

# BALANCED WORDS AND MAJORIZATION

OLIVER JENKINSON

ABSTRACT. When expressed in terms of base-2 expansions, balanced words are *majorized* by other words of the same slope. Consequently they have smaller standard deviation and larger geometric mean than all words with given arithmetic mean (or slope), they can be expressed as a doubly stochastic average of any such word, and they can be derived from any such word by a finite number of transfers.

## 1. INTRODUCTION

This article is concerned with words

$$w = w_1 w_2 \dots w_m \in \{0, 1\}^+ := \bigcup_{q \geq 1} \{0, 1\}^q,$$

and their base-2 expansions

$$E_2(w) = \sum_{k=1}^m w_k 2^{m-k}.$$

Write  $|w| = m$ , the *length* of  $w$ , and  $|w|_1 = \text{card}(\{1 \leq i \leq m : w_i = 1\})$  its *1-length*. Define the *cyclic shift*  $\sigma : \{0, 1\}^m \rightarrow \{0, 1\}^m$  by  $\sigma(w_1 \dots w_m) = w_2 \dots w_m w_1$ . A *cyclic subword* of  $w$  is any length- $q$  prefix of some  $\sigma^{i-1}(w)$ ,  $1 \leq i, q \leq m$ .

To any word  $w = w_1 \dots w_m$  we associate its *orbit*  $\mathcal{O}(w)$ , the vector

$$\mathcal{O}(w) = (\mathcal{O}_1(w), \dots, \mathcal{O}_m(w))$$

consisting of the iterated cyclic shifts  $w, \sigma(w), \dots, \sigma^{m-1}(w)$  arranged in lexicographic order<sup>1</sup>

$$\mathcal{O}_1(w) \leq \mathcal{O}_2(w) \leq \dots \leq \mathcal{O}_m(w). \quad (1)$$

Define the *base-2 orbit*

$$\mathcal{I}(w) = (\mathcal{I}_1(w), \dots, \mathcal{I}_m(w)),$$

where

$$\mathcal{I}_i(w) := E_2(\mathcal{O}_i(w)) \quad \text{for } 1 \leq i \leq m.$$

---

The author was supported by an EPSRC Advanced Research Fellowship.

<sup>1</sup>The lexicographic order on  $\{0, 1\}^m$  is defined by:  $w < w'$  if there exists  $j \in \{1, \dots, m\}$  with  $w_k = w'_k$  for all  $k = 1, \dots, j-1$ , and  $w_j < w'_j$ , and  $w \leq w'$  if either  $w < w'$  or  $w = w'$ .

We shall be interested in certain quantities depending on the base-2 orbit  $\mathcal{I}(w)$ , for example the product

$$P(w) := \prod_{i=1}^q \mathcal{I}_i(w).$$

For all words  $w$  with a fixed length and 1-length, we shall compare the products  $P(w)$ ; it turns out (see Theorem 1.2 below) that  $P(w)$  is as large as possible when  $w$  is *balanced*. By a balanced word<sup>2</sup> we mean one with the property that, if  $z$  and  $z'$  are cyclic subwords of  $w$  with  $|z| = |z'|$ , then  $||z|_1 - |z'|_1| \leq 1$ . Clearly,  $w$  is balanced if and only if each of its cyclic shift iterates is balanced.

**Example 1.1.** The word 0010101 is balanced, and

$$P(0010101) = 21 \times 37 \times 41 \times 42 \times 74 \times 82 \times 84 = 681991597728.$$

The unbalanced length-7 words with 1-length equal to 3 are 0000111, 0001011, 0001101, and 0010011, together with their cyclic shift iterates. The corresponding products are

$$P(0000111) = 7 \times 14 \times 28 \times 56 \times 67 \times 97 \times 112 = 111850181632,$$

$$P(0001011) = 11 \times 22 \times 44 \times 49 \times 69 \times 88 \times 98 = 310471658112,$$

$$P(0001101) = 13 \times 26 \times 35 \times 52 \times 70 \times 81 \times 104 = 362747548800,$$

$$P(0010011) = 19 \times 25 \times 38 \times 50 \times 73 \times 76 \times 100 = 500707000000,$$

each of which is smaller than  $P(0010101)$ .

More generally, if  $1 \leq p < q$  are coprime integers, let  $\mathcal{W}_{p,q}$  denote the set of all length- $q$  words whose 1-length equals  $p$ . There are precisely  $q$  balanced words in  $\mathcal{W}_{p,q}$ , all of which are in the same orbit (see e.g. [9, 24]), so if  $\mathbb{W}_{p,q}$  is defined to be the set of all orbits of words in  $\mathcal{W}_{p,q}$ , there is a unique balanced orbit in  $\mathbb{W}_{p,q}$ . Since  $\mathcal{I}(w)$  only depends on the orbit of  $w$ , it will be notationally convenient to represent orbits by one of their components; for consistency we shall always use  $\mathcal{O}_1(w)$ , the lexicographically smallest component. So for example the orbit  $(00101, 01001, 01010, 10010, 10100)$  will be represented notationally by 00101.

**Theorem 1.2.** *Suppose  $1 \leq p < q$  are coprime integers. For  $w \in \mathbb{W}_{p,q}$ , the product  $P(w) = \prod_{i=1}^q \mathcal{I}_i(w)$  is maximized precisely when  $w$  is balanced.*

**Example 1.3.** The simplest non-trivial instance of Theorem 1.2 is when  $(p, q) = (2, 5)$ , in which case  $\mathbb{W}_{2,5} = \{00011, 00101\}$ , with 00101 the balanced orbit, and  $P(00011) = 3 \times 6 \times 12 \times 17 \times 24 = 88128 < 162000 = 5 \times 9 \times 10 \times 18 \times 20 = P(00101)$ .

More generally, the balanced orbit in  $\mathbb{W}_{p,q}$  can be compared to orbits of words  $w$  whose *slope*  $sl(w) := |w|_1/|w|$  is equal to  $p/q$ . If  $w$  has length  $Q$  (necessarily a multiple of  $q$ ) then any length- $Q$  balanced word is the  $(Q/q)$ -fold concatenation<sup>3</sup>

<sup>2</sup>Finite balanced words, defined in [24], constitute a subclass of the collection of *Sturmian words* (see e.g. [1, 4, 32] for more on Sturmian words, and on infinite balanced words).

<sup>3</sup>In general the *k-fold concatenation* of a word  $w = w_1 \dots w_m$  is the length- $km$  word  $w^k$  whose  $i$ -th entry is  $w_{i \pmod{m}}$ .

$b^{Q/q}$  of a balanced word  $b \in \mathcal{W}_{p,q}$ . In particular, there is a unique length- $Q$  balanced orbit of slope  $p/q$ . The following result is a generalisation of Theorem 1.2:

**Theorem 1.4.** *If the unbalanced word  $w \in \{0,1\}^Q$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers, and  $b \in \mathcal{W}_{p,q}$  is balanced, then  $P(w) < P(b^{Q/q})$ .*

**Example 1.5.** The words 0011 and 001011 both have slope  $1/2$ , as does the balanced word 01. The product

$$P(0011) = 3 \times 6 \times 9 \times 12 = 1944$$

is smaller than

$$P(0101) = 5 \times 5 \times 10 \times 10 = 2500,$$

while

$$P(001011) = 11 \times 22 \times 25 \times 37 \times 44 \times 50 = 492470000$$

is smaller than

$$P(010101) = 21 \times 21 \times 21 \times 42 \times 42 \times 42 = 686128968.$$

An equivalent way of comparing words with the same slope but different length is to associate the vector

$$\mathcal{J}(w) := (\mathcal{J}_1(w), \dots, \mathcal{J}_{|w|}(w))$$

to  $w$ , where

$$\mathcal{J}_i(w) := \frac{\mathcal{I}_i(w)}{2^{|w|} - 1} \quad \text{for } 1 \leq i \leq |w|,$$

and consider the corresponding *geometric mean*

$$GM(w) := \left( \prod_{i=1}^{|w|} \mathcal{J}_i(w) \right)^{1/|w|} = \frac{\left( \prod_{i=1}^{|w|} \mathcal{I}_i(w) \right)^{1/|w|}}{2^{|w|} - 1}.$$

A short calculation shows that the slope of a word  $w$ , a priori a purely symbolic notion, is actually equal to the *arithmetic mean* of the  $\mathcal{J}_i(w)$ :

$$sl(w) = \frac{1}{|w|} \sum_{i=1}^{|w|} \mathcal{J}_i(w) =: AM(w).$$

So Theorem 1.4 can be interpreted as relating arithmetic and geometric means:

**Theorem 1.6.** *For all orbits of a given arithmetic mean, the geometric mean is maximized precisely when the orbit is balanced.*

**Example 1.7.** The words 01, 0011, 000111, 001011, and 001101 all have slope  $1/2$ , and the corresponding geometric means are

$$GM(01) = \frac{(1 \times 2)^{1/2}}{3} = \frac{\sqrt{2}}{3} \approx 0.471,$$

$$GM(0011) = \frac{(3 \times 6 \times 9 \times 12)^{1/4}}{15} \approx 0.442,$$

$$\begin{aligned}
GM(000111) &= \frac{(7 \times 14 \times 28 \times 35 \times 49 \times 56)^{1/6}}{63} \approx 0.401, \\
GM(001011) &= \frac{(11 \times 22 \times 25 \times 37 \times 44 \times 50)^{1/6}}{63} \approx 0.446, \\
GM(001101) &= \frac{(13 \times 19 \times 26 \times 38 \times 41 \times 52)^{1/6}}{63} \approx 0.450.
\end{aligned}$$

Another quantity of interest is the *standard deviation* of an orbit around its arithmetic mean, defined as

$$SD(w) = \sqrt{\frac{1}{|w|} \sum_{i=1}^{|w|} (\mathcal{J}_i(w) - AM(w))^2}. \quad (2)$$

It turns out that standard deviation is *minimized* by balanced orbits:

**Theorem 1.8.** *For all orbits of a given slope, the standard deviation is minimized precisely when the orbit is balanced.*

**Example 1.9.** The unique balanced orbit of slope  $1/2$  is  $01$ , with standard deviation

$$SD(01) = 1/6 \approx 0.166,$$

while standard deviations of some other orbits with slope  $1/2$  are:

$$SD(0011) = (20)^{-1/2} \approx 0.223,$$

$$SD(000111) = 5/18 \approx 0.277,$$

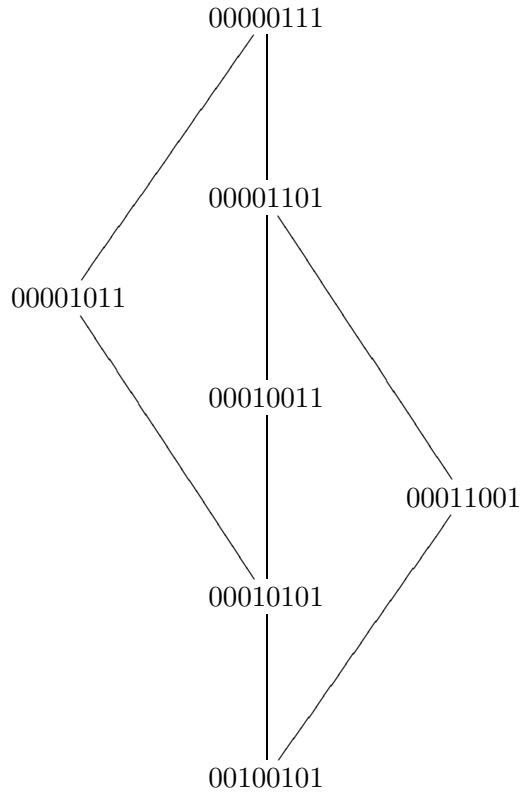
$$SD(001011) = SD(001101) = (103/7)^{1/2}/18 \approx 0.213.$$

Although it had previously been known that balanced words are characterised by various “extremal” properties<sup>4</sup>, the above results in terms of base-2 expansions are of a rather different nature. In fact they are all consequences of a more general result, Theorem 2.4, which is phrased in terms of *majorization*, and asserts that each  $\mathbb{W}_{p,q}$  is a poset whose minimal element is the balanced orbit. Majorization is a notion which is pervasive in various branches of mathematics<sup>5</sup>, and can itself be formulated in a number of different ways, each formulation leading to a different characterisation of balanced words which will be described in this paper. In order, these characterisations are in terms of: partial sums of the  $\mathcal{I}_i(w)$  (§2), symmetric mean value inequalities (§3), doubly stochastic matrices (§4), and transfers (§5).

<sup>4</sup>Notably, their *subword complexity* is as small as possible: a length- $q$  word has at most  $n+1$  length- $n$  cyclic subwords for each  $1 \leq n \leq q-1$  if and only if it is balanced.

<sup>5</sup>The monograph [27] by Marshall & Olkin describes many applications of majorization, e.g. to inequalities relating eigenvalues and singular values of linear operators [5, Ch. II], [15, Ch. VI], [27, Ch. 9], and to aspects of graph theory [27, Ch. 7]. Majorization is applied extensively in probability and statistics (see e.g. [27, Ch. 11–13], [36], [37, Ch. 6]), for example in comparison of experiments [6, 7], sampling theory [27, Ch. 12.A], hypothesis testing [27, Ch. 13.A], and reliability theory [27, Ch. 13.D], [36]. It is used as a measurement of inequality and diversity in a variety of contexts, for example in economics, to describe income distribution (if  $\mu \prec \nu$  then the distribution  $\nu$  is more equitable, see [11, 12]).

## 2. MAJORIZATION AND PARTIAL SUM INEQUALITIES

FIGURE 1. Hasse diagram for the poset  $(\mathbb{W}_{3,8}, \prec)$ 

The partial order of majorization (see e.g. [27] for a general reference) is defined as follows:

**Definition 2.1.** Given real numbers  $x_1 \leq \dots \leq x_Q$  and  $y_1 \leq \dots \leq y_Q$ , the vector  $y = (y_1, \dots, y_Q)$  is said to *majorize*  $x = (x_1, \dots, x_Q)$ , denoted  $x \prec y$ , if

$$\sum_{k=1}^i x_k \geq \sum_{k=1}^i y_k \quad \text{for } 1 \leq i \leq Q-1, \quad (3)$$

and

$$\sum_{k=1}^Q x_k = \sum_{k=1}^Q y_k. \quad (4)$$

We shall be interested in restricting majorization to vectors of the form  $\mathcal{I}(w)$ , for words  $w \in \{0, 1\}^+$ . We write  $w \prec w'$  (respectively  $\mathcal{O}(w) \prec \mathcal{O}(w')$ ), and say that  $w'$  majorizes  $w$  (respectively  $\mathcal{O}(w')$  majorizes  $\mathcal{O}(w)$ ), whenever  $\mathcal{I}(w) \prec \mathcal{I}(w')$ . This defines a partial order on the set of orbits of words in  $\{0, 1\}^+$ . Necessary conditions for two orbits to be comparable are that they have the same length and the same

1-length: the first condition is obvious, while the second follows from (4), because if  $w \in \{0, 1\}^Q$  then  $\sum_{i=1}^Q \mathcal{I}_i(w) = (2^Q - 1)|w|_1$ . In particular, each  $(\mathbb{W}_{p,q}, \prec)$  is a poset. For example  $(\mathbb{W}_{3,8}, \prec)$  is represented by the Hasse diagram in Figure 1.

**Example 2.2.** The computations below prove that the relations in  $(\mathbb{W}_{3,8}, \prec)$  are as in Figure 1; here the partial sums are denoted by

$$\mathcal{S}_i(w) := \sum_{k=1}^i \mathcal{I}_k(w).$$

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
00100101	37	37	00011001	25	25	00010101	21	21
00101001	41	78	00101001	35	60	00101010	42	63
01001001	73	151	00110010	50	110	01000101	69	132
01001010	74	225	01000110	70	180	01010001	81	213
01010010	82	307	01100100	100	280	01010100	84	297
10010010	146	453	10001100	140	420	10001010	138	435
10010100	148	601	10010001	145	565	10100010	162	597
10100100	164	765	11001000	200	765	10101000	168	765

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
00010011	19	19
00100110	38	57
00110001	49	106
01001100	76	182
01100010	98	280
10001001	137	417
10011000	152	569
11000100	196	765

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
00001101	13	13
00011010	26	39
00110100	52	91
01000011	67	158
01100001	104	262
10000101	134	396
10110000	161	557
11000010	208	765

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
00001011	11	11
00010110	22	33
00101100	44	77
01011000	88	165
01100001	97	262
10000101	133	395
10110000	176	571
11000010	194	765

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
00000111	7	7
00001110	14	21
00011100	28	49
00111000	56	105
01110000	112	217
10000011	131	348
11000001	193	541
11100000	224	765

In particular, the balanced orbit 00100101 is seen to be the least element in  $(\mathbb{W}_{3,8}, \prec)$ .

The extremal property of balanced orbits seen in Example 2.2 is quite general:

**Theorem 2.3.** *For any coprime integers  $1 \leq p < q$ , the unique balanced orbit  $b \in \mathbb{W}_{p,q}$  is the least element in  $(\mathbb{W}_{p,q}, \prec)$ . In other words, for any  $w \in \mathbb{W}_{p,q}$ ,*

$$\mathcal{S}_i(b) \geq \mathcal{S}_i(w) \quad \text{for all } 1 \leq i \leq q. \quad (5)$$

As in §1, it turns out that balanced orbits are actually extremal among all orbits of a given slope:

**Theorem 2.4.** *Suppose  $w \in \{0,1\}^{\mathbb{Q}}$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers. If  $b \in \mathbb{W}_{p,q}$  is balanced then  $b^{Q/q} \prec w$ . In other words,*

$$\mathcal{S}_i(b^{Q/q}) \geq \mathcal{S}_i(w) \quad \text{for all } 1 \leq i \leq Q. \quad (6)$$

*Proof.* The majorization  $b^{Q/q} \prec w$  means precisely that  $\mathcal{I}(b^{Q/q}) \prec \mathcal{I}(w)$ , and by a classical result of Hardy, Littlewood & Pólya (see [17, 18]) this is equivalent to

$$\sum_{i=1}^Q f(\mathcal{I}_i(b^{Q/q})) \leq \sum_{i=1}^Q f(\mathcal{I}_i(w)) \quad \text{for all convex functions } f : \mathbb{R} \rightarrow \mathbb{R}.$$

Since  $\mathcal{J}_i(w) = \mathcal{I}_i(w)/(2^{|w|} - 1)$ , this is equivalent to

$$\sum_{i=1}^Q f(\mathcal{J}_i(b^{Q/q})) \leq \sum_{i=1}^Q f(\mathcal{J}_i(w)) \quad \text{for all convex functions } f : (0,1) \rightarrow \mathbb{R}, \quad (7)$$

hence equivalent to

$$\frac{1}{q} \sum_{i=1}^q f(\mathcal{J}_i(b)) \leq \frac{1}{Q} \sum_{i=1}^Q f(\mathcal{J}_i(w)) \quad (8)$$

for all convex functions  $f : (0,1) \rightarrow \mathbb{R}$ , and by a simple approximation argument we may take  $f$  to be smooth.

Now, for an arbitrary smooth convex function  $f : (0,1) \rightarrow \mathbb{R}$ , we shall establish (8). If we define  $a(x) = (x+1)/2$  for  $x \in (0, 2\mathcal{J}_1(b)]$ , and  $a(x) = x/2$  for  $x \in (2\mathcal{J}_1(b), 1)$ , and set

$$g(x) := f(x) - x \int_0^1 \sum_{n \geq 0} \frac{f'(a^n(t))}{2^n} dt,$$

then for  $w \in \{0,1\}^{\mathbb{Q}}$  with slope  $p/q$ , the inequality (8) is clearly equivalent to

$$\frac{1}{q} \sum_{i=1}^q g(\mathcal{J}_i(b)) \leq \frac{1}{Q} \sum_{i=1}^Q g(\mathcal{J}_i(w)). \quad (9)$$

In fact we claim that (9) holds for *any*  $w \in \{0,1\}^+$ . To prove this, let  $h$  be the indefinite integral (unique up to an additive constant) of the bounded function  $t \mapsto \sum_{n \geq 0} g'(a^n(t))/2^n$ . The function  $h$  is Lipschitz, and if  $G := g + h - h \circ T$ , where  $T(x) := 2x \pmod{1}$ , then  $G' = 0$  Lebesgue almost everywhere on  $[\mathcal{J}_1(b), \mathcal{J}_1(b) + 1/2]$ . Therefore  $G$  is constant on  $[\mathcal{J}_1(b), \mathcal{J}_1(b) + 1/2]$ , and this constant,  $c$  say,

is its minimum value (cf. [22, 23]). It follows that  $\frac{1}{q} \sum_{i=1}^q G(\mathcal{J}_i(b)) = c$ , because  $\mathcal{J}_q(b) - \mathcal{J}_1(b) < 1/2$  (see [9, Prop. 1]), and hence that

$$\frac{1}{q} \sum_{i=1}^q G(\mathcal{J}_i(b)) = c \leq \frac{1}{Q} \sum_{i=1}^Q G(\mathcal{J}_i(w)). \quad (10)$$

However, clearly  $\frac{1}{|v|} \sum_{i=1}^{|v|} G(\mathcal{J}_i(v)) = \frac{1}{|v|} \sum_{i=1}^{|v|} g(\mathcal{J}_i(v))$  for all  $v \in \{0, 1\}^+$ , so (10) implies that (9) holds for all  $w \in \{0, 1\}^+$ , as required.  $\square$

**Example 2.5.** In the case of slope  $1/2$ , the calculations below show that  $01010101$  is majorized by both  $00110011$  and  $00101101$ , but that  $00110011$  and  $00101101$  are unrelated.

$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{O}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
01010101	85	85	00110011	51	51	00101101	45	45
01010101	85	170	00110011	51	102	01001011	75	120
01010101	85	255	01100110	102	204	01011010	90	210
01010101	85	340	01100110	102	306	01101001	105	315
10101010	170	510	10011001	153	459	10010110	150	465
10101010	170	680	10011001	153	612	10100101	165	630
10101010	170	850	11001100	204	816	10110100	180	810
10101010	170	1020	11001100	204	1020	11010010	210	1020

**Corollary 2.6.** *Suppose  $w \in \{0, 1\}^Q$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers, that  $w$  is not balanced, and that  $b \in \mathbb{W}_{p,q}$  is balanced. Then*

$$\mathcal{O}_1(w) < \mathcal{O}_1(b^{Q/q}) \quad \text{and} \quad \mathcal{O}_Q(w) > \mathcal{O}_Q(b^{Q/q}).$$

*In other words, the lexicographically smallest component of  $\mathcal{O}(w)$  is strictly smaller, in the lexicographic order, than the lexicographically smallest component of  $\mathcal{O}(b^{Q/q})$ , and the lexicographically largest component of  $\mathcal{O}(w)$  is strictly larger, in the lexicographic order, than the lexicographically largest component of  $\mathcal{O}(b^{Q/q})$ .*

*Proof.* Since  $w$  is not balanced,  $\mathcal{O}_i(w) \neq \mathcal{O}_i(b^{Q/q})$ , hence  $\mathcal{I}_i(w) \neq \mathcal{I}_i(b^{Q/q})$  for all  $1 \leq i \leq Q$ . In particular,

$$\mathcal{I}_1(w) \neq \mathcal{I}_1(b^{Q/q}) \quad \text{and} \quad \mathcal{I}_Q(w) \neq \mathcal{I}_Q(b^{Q/q}). \quad (11)$$

Setting  $i = 1$  and  $i = Q - 1$  in (6) gives

$$\mathcal{I}_1(w) \leq \mathcal{I}_1(b^{Q/q}) \quad \text{and} \quad \mathcal{I}_Q(w) \geq \mathcal{I}_Q(b^{Q/q}),$$

so by (11) these inequalities are *strict*. This corresponds to strict inequality, in the lexicographic order, between the corresponding words, as required.  $\square$

**Remark 2.7.** Corollary 2.6 was originally proved, via entirely different methods, by Bernhardt [3]; an alternative proof is due to Gambaudo, Lanford & Tresser [14]. In all other cases  $2 \leq i \leq Q - 2$ , the inequality (6) is new.

In fact it should be possible to replace (6) with *strict* inequalities (provided  $i \neq Q$ ):

**Conjecture 2.8.** Under the hypotheses of Corollary 2.6,

$$\mathcal{S}_i(b^{Q/q}) > \mathcal{S}_i(w) \quad \text{for all } 1 \leq i \leq Q - 1.$$

**Remark 2.9.** In general it is possible for partial sums of distinct  $v$  and  $w$  to be equal: for example we saw above that if  $v = 00001101$  and  $w = 00001011$  then  $S_5(v) = 262 = S_5(w)$ .

We can now prove the theorems which were stated in §1:

*Proof of Theorems 1.2, 1.4, 1.6, 1.8.* Suppose the non-balanced word  $w \in \{0, 1\}^Q$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers, and that  $b \in \mathbb{W}_{p,q}$  is balanced. The weak inequality (7) implies the strict inequality

$$\sum_{i=1}^Q f(\mathcal{J}_i(b^{Q/q})) < \sum_{i=1}^Q f(\mathcal{J}_i(w)).$$

whenever  $f : [0, 1] \rightarrow \mathbb{R}$  is *strictly* convex. In particular, the function  $f(x) = (x - p/q)^2$  is strictly convex, so Theorem 1.8 follows.

If, on the other hand,  $f : [0, 1] \rightarrow \mathbb{R}$  is strictly *concave*, then

$$\sum_{i=1}^Q f(\mathcal{J}_i(b^{Q/q})) > \sum_{i=1}^Q f(\mathcal{J}_i(w)).$$

In particular  $f(x) = \log x$  is strictly concave, so

$$\sum_{i=1}^Q \log \mathcal{J}_i(b^{Q/q}) > \sum_{i=1}^Q \log \mathcal{J}_i(w),$$

in other words

$$\prod_{i=1}^Q \mathcal{J}_i(b^{Q/q}) > \prod_{i=1}^Q \mathcal{J}_i(w).$$

From this we deduce Theorems 1.2 and 1.4, that the product  $P(w) = \prod_{i=1}^Q \mathcal{I}_i(w)$  is maximized precisely when  $w$  is balanced, as well as the formulation (Theorem 1.6) in terms of arithmetic and geometric means.  $\square$

### 3. SYMMETRIC MEANS AND SCHUR-CONVEXITY

The work of Hardy, Littlewood & Pólya [17, 18] on majorization was inspired by a theorem of Muirhead [30] on comparisons of so-called symmetric means. Before recalling the general form of Muirhead's theorem, we first state its consequences in our setting, when combined with Theorem 2.4. Let  $\mathcal{P}_m$  denote the set of permutations of  $\{1, \dots, m\}$ .

**Corollary 3.1.** *Suppose that  $1 \leq p < q$  are coprime integers, and let  $b$  denote the unique balanced element in  $\mathbb{W}_{p,q}$ . If  $w$  is any other element of  $\mathbb{W}_{p,q}$  then*

$$\sum_{\pi \in \mathcal{P}_q} \prod_{i=1}^q a_{\pi(i)}^{I_i(b)} \leq \sum_{\pi \in \mathcal{P}_q} \prod_{i=1}^q a_{\pi(i)}^{I_i(w)} \quad (12)$$

for all  $(a_1, \dots, a_q) \in \mathbb{R}_+^q$ , with equality if and only if  $a_1 = \dots = a_q$ .

**Example 3.2.** The balanced orbit in  $\mathbb{W}_{2,5}$  is  $b = 00101$ , with  $\mathcal{I}(b) = (5, 9, 10, 18, 20)$ , and the other member of  $\mathbb{W}_{2,5}$  is  $w = 00011$ , with  $\mathcal{I}(w) = (3, 6, 12, 17, 24)$ . Corollary 3.1 implies that

$$\sum_{\pi \in \mathcal{P}_5} a_{\pi(1)}^5 a_{\pi(2)}^9 a_{\pi(3)}^{10} a_{\pi(4)}^{18} a_{\pi(5)}^{20} \leq \sum_{\pi \in \mathcal{P}_5} a_{\pi(1)}^3 a_{\pi(2)}^6 a_{\pi(3)}^{12} a_{\pi(4)}^{17} a_{\pi(5)}^{24},$$

for all  $(a_1, \dots, a_5) \in \mathbb{R}_+^5$ , with equality if and only if  $a_1 = \dots = a_5$ .

As usual, there is a more general version which applies to all words  $w \in \{0, 1\}^+$ :

**Corollary 3.3.** *Suppose  $w \in \{0, 1\}^Q$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers. If  $b \in \mathbb{W}_{p,q}$  is balanced then*

$$\sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{I_i(b^{Q/q})} \leq \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{I_i(w)} \quad (13)$$

for all  $(a_1, \dots, a_Q) \in \mathbb{R}_+^Q$ , with equality if and only if  $a_1 = \dots = a_Q$ .

**Example 3.4.** 0011 has slope  $1/2$ , with  $\mathcal{I}(w) = (3, 6, 9, 12)$ , and the balanced orbit in  $\mathbb{W}_{1,2}$  is  $b = 01$ , with  $\mathcal{I}(b^2) = I(0101) = (5, 5, 10, 10)$ . So Corollary 3.3 implies that

$$\sum_{\pi \in \mathcal{P}_4} a_{\pi(1)}^5 a_{\pi(2)}^5 a_{\pi(3)}^{10} a_{\pi(4)}^{10} \leq \sum_{\pi \in \mathcal{P}_4} a_{\pi(1)}^3 a_{\pi(2)}^6 a_{\pi(3)}^9 a_{\pi(4)}^{12}, \quad (14)$$

for all  $(a_1, a_2, a_3, a_4) \in \mathbb{R}_+^4$ , with equality if and only if  $a_1 = \dots = a_4$ .

The general form of Muirhead's theorem (see [30], or [27, Ch. 3.G]) asserts that if  $(x_1, \dots, x_Q) \prec (y_1, \dots, y_Q)$ , then

$$\sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{x_i} \leq \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{y_i}$$

for all  $(a_1, \dots, a_Q) \in \mathbb{R}_+^Q$ , with equality if and only if  $a_1 = \dots = a_Q$ . A quantity of the form

$$\frac{1}{Q!} \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{x_i} \quad (15)$$

(note the normalisation factor  $1/Q!$  in (15)) is known as a *symmetric mean*<sup>6</sup> (symmetric in the variables  $a_1, \dots, a_Q$ ). The best known examples of symmetric means are the arithmetic and geometric means: if  $(x_1, \dots, x_Q) = (0, \dots, 0, 1)$  then

$$\frac{1}{Q!} \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{x_i} = \frac{a_1 + \dots + a_Q}{Q}, \quad (16)$$

---

<sup>6</sup>Sometimes this terminology is reserved for the case when  $x_1 + \dots + x_Q = 1$ .

while if  $(x_1, \dots, x_Q) = (Q^{-1}, \dots, Q^{-1})$  then

$$\frac{1}{Q!} \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{x_i} = (a_1 \cdots a_Q)^{1/Q}. \quad (17)$$

The classical *arithmetic-geometric mean inequality* asserts that (16) is larger than (17) for all  $(a_1, \dots, a_Q) \in \mathbb{R}_+^q$ , with equality if and only if  $a_1 = \dots = a_Q$ . So Muirhead's theorem can be seen as a generalisation of this well known inequality (note that clearly  $(1/Q, \dots, 1/Q) \prec (0, \dots, 0, 1)$ ).

In fact Muirhead's theorem can nowadays be viewed as a special consequence of *Schur-convexity* (named after the pioneering work of Schur [33]). A real-valued function  $\varphi$  defined on a subset  $\mathcal{A}$  of  $\mathbb{R}^Q$  is called *Schur-convex* if  $\varphi(x) \leq \varphi(y)$  whenever  $x, y \in \mathcal{A}$  are such that  $x \prec y$ . In other words, the Schur-convex functions are precisely those which are isotonic with respect to the partial order of majorization. One of the simplest constructions of a Schur-convex function  $\varphi$  (on  $[0, 1]^Q$ , say) is to choose any convex function  $f : [0, 1] \rightarrow \mathbb{R}$ , and define  $\varphi(x_1, \dots, x_Q) = \sum_{i=1}^Q f(x_i)$  (cf. the Hardy-Littlewood-Pólya theorem mentioned in §2).

Another well-known sufficient condition for a function  $\varphi$  to be Schur-convex is that it be symmetric and convex (see e.g. [27, Prop. 3.C.2]). Indeed for such a  $\varphi$ , if  $t_1, \dots, t_Q \in \mathbb{R}$  then the function

$$(x_1, \dots, x_Q) \mapsto \sum_{\pi \in \mathcal{P}_Q} \varphi(t_{\pi(1)}x_1, \dots, t_{\pi(Q)}x_Q) \quad (18)$$

is also Schur-convex [27, p. 86], and the particular choices

$$t_i = \log a_i \quad \text{and} \quad \varphi(z_1, \dots, z_Q) = \exp \sum_{i=1}^Q z_i$$

yield

$$\sum_{\pi \in \mathcal{P}_Q} \varphi(t_{\pi(1)}x_1, \dots, t_{\pi(Q)}x_Q) = \sum_{\pi \in \mathcal{P}_Q} \prod_{i=1}^Q a_{\pi(i)}^{x_i},$$

which is precisely the quantity treated in Muirhead's theorem.

In order to generalise Corollaries 3.1 and 3.3, our Theorem 2.3 may be combined with various functions which are known to be Schur-convex (see e.g. [27] for examples of such functions) in order to derive new characterisations of balanced words.

#### 4. DOUBLY STOCHASTIC AVERAGES

A square matrix with non-negative entries is called *doubly stochastic* if each of its row and column sums is equal to one. Given vectors  $x, y \in \mathbb{R}^Q$ , we say that  $x$  is a *doubly stochastic average* of  $y$  if  $x = Ay$  for some doubly stochastic matrix  $A$ .

Given  $v, w \in \{0, 1\}^+$ , we say that  $v$  (respectively  $\mathcal{O}(v)$ ) is a *doubly stochastic average* of  $w$  (respectively  $\mathcal{O}(w)$ ) if  $\mathcal{I}(v^k)$  is a doubly stochastic average of  $\mathcal{I}(w^l)$  for some  $k, l \geq 1$ .

If  $x$  is a doubly stochastic average of  $y$ , with doubly stochastic matrix  $A$  such that  $x_i = \sum_j A(i, j)y_j$  for each  $i$ , then

$$\begin{aligned} \sum_i f(x_i) &= \sum_i f\left(\sum_j A(i, j)y_j\right) \\ &\leq \sum_i \sum_j A(i, j)f(y_j) \\ &= \sum_j f(y_j) \sum_i A(i, j) = \sum_j f(y_j), \end{aligned} \tag{19}$$

whenever  $f$  is convex, by Jensen's inequality (which is applicable because  $\sum_j A(i, j) = 1$  for each  $i$ ). This simple argument, which seems to have first been articulated by Schur [33], implies that  $x \prec y$  whenever  $x$  is a doubly stochastic average of  $y$ , by the Hardy-Littlewood-Pólya characterisation of majorization used in the proof of Theorem 2.4. In fact Hardy, Littlewood & Pólya [17, 18] (see also [27, 28, 35]) proved that the converse is true as well, so by Theorem 2.3 we have:

**Corollary 4.1.** *Suppose  $w \in \{0, 1\}^+$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers. The unique balanced orbit in  $\mathbb{W}_{p,q}$  is a doubly stochastic average of  $w$ .*

**Example 4.2.** The unique balanced orbit of slope  $1/2$  is  $b = 01$ . Now  $0011$  has slope  $1/2$ , with  $\mathcal{I}(0011) = (3, 6, 9, 12)$ , and  $\mathcal{I}(b^2) = \mathcal{I}(0101) = (5, 5, 10, 10)$ . Since

$$\begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 0 & 1/3 & 0 & 2/3 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \\ 9 \\ 12 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \\ 10 \\ 10 \end{pmatrix},$$

the balanced orbit  $01$  is a doubly stochastic average of  $0011$ .

The word  $000111$  also has slope  $1/2$ , with

$$\mathcal{I}(000111) = (7, 14, 28, 35, 49, 56),$$

and

$$\mathcal{I}(b^3) = \mathcal{I}(010101) = (21, 21, 21, 42, 42, 42).$$

Since

$$\frac{1}{7} \begin{pmatrix} 3 & 1 & 1 & 1 & 1 & 0 \\ 3 & 1 & 2 & 0 & 0 & 1 \\ 1 & 3 & 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 & 2 & 2 \\ 0 & 0 & 3 & 0 & 2 & 2 \\ 0 & 1 & 0 & 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} 7 \\ 14 \\ 28 \\ 35 \\ 49 \\ 56 \end{pmatrix} = \begin{pmatrix} 21 \\ 21 \\ 21 \\ 42 \\ 42 \\ 42 \end{pmatrix},$$

the balanced orbit  $01$  is a doubly stochastic average of  $000111$ .

## 5. TRANSFERS

If  $x \prec y$  then the doubly stochastic matrix  $A$  satisfying  $x = Ay$  is in general not unique; indeed little is known about the set of doubly stochastic matrices realising a given majorization (see [27, p. 40]). One means of constructing  $A$  is provided by the following notion:

**Definition 5.1.** Given  $x \in \mathbb{Z}_+^Q$ , a *transfer* is a linear transformation  $T : \mathbb{R}^Q \rightarrow \mathbb{R}^Q$  such that, for some  $1 \leq i, j \leq Q$  with  $x_j > x_i$ , the vector  $x' := T(x)$  satisfies  $x'_i = x_i + 1$ ,  $x'_j = x_j - 1$ , and  $x'_k = x_k$  for  $k \neq i, j$ .

In other words, a transfer *evens up* the components of a vector, by simultaneously decreasing by 1 the value of a “large” component and increasing by 1 a “small” component<sup>7</sup>; in [35] they are described as *Robin Hood transformations*. Transfers were employed by Muirhead [30] in the proof of his theorem (see §3), as well as by Hardy, Littlewood & Pólya [17, 18], and Folkman & Fulkerson [13]. Transfers are also of interest in economics, in the theory of income distribution (see e.g. [11]), their consideration in this context dating back to the work of Dalton [12] (see also [26, 31]).

Explicitly, the linear transformation effecting the transfer in Definition 5.1 is given by the  $Q \times Q$  (doubly stochastic) matrix  $A$  defined by

$$A(i, i) = A(j, j) = 1 - \frac{1}{x_j - x_i},$$

$$A(i, j) = A(j, i) = \frac{1}{x_j - x_i},$$

and

$$A(k, l) = \delta_{kl} \quad \text{for } k, l \in \{1, \dots, Q\} \setminus \{i, j\}.$$

Muirhead [30] (see also [27, p. 135]) proved that if  $x \prec y$  then  $x$  can be obtained from  $y$  by a (finite) number of transfers<sup>8</sup>. Therefore:

**Corollary 5.2.** *Let  $b$  denote the unique balanced orbit in  $\mathbb{W}_{p,q}$ , where  $1 \leq p < q$  are coprime integers. For any  $w \in \mathbb{W}_{p,q}$ , the base-2 orbit  $\mathcal{I}(b)$  can be derived from  $\mathcal{I}(w)$  by a finite number of transfers.*

**Example 5.3.** the members of  $\mathbb{W}_{2,5}$  are  $w = 00011$  and  $b = 00101$ . The vector  $\mathcal{I}(b) = (5, 9, 10, 18, 20)$  can be derived from  $\mathcal{I}(w) = (3, 6, 12, 17, 24)$  by applying six transfers, for example as follows:

$$\begin{aligned} (3, 6, 12, 17, 24) &\rightarrow (4, 6, 12, 17, 23) \rightarrow (5, 6, 12, 17, 22) \\ &\rightarrow (5, 7, 12, 17, 21) \rightarrow (5, 7, 12, 18, 20) \\ &\rightarrow (5, 8, 11, 18, 20) \rightarrow (5, 9, 10, 18, 20). \end{aligned}$$

<sup>7</sup>This notion of transfer can obviously be generalised: the amount transferred between the two components need not equal 1, and the vectors involved could have real rather than integer components.

<sup>8</sup>Clearly the converse also holds: a transfer is given by a doubly stochastic matrix, and a product of doubly stochastic matrices is again doubly stochastic, so Schur’s calculation (19), together with the Hardy-Littlewood-Pólya criterion of §2, implies the majorization.

The doubly stochastic matrices representing these transfers are, in order,

$$\begin{pmatrix} \frac{20}{21} & 0 & 0 & 0 & \frac{1}{21} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{21} & 0 & 0 & 0 & \frac{20}{21} \end{pmatrix}, \begin{pmatrix} \frac{18}{19} & 0 & 0 & 0 & \frac{1}{19} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{19} & 0 & 0 & 0 & \frac{18}{19} \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{15}{16} & 0 & 0 & \frac{1}{16} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{16} & 0 & 0 & \frac{15}{16} \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1/4 & 3/4 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4/5 & 1/5 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and their product (in reverse order) is

$$\begin{pmatrix} 19/21 & 0 & 0 & 0 & 2/21 \\ 1/280 & 9/16 & 2/5 & 0 & 19/560 \\ 1/420 & 3/8 & 3/5 & 0 & 19/840 \\ 5/224 & 1/64 & 0 & 3/4 & 95/448 \\ 15/224 & 3/64 & 0 & 1/4 & 285/448 \end{pmatrix},$$

a doubly stochastic matrix which transforms  $\mathcal{I}(w)$  to  $\mathcal{I}(b)$ .

Analogously, any orbit can be transformed, via a sequence of transfers, into the balanced orbit of the same slope:

**Corollary 5.4.** *Suppose  $w \in \{0, 1\}^{\mathbb{Q}}$  has slope  $p/q$ , where  $1 \leq p < q$  are coprime integers. If  $b \in \mathbb{W}_{p,q}$  is balanced then  $\mathcal{I}(b^{\mathbb{Q}/q})$  can be derived from  $\mathcal{I}(w)$  by a finite number of transfers.*

**Remark 5.5.** It would be useful to identify a class of operations  $w \mapsto w'$  on words (or the corresponding orbits) which guarantee that  $w' \prec w$ . Note that transfers are *not* of this type: if  $v \prec w$  we may derive  $\mathcal{I}(v)$  from  $\mathcal{I}(w)$  via a finite number of transfers, but the *intermediate* vectors in general have no symbolic interpretation, i.e. they are not the base-2 orbit  $\mathcal{I}(u)$  for any word  $u$ .

A natural idea is to define an operation which exchanges the position of some (cyclically) adjacent symbols 0 and 1. For example 0011 can be balanced by exchanging the middle letters, to get 0101, while 00011 can be balanced by swapping the 3rd and 4th letters, obtaining 00101. The word 000111 can be transformed to 001011 by swapping the middle two letters (note that 001011  $\prec$  000111); then we swap first and last letters to obtain 101010, which is balanced. Of course these balancing operations can be reversed, in which case their effect is to *unbalance* rather than balance; therefore criteria for recognising *which* adjacent symbols should be swapped are necessary. Moreover, if this notion is to bear fruit then some refinement is necessary, since there exist unbalanced words which cannot be transformed to the balanced word via a sequence of such operations. For example if  $w = 00101101$  then each of the six possible exchanges of adjacent 0 and 1 yields a word which either majorizes  $w$ , or is incomparable to  $w$ .

## 6. BALANCED SEQUENCES

If the word  $w \in \{0, 1\}^+$  is balanced, the (infinite) 0-1 sequence

$$w^\infty = w \cdot w \cdot w \cdots$$

is a particular example of a *balanced sequence*, i.e. a 0-1 sequence such that  $||z|_1 - |z'|_1| \leq 1$  for any two finite subwords  $z, z'$  with  $|z| = |z'|$  (see e.g. [1, Chs. 9, 10], [4] or [32, Ch. 6]). If  $\omega = (\omega_i)_{i=1}^\infty$  is any balanced sequence then its *slope*  $\lim_{n \rightarrow \infty} |\omega_1 \dots \omega_n|_1 / n$  is always defined. The sequence  $w^\infty$  has rational slope, but there exist balanced sequences with irrational slope, often called *Sturmian sequences*.<sup>9</sup> Indeed if  $\rho \in (0, 1)$  is irrational then there are uncountably many balanced sequences with slope  $\rho$ .

Although the results of this paper have natural analogues in the case of irrational slope, in order to make uniqueness statements it is more convenient to formulate these results in terms of balanced *measures*. For  $\rho \in [0, 1]$ , the *balanced measure of slope  $\rho$*  is a probability measure on the set of all 0-1 sequences, invariant under the left shift map, and giving full measure to the set of all balanced sequences of slope  $\rho$ . We will not elaborate on all of the irrational slope versions of the results in this paper (though the interested reader should be able to re-construct them in conjunction with [23]), instead restricting attention to the following analogue of Theorem 1.8 (here by the slope of a general shift-invariant probability measure we mean the measure of the set of all 0-1 sequences  $\omega$  whose first symbol  $\omega_1$  equals 1):

**Theorem 6.1.** *For each  $\rho \in [0, 1]$ , the balanced measure of slope  $\rho$  has smaller standard deviation than any other shift-invariant probability measure of slope  $\rho$ .*

The standard deviation of the balanced measure of slope  $\rho$  varies continuously with  $\rho$ ; this is depicted in Figure 2 in the range  $\rho \in [1/5, 1/2]$ .

Clearly the smallest possible standard deviation of a balanced measure is 0: this occurs for the balanced measures of slopes 0 and 1, which are both concentrated on a single point. An open problem, however, is to determine the balanced measure<sup>10</sup> with *largest* standard deviation.

---

<sup>9</sup>The terminology is unfortunately a little confused. Morse & Hedlund [29] considered what they called *Sturmian trajectories* (in our setting the term *trajectory* means a 0-1 sequence), and these correspond precisely to what we are calling *balanced sequences*. So, following [29], it may seem more natural to talk of *Sturmian sequences* instead of balanced sequences, and indeed many authors adopt this convention (see e.g. [8, 9, 10, 21, 25, 34]). However, many other authors reserve the term *Sturmian sequence* to denote a balanced sequence with irrational slope. The reason for this seems to stem from an emphasis on the irrational case, both initially (Morse & Hedlund were interested in examples of recurrent behaviour which was not periodic, see also [20]), but also subsequently (various properties, for example the characterisation in terms of smallest subword complexity, are more conveniently formulated in the irrational case). Our use of the term *balanced word* throughout this article is consistent with [24], and differs from the notion of a *Sturmian word* (see [24] for further discussion).

<sup>10</sup>Of course there are (at least) *two* balanced measures with largest standard deviation, by symmetry: the balanced measures of slope  $\rho$  and  $1 - \rho$  have equal standard deviation. For this reason we shall henceforth restrict to  $\rho \in [0, 1/2]$ .

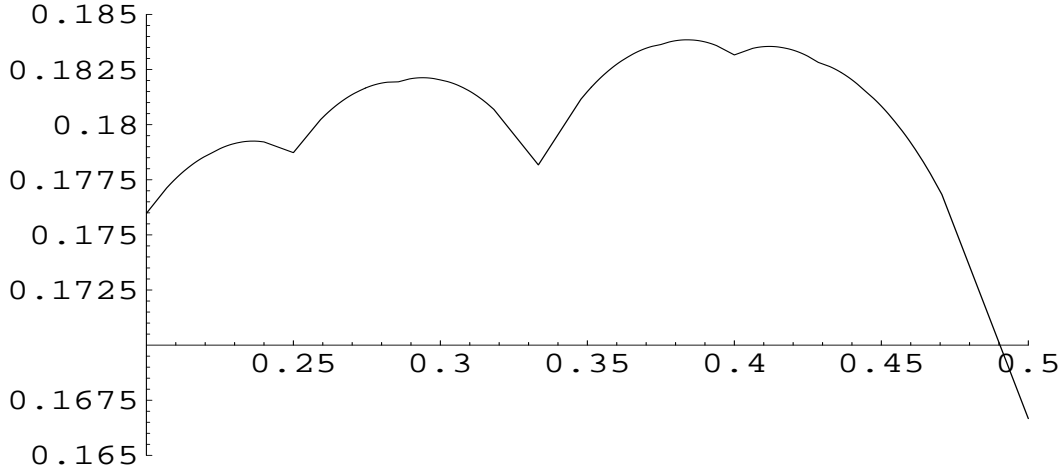


FIGURE 2. Standard deviation of balanced measure of slope  $\varrho$ , for  $\varrho \in [1/5, 1/2]$

**Problem 6.2.** Which balanced measure has largest standard deviation around its mean?

We have carried out numerical computations related to this problem. Namely, for a large number of rational values of  $\varrho$  we have computed the standard deviation (2) of the balanced measure (in this case an orbit) of slope  $\varrho$ . The rational values in question correspond to higher levels of the Farey tree (see e.g. [19]; the first four levels of the Farey tree are given below).

Level										
1			$\frac{0}{1}$			$\frac{1}{1}$				
2			$\frac{0}{1}$	$\frac{1}{2}$	$\frac{1}{1}$					
3			$\frac{0}{1}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{1}$			
4	$\frac{0}{1}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{1}{1}$	

More precisely, we computed standard deviations corresponding to the first 21 Farey levels; the rationals  $\varrho$  whose corresponding balanced orbit has largest standard deviation at each Farey level are recorded in the table below. For the first several levels, these maximizing rationals are  $\frac{0}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{3}{8}, \frac{5}{13}$ , i.e. continued fraction convergents to  $1 - \gamma$ , where  $\gamma = \frac{1}{2}(\sqrt{5} - 1)$  is the golden ratio<sup>11</sup>. However this pattern does not continue, and in particular it is *not* the case that the balanced measures of slope  $\gamma$  and  $1 - \gamma$  have largest standard deviation amongst all balanced measures.

<sup>11</sup>By symmetry, the continued fraction convergents  $\frac{1}{1}, \frac{1}{2}, \frac{2}{3}, \frac{3}{5}, \frac{5}{8}, \frac{8}{13}$  to  $\gamma$  also correspond to largest standard deviation on these levels.

Level	$\varrho$	Standard deviation
1	0	0 · 0000000000000000
2	1/2	0 · 1666666666666667
3	1/3	0 · 1781741612749496
4	2/5	0 · 1831621879558502
5	3/8	0 · 1836338891075342
6–9	5/13	0 · 1838465249738321
10	23/60	0 · 1838472388787628
11	28/73	0 · 1838477660379445
12	33/86	0 · 1838479659150774
13–14	38/99	0 · 1838480248997053
15	81/211	0 · 1838480274162965
16–20	119/310	0 · 1838480279673947
21	676/1761	0 · 1838480279684007

The above computations suggest that a maximum standard deviation, slightly larger than  $0 \cdot 1838480279684$ , is attained when  $\varrho$  is close to  $676/1761 \approx 0 \cdot 383873$  (and, by symmetry, when  $\varrho$  is close to  $1085/1761$ ).

**Remark 6.3.** Hajek [16] has identified a class of functions, which he called *multimodular*, with the property that the balanced measure of slope  $\varrho$  has smaller ergodic average than other shift-invariant measures of slope  $\varrho$ . While the results in this paper are similar in spirit to [16], their substance is very different: multimodular functions are simple functions defined on an abstract shift space in terms of multi-dimensional convex functions, whereas our convex functions are one-dimensional, and our results depend on base-2 expansions rather than on purely symbolic properties. In particular, the results of this paper cannot be deduced from those of [16], nor the various multimodular generalisations (see e.g. [2]).

#### REFERENCES

- [1] J-P. Allouche & J. Shallit, *Automatic sequences. Theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003.
- [2] E. Altman, B. Gaujal & A. Hordijk, *Discrete-event control of stochastic networks: multimodularity and regularity*, LNM 1829, Springer, 2003.
- [3] C. Bernhardt, A class of endomorphisms of the circle, *Proc. London Math. Soc.*, **45** (1982), 258–280.
- [4] J. Berstel & P. Séébold, Sturmian words, in: *Algebraic Combinatorics on Words* (Chapter 2), vol. 90 of *Encyclopedia of Mathematics and its Applications*, 2002, Cambridge University Press, Cambridge, UK, pp. 45–110.
- [5] R. Bhatia, *Matrix analysis*, Springer-Verlag, New York, 1997.
- [6] D. Blackwell, Comparison of experiments, *Proc. Second Berkeley Symp. Math. Statist. Prob.*, pp. 93–102, Univ. of California Press, Berkeley, 1951.
- [7] D. Blackwell, Equivalent comparisons of experiments, *Ann. Math. Statist.*, **24** (1953), 265–272.
- [8] T. Bousch, Le poisson n’a pas d’arêtes, *Ann. Inst. Henri Poincaré (Proba. et Stat.)* **36**, (2000), 489–508.
- [9] S. Bullett and P. Sentenac, Ordered orbits of the shift, square roots, and the devil’s staircase, *Math. Proc. Camb. Phil. Soc.*, **115** (1994), 451–481.

- [10] E. M. Coven & G. A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory*, **7** (1973), 138–153.
- [11] J. Creedy, *Dynamics of income distribution*, Basil Blackwell, 1985.
- [12] H. Dalton, The measurement of the inequality of incomes, *Econom. J.*, **30** (1920), 348–361.
- [13] J. H. Folkman & D. R. Fulkerson, Edge colorings in bipartite graphs, in *Combinatorial mathematics and its applications* (R. C. Bose & T. A. Dowling, eds.), Chapter 31, pp. 561–577, Univ. of North Carolina Press, Chapel Hill, 1969.
- [14] J.-M. Gambaudo, O. Lanford III, & C. Tresser, Dynamique symbolique des rotations, *C. R. Acad. Sc. Paris*, **299** (1984), 823–826.
- [15] I. Gohberg, S. Goldberg and M. A. Kaashoek, *Classes of linear operators vol. 1*, Birkhäuser, Berlin, 1990
- [16] B. Hajek, Extremal splittings of point processes, *Math. of Operations Research*, **10** (1985), 543–556.
- [17] G. H. Hardy, J. E. Littlewood, & G. Pólya, Some simple inequalities satisfied by convex functions, *Messenger Math.*, **58** (1929), 145–152.
- [18] G. H. Hardy, J. E. Littlewood, & G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, (1st ed., 1934; 2nd ed., 1952).
- [19] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (5th edition), Oxford, 1979.
- [20] G. A. Hedlund, Sturmian minimal sets, *Amer. J. Math.*, **66** (1944), 605–620.
- [21] O. Jenkinson, Frequency locking on the boundary of the barycentre set, *Experimental Mathematics*, **9** (2000), 309–317.
- [22] O. Jenkinson, Optimization and majorization of invariant measures, *Electron. Res. Announc. Amer. Math. Soc.*, **13** (2007), 1–12.
- [23] O. Jenkinson, A partial order on  $\times 2$ -invariant measures, *Math. Res. Lett.*, to appear.
- [24] O. Jenkinson & L. Q. Zamboni, Characterisations of balanced words via orderings, *Theoret. Comp. Sci.*, **310** (2004), 247–271.
- [25] K. Keller, *Invariant factors, Julia equivalences and the (abstract) Mandelbrot set*, LNM 1732, Springer, 2000.
- [26] M. O. Lorenz, Methods of measuring concentration of wealth, *J. Amer. Statist. Assoc.*, **9** (1905), 209–219.
- [27] A. W. Marshall & I. Olkin, *Inequalities: theory of majorization and its applications*, Mathematics in science and engineering vol. 143, Academic Press, 1979.
- [28] L. Mirsky, Results and problems in the theory of doubly stochastic matrices, *Z. Wahrscheinlichkeitstheorie*, **1** (1963), 319–334.
- [29] M. Morse & G. A. Hedlund, Symbolic Dynamics II. Sturmian Trajectories, *Amer. J. Math.*, **62** (1940), 1–42.
- [30] R. F. Muirhead, Some methods applicable to identities and inequalities of symmetric algebraic functions of  $n$  letters, *Proc. Edinburgh Math. Soc.*, **21** (1903), 144–157.
- [31] A. C. Pigou, *Wealth and welfare*, Macmillan, New York, 1912.
- [32] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, LNM 1794, Springer, 2002.
- [33] I. Schur, Über eine Klasse von Mittelbildungen mit Anwendungen die Determinanten, *Theorie Sitzungsber. Berlin. Math. Gesellschaft*, **22** (1923), 9–20 (*Issai Schur collected works*, (A. Brauer & H. Rohrbach, eds.) Vol. II, pp. 416–427, Springer-Verlag, Berlin, 1973).
- [34] C. Series, The geometry of Markoff numbers, *Math. Intelligencer*, **7**, no. 3 (1985), 20–29.
- [35] J. M. Steele, *The Cauchy-Schwarz master class. An introduction to the art of mathematical inequalities*, MAA Problem Books Series, Mathematical Association of America, Cambridge University Press, Cambridge, 2004.
- [36] D. Stoyan, *Comparison methods for queues and other stochastic methods*, Wiley, 1983.
- [37] Y. L. Tong, *Probability inequalities in multivariate distributions*, Academic Press, 1980.

SCHOOL OF MATHEMATICAL SCIENCES, QUEEN MARY, UNIVERSITY OF LONDON, MILE END ROAD, LONDON, E1 4NS, UK.

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