be that Theorem D holds for $\frac{p'}{q'}$, and we will deduce that it also holds for $\frac{p}{q}$. The proof will differ according to whether $0 < \frac{p}{q} < \frac{1}{2}$ or $\frac{1}{2} < \frac{p}{q} < 1$. We shall consider only the first of these cases, leaving the almost identical proof of the second case as an exercise for the reader.

Suppose then that $0 < \frac{p}{q} < \frac{1}{2}$, and that $\frac{p}{q} = [n_0, n_1, \dots, n_{2k+1}]$ is its even length continued fraction expansion. We recall this means that $n_0 \ge 2$, $\frac{p'}{q'} = [n_0 - 1, n_1, \dots, n_{2k+1}]$ (and hence also $p' \neq 1$), $p = p'$ and $q = p' + q'$.

Let

$$z_{p',q'} = \begin{cases} \tau_0^{n_0-2} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k+1}-1}(0) & \text{if } n_{2k+1} \ge 2 \\ \tau_0^{n_0-2} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}}(0) & \text{if } n_{2k+1} = 1 \end{cases},$$

and set $\mathfrak{m}_{p',q'} = |z_{p',q'}|$. Let $w'_{(0)} <_L w'_{(1)} <_L \cdots <_L w'_{(q'-1)}$ be the q' cyclic permutations of $w_{p',q'}$ in increasing lexicographic order, and for $0 \le i \le q'-1$ we write $w'_{(i)} = u'_{(i)}v'_{(i)}$ where $u'_{(i)}$ is the prefix of $w'_{(i)}$ of length $\mathfrak{m}_{p',q'}$. From the inductive hypothesis we have:

(1′) $|u'_{(i)}|_1 = |u'_{(0)}|_1$ for $0 \le i < q'-1$, and $|u'_{(q'-1)}|_1 = |u'_{(0)}|_1 + 1$.

(2′) $u'_{(q'-1)}$ begins and ends in a 1, and $v'_{(0)}, \dots, v'_{(p'-1)}$ all end in a 1, and $v'_{(p')}, \dots, v'_{(q'-1)}$ all end in a 0.

(3′) $u'_{(0)}, \dots, u'_{(p'-2)}$ all end in a 1, $u'_{(p'-1)}$ ends in 10, and $u'_{(0)} = 0z_{p',q'}0^{-1}$.

(4′) $w'_{(i+1)} = v'_{(i)}u'_{(i)}$ for $0 \le i < q'-1$, and $w'_{(0)} = v'_{(q'-1)}u'_{(q'-1)}$.

(5′) $u'_{(q'-1)}$ and $v'_{(q'-1)}$ are both palindromes.

Let $w_{(0)} <_L w_{(1)} <_L \cdots <_L w_{(q)}$ be the q cyclic permutations of the post-minimal balanced word $w_{p,q}$. From Lemma 10 and (2′) we have

$$\begin{aligned}
w_{(0)} &= \tau_0(w'_{(0)}) = \tau_0(u'_{(0)})\tau_0(v'_{(0)})\\
w_{(1)} &= \tau_0(w'_{(1)}) = \tau_0(u'_{(1)})\tau_0(v'_{(1)})\\
&\vdots\\
w_{(q'-2)} &= \tau_0(w'_{(q'-2)}) = \tau_0(u'_{(q'-2)})\tau_0(v'_{(q'-2)})\\
w_{(q'-1)} &= \tau_0(w'_{(q'-1)}) = \tau_0(u'_{(q'-1)})\tau_0(v'_{(q'-1)})\\
w_{(q')} &= 1\tau_0(w'_{(0)})1^{-1} = 1\tau_0(u'_{(0)})\tau_0(v'_{(0)})1^{-1}\\
&\vdots\\
w_{(q'+p'-1)} &= 1\tau_0(w'_{(p'-1)})1^{-1} = 1\tau_0(u'_{(p'-1)})\tau_0(v'_{(p'-1)})1^{-1}.
\end{aligned}$$

Recall that $\mathfrak{m}_{p,q} = |z_{p,q}|$ where $z_{p,q} = \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k+1}-1}(0)$ if $n_{2k+1} \geq 2$, and $z_{p,q} = \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}}(0)$ if $n_{2k+1} = 1$. Then

$$\mathfrak{m}_{p,q} = |z_{p,q}| = |\tau_0(z_{p',q'})| = |\tau_0(u'_{(0)})|,$$

where the last equality follows from (3′). We now compute the length-$\mathfrak{m}_{p,q}$ prefixes $u_{(i)}$ of $w_{(i)}$, and show they satisfy the conditions (1), (2), (3), and (4) of the proposition. Using (1′), (2′), and (3′) we deduce that

$$\begin{aligned}
u_{(0)} &= \tau_0(u'_{(0)})\\
&\vdots\\
u_{(q'-2)} &= \tau_0(u'_{(q'-2)})\\
u_{(q'-1)} &= \tau_0(u'_{(q'-1)})1^{-1}\\
u_{(q')} &= 1\tau_0(u'_{(0)})1^{-1}\\
&\vdots\\
u_{(q'+p'-2)} &= 1\tau_0(u'_{(p'-2)})1^{-1}\\
u_{(q'+p'-1)} &= 1\tau_0(u'_{(p'-1)})0^{-1}.
\end{aligned}$$

It is clear from the above expressions that each $u_{(i)}$ is a prefix of $w_{(i)}$ of length $\mathfrak{m}_{p,q}$, and that $|u_{(i)}|_1 = |u_{(0)}|_1$ for $0 \leq i \leq q-2 = p'+q'-2$ while $|u_{(q-1)}|_1 = |u_{(0)}|_1 + 1$, thus establishing (1).

Since $u'_{(p'-1)}$ ends in 10 it follows that $u_{(q-1)}$ begins and ends in a 1. Since $p = p'$, and $v'_{(0)}, \ldots, v'_{(p'-1)}$ all end in a 1, we deduce that $w_{(0)}, \ldots, w_{(p'-1)}$ all end in a 1, and hence that $v_{(0)}, \ldots, v_{(p'-1)}$ all end in a 1. This establishes (2).

Clearly $u_{(0)} = \tau_0(u'_{(0)}) = 0\tau_0(z_{p',q'})0^{-1} = 0z_{p,q}0^{-1}$, and since $u_{(i)} = \tau_0(u'_{(i)})$ for $i \leq 0 \leq p-1 = p'-1$ we see that (3) follows directly from (3′).

To verify (4) we compute the prefixes $v_{(i)}$ and show that $u_{(i+1)}v_{(i+1)} = v_{(i)}u_{(i)}$. We have

$$\begin{aligned}
v_{(0)} &= \tau_0(v'_{(0)}) \\
&\vdots \\
v_{(q'-2)} &= \tau_0(v'_{(q'-2)}) \\
v_{(q'-1)} &= 1\tau_0(v'_{(q'-1)}) \\
v_{(q')} &= 1\tau_0(v'_{(0)})1^{-1} \\
&\vdots \\
v_{(q'+p'-2)} &= 1\tau_0(v'_{(p'-2)})1^{-1} \\
v_{(q'+p'-1)} &= 0\tau_0(v'_{(p'-1)})1^{-1}.
\end{aligned}$$

For $0 \le i \le q'-2$ we have

$$w_{(i+1)} = \tau_0(w'_{(i+1)}) = \tau_0(v'_{(i)}u'_{(i)}) = v_{(i)}u_{(i)}.$$

Next

$$w_{(q'-1)} = 1\tau_0(w'_{(0)})1^{-1} = 1\tau_0(v'_{(q'-1)}u'_{(q'-1)})1^{-1} = v_{(q'-1)}u_{(q'-1)}.$$

Next for $0 \le i \le p'-2$ we have

$$\begin{aligned}
w_{(q'+i+1)} &= 1\tau_0(w'_{(i+1)})1^{-1} \\
&= 1\tau_0(v'_{(i)}u'_{(i)})1^{-1} \\
&= 1\tau_0(v'_{(i)})1^{-1} \cdot 1\tau_0(u'_{(i)})1^{-1} \\
&= v_{(q'+i)}u_{(q'+i)}\,.
\end{aligned}$$

To establish (5) we will need two additional lemmas:

Lemma D.1. *Under the hypotheses of Theorem D, if $p \neq 1$, then*

$$w_{p,q} = \begin{cases} \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k+1}-1}(0)\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}} \circ \tau_0^{n_{2k}-1} \circ \tau_1(0) & \text{if } n_{2k+1} \ge 2 \\ \tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}}(0)\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}} \circ \tau_0^{n_{2k}-1} \circ \tau_1(0) & \text{if } n_{2k+1} = 1 \end{cases}.$$

Proof. If $n_{2k+1} \ge 2$ then the right hand side of the first equation equals

$$\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_0^{n_{2k}-1} \left(\tau_0 \circ \tau_1^{n_{2k+1}-1}(0)\tau_1(0)\right).$$

But

$$\begin{aligned}
\tau_0 \circ \tau_1^{n_{2k+1}-1}(0)\tau_1(0) &= (01)^{n_{2k+1}-1}010 \\
&= (01)^{n_{2k+1}}0 \\
&= \tau_0 \circ \tau_1^{n_{2k+1}}(0),
\end{aligned}$$

thereby establishing the first case. Similarly if $n_{2k+1} = 1$, then the right hand side of the second equation is equal to

$$\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \cdots \circ \tau_1^{n_{2k-1}} \left(0\tau_0^{n_{2k}-1} \circ \tau_1(0)\right).$$

But

$$\begin{aligned} 0\tau_0^{n_{2k}-1} \circ \tau_1(0) &= 00^{n_{2k}-1}10 \\ &= 0^{n_{2k}}10 \\ &= \tau_0^{n_{2k}} \circ \tau_1(0) \\ &= \tau_0^{n_{2k}} \circ \tau_1^{n_{2k+1}}(0), \end{aligned}$$

thereby establishing the second case.

In either of the factorisations of $w_{p,q}$ in Lemma D.1, the first factor is of length $\mathfrak{m}_{p,q}$. Using the well known fact (see for instance Proposition III.7 in [RiZa]) that for each choice of positive integers $n_0, n_1, \ldots, n_{2k+1}$ there exists a palindrome x (depending on $n_0, n_1, \ldots, n_{2k+1}$) such that $\tau_0^{n_0-1} \circ \tau_1^{n_1} \circ \tau_0^{n_2} \circ \cdots \circ \tau_1^{n_{2k+1}}(0) = x10$. Moreover the palindrome x is a Sturmian bispecial word. Therefore we can combine Lemmas 3 and D.1 to obtain

Lemma D.2. *Under the hypotheses of Theorem D, if $p \neq 1$, then there exist palindromes x and y such that $w_{p,q} = x10y10$, and $\mathfrak{m}_{p,q} = |x10|$.*

We now return to the proof of part (5) of Theorem D. By Lemma D.2 we know there exist palindromes x and y such that $w_{p,q} = x10y10$, with $\mathfrak{m}_{p,q} = |x10|$. By Proposition 6 we deduce that $w_{(q-1)} = 1(x10y)0$ so that $u_{(q-1)} = 1x1$ and $v_{(q-1)} = 0y0$, and hence $u_{(q-1)}$ and $v_{(q-1)}$ are both palindromes, as required. This concludes our proof of Theorem D. $\square$

Proof of the "only if" part of Theorem C. Let w be the balanced orbit in $\mathbb{W}_{p,q}$, and let $w_{(0)} <_L w_{(1)} <_L \ldots <_L w_{(q-1)}$ denote its lexicographic ordering. From part (4) of Theorem D we have $w_{(i+1)} = v_{(i)}u_{(i)}$ for all $0 \leq i \leq q-1$. But $u_{(i)}v_{(i)} = w_{(i)}$ by definition, and $u_{(i)}$ is of length $\mathfrak{m} = \mathfrak{m}_{p,q}$, so we see that $w_{(i+1)} = \sigma^{\mathfrak{m}}(w_{(i)})$ for all $0 \leq i \leq q-1$. That is, w has the lexicographic constant shift property, and the lexicographic shift constant is precisely $\mathfrak{m}_{p,q}$.

Since $w_{(i+1)} = \sigma^{\mathfrak{m}}(w_{(i)})$, and the first $q-p$ of the $w_{(i)}$ begin in 0, we deduce that:

Corollary. *The arithmetic sequence $\{r\,\mathfrak{m}_{p,q}\}_{r=0}^{q-p-1}$ taken modulo q gives the positions of the 0's in the lexicographically minimal balanced word in $\mathcal{W}_{p,q}$. Similarly, if $w' \in \mathcal{W}_{p,q}$ denotes the lexicographically smallest balanced word whose first digit is a 1, then the arithmetic sequence $\{r\,\mathfrak{m}_{p,q}\}_{r=0}^{p-1}$ taken modulo q gives the positions of the 1's in w'.*

References

[BeSe] J. Berstel, P. Séébold, *Sturmian words*, in: *Algebraic Combinatorics on Words* (Chapter 2), vol. 90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK, 2002, pp. 45–110.

[Bou] T. Bousch, *Le poisson n'a pas d'arêtes*, Ann. Inst. Henri Poincaré (Proba. et Stat.) **36** (2000), 489–508.

[BuSe] S. Bullett & P. Sentenac, *Ordered orbits of the shift, square roots, and the devil's staircase,*, Math. Proc. Camb. Phil. Soc. **115** (1994), 451–481.

[dLMi] A. de Luca & F. Mignosi, *Some combinatorial properties of Sturmian words*, Theor. Comp. Sci. **136** (1994), 361–385.

[Ga] S. Gan, *Sturmian sequences and the lexicographic world*, Proc. of the A.M.S. **129**, 1445–1451.

[HaWr] G. H. Hardy & E. M. Wright, *An introduction to the theory of numbers, 5th edition*, Oxford University Press, 1979.

[Jen] O. Jenkinson, *Frequency locking on the boundary of the barycentre set*, Experimental Mathematics **9** (2000), 309–317.

[JuPi] J. Justin & G. Pirillo, *On a characteristic property of Arnoux Rauzy sequences*, Theoretical Informatics & Applications, to appear.

[Kel] K. Karsten, *Symbolic dynamics for angle-doubling on the circle. III: Sturmian sequences and the quadratic map.*, Ergod. Th. & Dyn. Sys. **14** (1994), 787–805.

[MoHe] M. Morse & G. A. Hedlund, *Symbolic dynamics, II: Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42.

[Pir1] G. Pirillo, *A new characteristic property of the palindrome prefixes of a standard Sturmian word*, (see http://www.mat.univie.ac.at/∼slc/), Seminaire Lotharingien de Combinatoire **43** (1999), 3 pp.

[Pir2] G. Pirillo, *A curious characteristic property of standard Sturmian words*, Algebraic Combinatorics and Computer Science, A tribute to Gian-Carlo Rota, (H. Crapo and D. Senato Eds), Springer Italia, Milan, 2001, pp. 541–546.

[RiZa] R. Risley, L.Q. Zamboni, *A generalization of Sturmian sequences: combinatorial structure and transcendence*, Acta Arith. **XCV.2** (2000), 167–184.

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

O. Jenkinson, School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK.

E-mail address: `omj@maths.qmul.ac.uk`

L. Q. Zamboni, Department of Mathematics, University of North Texas, Denton, TX 76203-1430, USA.

E-mail address: `luca@unt.edu`