

## OPTIMIZATION AND MAJORIZATION OF INVARIANT MEASURES

OLIVER JENKINSON

(Communicated by Klaus Schmidt)

ABSTRACT. The set of  $\times 2$ -invariant measures can be equipped with the partial order of majorization, describing relative dispersion. The minimal elements for this order are precisely the Sturmian measures of Morse and Hedlund. This yields new characterisations of Sturmian measures, and has applications to the ergodic optimization of convex functions.

### 1. INTRODUCTION

The dynamical system given by iterating the “ $\times 2$  map”  $T(x) = 2x \pmod{1}$  is a textbook favourite (see e.g. [30, §§1.7, 2.4], [41, §2.1], [48, §2.1]). It is a standard example of a uniformly hyperbolic system, leading naturally to notions in symbolic dynamics and more general hyperbolic theory. As a simple dynamically interesting endomorphism of a compact abelian group it is a prototypical model of an algebraic dynamical system (cf. [14, 46]), and after homothetic conjugacy to the map  $z \mapsto z^2$  on the unit circle in  $\mathbb{C}$  it also serves as the basic example of a holomorphic dynamical system acting on its Julia set (see e.g. [3, §1.3], [39, §4]).

The dynamics of  $T$  is complicated, a fact reflected by the abundance and diversity of its invariant probability measures. Nevertheless, its simple definition offers the possibility of a finer analysis of these measures than is possible in the general case. In particular, it turns out to be possible to describe the *relative dispersion* of  $\times 2$ -invariant measures. This notion, described below, has a sense for any dynamical system defined on a compact convex set; for the  $\times 2$  map, however, it has a particularly explicit flavour.

For reasons which will become apparent, we extend  $T$  to a self-map of  $X = [0, 1]$  in the usual way, by defining  $T(1) = 1$ . The simplex  $\mathcal{M}$  of  $T$ -invariant Borel probability measures can then be equipped with a partial order, *majorization*, defined as follows. If  $\mathcal{N}$  denotes the set of Borel probability measures on  $X$ , and  $\mathcal{N}_\varrho$  the set of  $\mu \in \mathcal{N}$  whose barycentre  $b(\mu) := \int x d\mu(x)$  equals  $\varrho$ , we say that  $\mu$  is *majorized* by  $\nu$ , and write  $\mu \prec \nu$ , if  $\nu = \int D_x d\mu(x)$  for some family  $(D_x)_{x \in X}$  with  $D_x \in \mathcal{N}_x$  for each  $x \in X$ . The family  $(D_x)_{x \in X}$  is the *Markov kernel*, while the associated *dilation* (or *dilatation*, or *balayage*) is the map  $D : \mathcal{N} \rightarrow \mathcal{N}$  defined

---

Received by the editors September 15, 2006.

2000 *Mathematics Subject Classification*. Primary 37A05, 37D20, 37E05; Secondary 37B10, 37E45, 37F15, 46A55.

*Key words and phrases*. Invariant measures, majorization, dilation, ergodic optimization.

The author was supported by an EPSRC Advanced Research Fellowship.

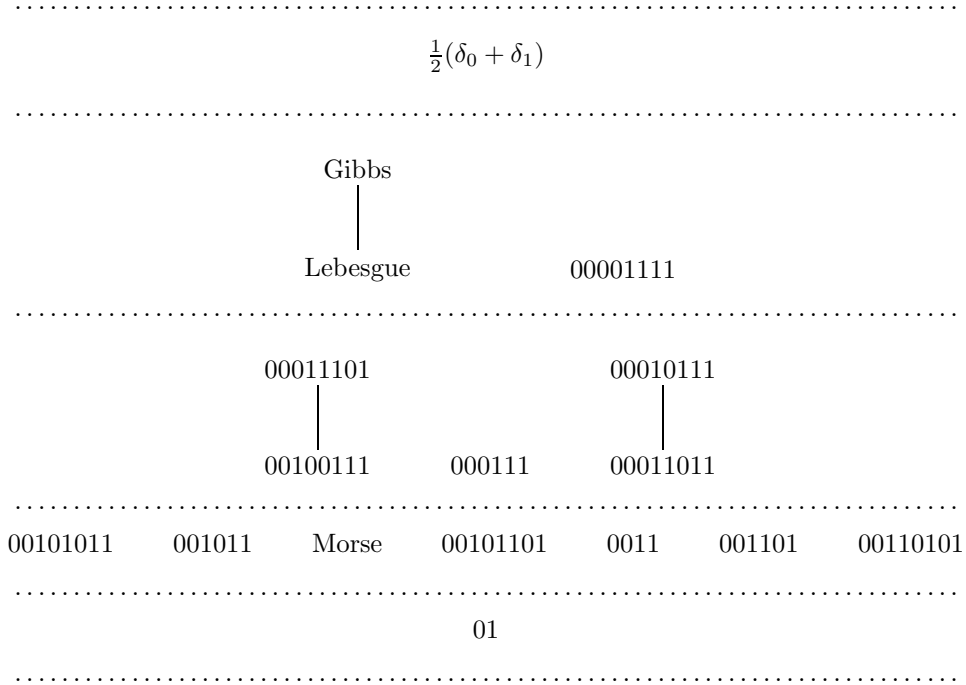


FIGURE 1. Part of  $(\mathcal{M}_{1/2}, \prec)$ . Symbolic codes denote periodic orbit measures (see §2 for more detail) up to period 8. The least element in  $(\mathcal{M}_{1/2}, \prec)$  is carried by the period-2 orbit; the greatest element is non-ergodic, giving weight 1/2 to each of the fixed points 0 and 1. Morse denotes the unique invariant probability measure carried by the shift orbit closure of the Morse sequence  $0110100110010110\dots$ , while Gibbs denotes a particular Gibbs measure (see §3). The diagram is ordered upwards: a given measure is majorized by all those in higher bands (bands are demarcated by dotted horizontal lines). Measures in the same band are incomparable unless a vertical line connects them, in which case the lower measure is majorized by the higher one.

by  $D\mu = \int D_x d\mu(x)$ . See e.g. [12, 37, 42] for more details. Intuitively, a dilation spreads out the mass of a probability measure: if  $\mu \prec \nu$  then the mass of  $\nu$  is, in a certain precise sense, closer to the boundary of  $X$  than that of  $\mu$ .

This relation is a partial order on  $\mathcal{N}$ , and the ordered set  $(\mathcal{N}, \prec)$  is the disjoint union of the  $(\mathcal{N}_\varrho, \prec)$ , because dilations preserve barycentres. Similarly  $(\mathcal{M}, \prec) = \coprod_{\varrho \in [0,1]} (\mathcal{M}_\varrho, \prec)$ , where  $\mathcal{M}_\varrho := \mathcal{M} \cap \mathcal{N}_\varrho$ . If  $T$  is identified, in the usual way, with the shift map on the set  $\{0, 1\}^{\mathbb{N}}$ , then  $\mathcal{M}_\varrho$  is precisely the set of invariant measures giving weight  $\varrho$  to the symbol “1”; for example Bernoulli measure  $B(1 - \varrho, \varrho)$  lies in  $\mathcal{M}_\varrho$  (and maximizes entropy therein). Figure 1 illustrates part of  $(\mathcal{M}_{1/2}, \prec)$  (cf. the sparser diagram in [26]).

The maximal elements of  $(\mathcal{N}, \prec)$  are precisely the convex combinations  $\nu_\varrho := \varrho\delta_1 + (1 - \varrho)\delta_0$  of the Dirac masses at 0 and 1, since  $\nu_\varrho = \int \nu_x d\mu(x)$  for all  $\mu \in \mathcal{N}_\varrho$ .

So each  $(\mathcal{N}_\varrho, \prec)$  has a greatest element, its unique member carried<sup>1</sup> by  $\partial X$ . But 0 and 1 are both fixed points for  $T$ , so the measure  $\nu_\varrho$  is also the greatest element in  $\mathcal{M}_\varrho$ . Each  $(\mathcal{N}_\varrho, \prec)$  also has a least element, the Dirac measure  $\delta_\varrho$ , but these are not invariant (unless  $\varrho = 0$  or 1), so determining the minimal elements of  $(\mathcal{M}, \prec)$  is a non-trivial problem. In particular it is unclear, a priori, whether or not each  $(\mathcal{M}_\varrho, \prec)$  has a least element.

It turns out that each  $(\mathcal{M}_\varrho, \prec)$  does have a least element  $S_\varrho$ , which can be identified explicitly. Indeed the family  $S_\varrho$  is well known, having been studied, in various contexts and for varying reasons, since the early days of symbolic dynamics:

**Theorem 1.1.** *For every  $\varrho \in [0, 1]$ , the ordered set  $(\mathcal{M}_\varrho, \prec)$  has a least element. This least element is the Sturmian measure of rotation number  $\varrho$ .*

Sturmian measures, or rather the symbol sequences (*Sturmian sequences*<sup>2</sup>) corresponding to points in their support, were defined by Morse & Hedlund [40]. For each  $\varrho \in (0, 1)$  there is a unique non-empty minimal closed invariant set  $X_\varrho$  on which  $T$  is combinatorially equivalent to rotation by angle  $\varrho$ ; the set  $X_\varrho$  is a periodic orbit if and only if  $\varrho$  is rational (see e.g. [11, 53]). The *Sturmian measure* of rotation number  $\varrho$ , denoted  $S_\varrho$ , is the unique invariant probability measure carried by  $X_\varrho$ ; equivalently,  $S_\varrho$  is the push forward of Lebesgue measure on  $X$  under the map  $h(x) = \sum_{n \geq 0} \chi_{[1-\varrho, 1)}(\{x + n\varrho\})/2^{n+1}$ , where  $\{\cdot\}$  denotes reduction modulo 1 (see e.g. [7, 24]). If  $\varrho = 0$  or 1 then  $S_\varrho$  is defined to be  $\delta_\varrho$ .

Each Sturmian measure is ergodic, and has zero entropy; indeed Sturmian sequences are characterised by having smallest possible symbolic complexity: the number of distinct length- $n$  subwords of a Sturmian sequence equals  $n + 1$  for all  $n$  smaller than the period  $q$  of the sequence ( $q = \infty$  if  $\varrho$  is irrational). Another symbolic property characterising Sturmian sequences is that they are *balanced*: there exist integers  $p_1, p_2, \dots$  such that every length- $n$  subword of the sequence contains either  $p_n$  or  $p_n + 1$  occurrences of the symbol “1”. Sturmian sequences are used extensively in computer science (where, as previously noted, the term *balanced* is preferred), their principal application being to the discrete approximation of lines (see e.g. [1, 36, 49]). Any  $\times 2$ -orbit contained in  $X_\varrho$  is called a *Sturmian orbit* of rotation number  $\varrho$ . For rational  $\varrho$  the Sturmian orbit is unique, and periodic, while for irrational  $\varrho$  there are many distinct Sturmian orbits, each of which is dense in the Cantor set  $X_\varrho$ . A Sturmian orbit is characterised by being *ordered*: the cyclic ordering of its points is invariant under  $T$  (see e.g. [11]). Another characterisation is that Sturmian orbits are precisely those which are contained in some length-1/2 sub-interval of  $X$  (see e.g. [11, 19, 53]), a fact exploited in the proof (see below) of Theorem 1.1. Among periodic orbits, all those of sufficiently low period are Sturmian (the period-4 orbit  $\{1/5, 2/5, 3/5, 4/5\}$  is the lowest period example of one which is not), but as  $q \rightarrow \infty$  the proportion of period- $q$  orbits which are Sturmian tends to zero. One important application of Sturmian orbits arises in complex dynamics, in the combinatorial description of the Mandelbrot set (see e.g. [11, 17, 32] for more details).

A classical result, due in its simplest form to Hardy, Littlewood, & Pólya [22], is that majorization may be re-formulated in terms of convex functions, as we now

<sup>1</sup>We say that  $Y \subset X$  carries a measure  $\mu$  whenever the support of  $\mu$  lies in  $Y$ .

<sup>2</sup>Many authors use the term *balanced* instead of Sturmian, reserving the latter terminology only for the case where  $\varrho$  is irrational. Further details on the construction and properties of these sequences can be found in e.g. [1, 11, 36, 44].

describe. Suppose  $\mu \prec \nu$  with  $\nu = D\mu$  for some Markov kernel  $(D_x)_{x \in X}$ . Since each  $D_x \in \mathcal{N}_x$ , Jensen's inequality implies that  $f(x) \leq D_x(f)$  for every convex function  $f : X \rightarrow \mathbb{R}$ , so  $\nu(f) = \int D_x(f) d\mu(x) \geq \int f(x) d\mu(x) = \mu(f)$ . In fact the converse is also true (see [12, Thm. 2]): if  $\mu(f) \leq \nu(f)$  for every convex  $f$ , then  $\mu \prec \nu$ . So, defining a measure  $\mu \in \mathcal{M}_\varrho$  to be *f-minimizing in  $\mathcal{M}_\varrho$*  if  $\mu(f) = \inf_{m \in \mathcal{M}_\varrho} m(f)$ , Theorem 1.1 is equivalent to:

**Theorem 1.2.** *If  $f : X \rightarrow \mathbb{R}$  is convex then for every  $\varrho \in [0, 1]$ , the Sturmian measure of rotation number  $\varrho$  is f-minimizing in  $\mathcal{M}_\varrho$ .*

Since  $D_x(f) > f(x)$  whenever  $f$  is strictly convex and  $D_x$  is not a Dirac point mass, we deduce:

**Corollary 1.3.** *If  $f : X \rightarrow \mathbb{R}$  is strictly convex then for every  $\varrho \in [0, 1]$ , the Sturmian measure of rotation number  $\varrho$  is the unique f-minimizing measure in  $\mathcal{M}_\varrho$ .*

In the case where an invariant measure is supported on a single periodic orbit, its barycentre is simply the arithmetic mean of the points in the orbit; setting  $f(x) = -\log x$  in Corollary 1.3 shows that, among orbits with the same arithmetic mean, the *geometric* mean is maximized precisely when the orbit is Sturmian:

**Corollary 1.4.** *Among all periodic orbits with a given arithmetic mean, the Sturmian orbit has strictly largest geometric mean.*

For any measure  $\mu \in \mathcal{M}$ , its *variance* (around the mean  $b(\mu)$ ) is defined as  $\text{var}(\mu) = \int (x - b(\mu))^2 d\mu(x)$ . Therefore Corollary 1.3 also implies:

**Corollary 1.5.** *For every  $\varrho \in [0, 1]$ , the Sturmian measure  $S_\varrho$  is the unique measure with smallest variance in  $\mathcal{M}_\varrho$ .*

For a bounded Borel function  $f : X \rightarrow \mathbb{R}$ , a measure  $\mu \in \mathcal{M}$  is said to be *minimizing* if  $\mu(f) = \inf_{m \in \mathcal{M}} m(f)$  and *maximizing* if  $\mu(f) = \sup_{m \in \mathcal{M}} m(f)$ ; see e.g. [8, 9, 15, 25].<sup>3</sup> The problem of determining minimizing and maximizing measures for specific  $f$  was the original motivation for Theorem 1.2, since it implies:

**Corollary 1.6.** *Every convex function  $f : X \rightarrow \mathbb{R}$  has a Sturmian minimizing measure. If  $f$  is strictly convex then it has a unique minimizing measure, and this measure is Sturmian.*

More precisely,  $S_{\varrho_0}$  is *f-minimizing* if and only if  $\varrho_0$  is a minimum for the (itself convex) function  $\varrho \mapsto S_\varrho(f)$  (see Figure 2 for a particular example). An equivalent formulation of Corollary 1.6 is:<sup>4</sup>

**Corollary 1.7.** *Every concave function  $f : X \rightarrow \mathbb{R}$  has a Sturmian maximizing measure. If  $f$  is strictly concave then it has a unique maximizing measure, and this measure is Sturmian.*

Quadratic polynomials are either strictly convex or strictly concave, so:

<sup>3</sup>These notions are “global”: there is no constraint on the barycentre of the measure. Of course a measure which is *f-minimizing* in some  $\mathcal{M}_\varrho$  need not be *f-minimizing*, but clearly any *f-minimizing* measure  $\mu$  is also *f-minimizing* in  $\mathcal{M}_{b(\mu)}$ .

<sup>4</sup>Note that the maximizing measures for a convex function can trivially be determined, since provided the function is not constant, it attains its global maximum at one, or possibly both, of the fixed points 0 and 1, and at no other points.

**Corollary 1.8.** *Let  $f : X \rightarrow \mathbb{R}$  be a quadratic polynomial  $f(x) = ax^2 + bx + c$ . If  $a > 0$  then there is a unique  $f$ -minimizing measure, and this measure is Sturmian. If  $a < 0$  then there is a unique  $f$ -maximizing measure, and this measure is Sturmian.*

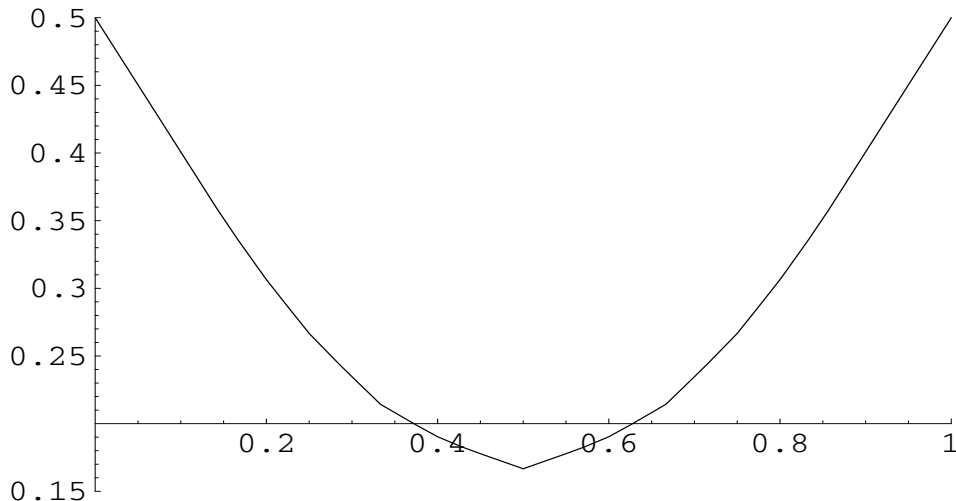


FIGURE 2. Function  $\varrho \mapsto S_\varrho(f)$  for  $f(x) = |x - 1/2|$ .

The above corollaries enlarge the list of functions whose minimizing and maximizing measures are known to be Sturmian. Bousch [7] has shown that any degree-one trigonometric polynomial has Sturmian minimizing and maximizing measures. Corollary 1.8 is comparable to [7] in that quadratic polynomials are the simplest non-trivial functions on the interval, while trigonometric polynomials of degree one are the simplest non-trivial functions on the circle.

Proofs of all the above results can be found in [26, 27]. Here we provide a sketch proof of Theorem 1.1, or rather the equivalent Theorem 1.2, asserting that  $S_\varrho(f) \leq \mu(f)$  for every convex  $f$  and every  $\mu \in \mathcal{M}_\varrho$ ,  $\varrho \in (0, 1)$ . First note that we may assume  $f$  to be  $C^2$ , since such convex functions are weakly dense among all convex functions. Moreover we are at liberty to tilt the graph of  $f$ : it suffices to find  $\theta = \theta(\varrho) \in \mathbb{R}$  such that  $S_\varrho$  is  $f_\theta$ -minimizing in  $\mathcal{M}_\varrho$ , where  $f_\theta(x) := f(x) + \theta x$ . In fact  $\theta$  can be chosen with the stronger property that  $S_\varrho$  is a (globally) minimizing measure for  $f_\theta$ , as we now explain.

Recalling [11, 19, 53] that there exists  $\gamma \in [0, 1/2]$  such that the length-1/2 interval  $H_\gamma = [\gamma, \gamma + 1/2]$  carries  $S_\varrho$ , define  $\theta := -\int \sum_{n=1}^{\infty} 2^{-n} f' \circ \tau^n$ , where  $\tau : [0, 1] \rightarrow [0, 1]$  is the “inverse” of  $T|_{H_\gamma}$  (precisely,  $\tau(x)$  equals  $x/2$  for  $x \in [2\gamma, 1]$ , and  $(x+1)/2$  otherwise). The function  $f_\theta$  can then be *flattened* on  $H_\gamma$  (cf. [7, 20]): if  $\varphi_\theta$  is the (Lipschitz continuous) function whose (Lebesgue almost everywhere) derivative is  $\sum_{n=1}^{\infty} 2^{-n} f'_\theta \circ \tau^n$ , it can be shown that  $f_\theta + \varphi_\theta - \varphi_\theta \circ T =: \tilde{f}_\theta$  is constant on  $H_\gamma$ , with constant value  $c$ , say. Of course a measure  $\mu \in \mathcal{M}$  is  $f_\theta$ -minimizing if and only if it is  $\tilde{f}_\theta$ -minimizing, since  $\mu(\varphi_\theta - \varphi_\theta \circ T) = 0$ , so it is enough to show that  $S_\varrho$  is  $\tilde{f}_\theta$ -minimizing. This is certainly the case if it can be shown that  $c$  is the minimum value of  $\tilde{f}_\theta$ , since then  $\mu(\tilde{f}_\theta) \geq c$  for  $\mu \in \mathcal{M}$ , and  $S_\varrho(\tilde{f}_\theta) = c$  because  $S_\varrho$  is

carried by the “flat” part of  $\tilde{f}_\theta$ . To prove that the minimum value of  $\tilde{f}_\theta$  is indeed  $c$ , we show that its value at each  $s \in H_\gamma$  is smaller than at either  $s + 1/2$  (if  $s \leq 1/2$ ) or  $s - 1/2$  (if  $s > 1/2$ ). Concentrating on the first case (the second case is similar), this amounts to proving that  $\Delta_\theta(s) := (f_\theta + \varphi_\theta)(s) - (f_\theta + \varphi_\theta)(s + 1/2) \leq 0$ .

By the fundamental theorem of calculus, the difference  $\Delta_\theta(s)$  may be written as  $\int C_s(x) f'_\theta(x) dx$  for some  $L^1$  function  $C_s$  whose primitive  $B_s(t) := \int_0^t C_s(x) dx$  vanishes at both 0 and 1, so that integration by parts gives  $\Delta_\theta(s) = -\int B_s(x) f''_\theta(x) dx$ . But  $f''_\theta = f'' \geq 0$ , so the required non-positivity of  $\Delta_\theta(s)$  follows from the non-negativity of  $B_s$ . This latter fact can be established by carefully analysing the sign changes of  $C_s$ , thereby completing the proof of Theorem 1.2.

It turns out that the uniqueness result in Corollary 1.3 can, with more work (see [27]), be extended to certain convex functions which are not strictly convex. Indeed the only obstruction to  $S_\varrho$  being the unique  $f$ -minimizing measure in  $\mathcal{M}_\varrho$  is the presence of long affine segments in the graph of  $f$ :

**Theorem 1.9.** *Let  $\varrho \in [0, 1]$ . If  $f : X \rightarrow \mathbb{R}$  is convex, and not affine on any interval which carries more than one measure from  $\mathcal{M}_\varrho$ , then the Sturmian measure  $S_\varrho$  is the unique  $f$ -minimizing measure in  $\mathcal{M}_\varrho$ .*

Theorem 1.9 implies an analogue of Corollary 1.5, namely that *mean absolute deviation*  $\int |x - b(\mu)| d\mu(x)$  is minimized precisely when  $\mu$  is Sturmian; this is because the affine segments of  $x \mapsto |x - \varrho|$  do not carry any measures from  $\mathcal{M}_\varrho$ .

A more explicit version of Theorem 1.9 is:

**Corollary 1.10.** *If  $f : X \rightarrow \mathbb{R}$  is convex, and not affine on any interval of length strictly greater than  $1/2$ , then for each  $\varrho \in [0, 1]$ , the Sturmian measure  $S_\varrho$  is the unique  $f$ -minimizing measure in  $\mathcal{M}_\varrho$ . In particular, the minimizing measure for  $f$  is unique, and Sturmian.*

In particular, if a convex function is symmetric about the point  $1/2$ , and attains its minimum only there, then any affine segments are sufficiently small to guarantee uniqueness of the minimizing measure in every  $\mathcal{M}_\varrho$ . The (global) minimizing measure can therefore be identified precisely:

**Corollary 1.11.** *If a convex function is symmetric about the point  $1/2$ , and attains its minimum only there, then its unique minimizing measure is the one carried by the period-2 orbit  $\{1/3, 2/3\}$ .*

For non-symmetric convex functions it is harder to find simple general conditions permitting the precise identification of the minimizing measure(s). For any specific such  $f$ , however, the reduction to a one-parameter family by Theorem 1.2 and Corollary 1.3 is a significant simplification; minima of  $\varrho \mapsto S_\varrho(f)$  can be found numerically, and if necessary confirmed rigorously, for example using ideas from [7, 20].

## 2. CHARACTERISATIONS OF PERIODIC STURMIAN ORBITS

Although notions of majorization are present in the work of early 20th century economists [16, 35, 43], the concept was put on a proper mathematical footing by Hardy, Littlewood, & Pólya in their celebrated 1929 paper [22] and subsequent monograph [23], and independently by Karamata [28]. In these investigations, majorization is formulated in terms of finite sets of real numbers (corresponding

to atoms of equal mass) rather than general probability measures. Even in this simple setting, majorization has proved to be an enormously useful concept, finding application in diverse branches of mathematics (see e.g. [37]). For example in pure mathematics it is the basis for numerous inequalities (see [37, Pt. II]), while in various applications it is regarded as the proper way of measuring diversity, providing a more subtle indicator of relative dispersion than simple consideration of variance around the mean (see e.g. [37, 45, 51]).

In our dynamical setting, atomic measures correspond to periodic orbits: if  $\{x, T(x), \dots, T^{q-1}(x)\}$  is an orbit of least period  $q$ , it carries a unique  $T$ -invariant probability measure, namely  $\mu = q^{-1} \sum_{i=0}^{q-1} \delta_{T^i(x)}$ . Every period- $q$  point  $x$  has a binary expansion  $x = \sum_{n \geq 1} x_n / 2^n$  whose corresponding symbol sequence  $(x_n)_{n=1}^{\infty}$  is periodic, of period  $q$ ; so we may represent  $x$  by the length- $q$  binary word  $x_1 \dots x_q$ . Each point  $T^i(x)$  in the  $T$ -orbit of  $x$  is represented by the length- $q$  binary word  $x_{i+1} \dots x_q x_1 \dots x_i$  obtained by cyclically shifting  $x_1 \dots x_q$ . Of these  $q$  words, we choose to notationally represent the whole periodic orbit, and the invariant measure  $\mu$  supported upon it, by the one which is lexicographically smallest. So e.g. 0011 denotes the invariant measure carried by  $\{1/5, 2/5, 3/5, 4/5\}$  (see Figure 1 for more usage of this notation). The barycentre  $b(\mu)$  is just the arithmetic mean  $q^{-1} \sum_{i=0}^{q-1} T^i(x)$ , and equals  $p/q$ , where  $p$  is the number of 1's in the periodic word representing the measure ( $p$  need not be coprime to  $q$ ).

If  $\mu, \nu \in \mathcal{N}$  can be written as  $\mu = Q^{-1} \sum_{i=1}^Q \delta_{\mu_i}$ ,  $\nu = Q^{-1} \sum_{i=1}^Q \delta_{\nu_i}$ , with  $\mu_1 \leq \dots \leq \mu_Q$  and  $\nu_1 \leq \dots \leq \nu_Q$ , and  $b(\mu) = b(\nu)$ , then a result of Hardy, Littlewood, & Pólya [22, 23] asserts that  $\mu \prec \nu$  if and only if

$$(2.1) \quad \sum_{i=1}^n \mu_i \geq \sum_{i=1}^n \nu_i \quad \text{for all } 1 \leq n \leq Q-1.$$

So Theorem 1.1 has the following corollary:

**Corollary 2.1.** *For a periodic orbit  $\mu_1 < \dots < \mu_q$  of least period  $q$ , if the integer  $p = \sum_{i=1}^q \mu_i$  is coprime to  $q$  then*

$$(2.2) \quad \sum_{i=1}^n s_i \geq \sum_{i=1}^n \mu_i \quad \text{for all } 1 \leq n \leq q-1,$$

where  $s_1 < \dots < s_q$  are the points in the Sturmian orbit of rotation number  $p/q$ .

**Example 2.2.** The invariant measure  $\mu = 00011$  carried by the period-5 orbit

$$3/31 < 6/31 < 12/31 < 17/31 < 24/31$$

majorizes the Sturmian measure  $S_{2/5} = 00101$  carried by the periodic orbit

$$5/31 < 9/31 < 10/31 < 18/31 < 20/31,$$

a fact which may be verified by observing that (2.2) corresponds to

$$5 \geq 3, \quad 14 \geq 9, \quad 24 \geq 21, \quad 42 \geq 38.$$

The coprime assumption in Corollary 2.1 was just for simplicity. The following version is valid for arbitrary periodic orbits:

**Corollary 2.3.** *Let  $\mu_1 < \dots < \mu_Q$  be a periodic orbit of least period  $Q$ , with  $r$  the highest common factor of  $P = \sum_{i=1}^Q \mu_i$  and  $Q$ . Writing  $P = rp$ ,  $Q = rq$ , let  $s_1 <$*

$\dots < s_q$  be the Sturmian orbit of rotation number  $p/q$ , and define  $s'_{i+(j-1)q} = s_j$  for  $1 \leq i \leq q$ ,  $1 \leq j \leq r$ . Then

$$(2.3) \quad \sum_{i=1}^n s'_i \geq \sum_{i=1}^n \mu_i \quad \text{for all } 1 \leq n \leq Q-1.$$

*Remark 2.4.* The inequalities  $s_1 \geq \mu_1$  and  $s_q \leq \mu_Q$  (i.e. the extreme cases  $n = 1$  and  $Q-1$  of (2.3)) can be proved using different methods (see [4, 19]).

Another reformulation of majorization due to Hardy, Littlewood, & Pólya [22], partly anticipated by Schur [47], asserts that  $Q^{-1} \sum_{i=1}^Q \delta_{\mu_i} = \mu \prec \nu = Q^{-1} \sum_{i=1}^Q \delta_{\nu_i}$  if and only if  $(\mu_1, \dots, \mu_Q) = (\nu_1, \dots, \nu_Q)A$  for a matrix  $A$  which is doubly stochastic (i.e. has non-negative entries, with each row and column sum equal to 1). As a corollary of Theorem 1.1 we therefore have:

**Corollary 2.5.** *With the hypotheses and notation of Corollary 2.3, there exists a doubly stochastic  $Q \times Q$  matrix  $A$  such that  $(s'_1, \dots, s'_Q) = (\mu_1, \dots, \mu_Q)A$ .*

### 3. CHARACTERISATIONS OF STURMIAN MEASURES

Some indication of the importance of majorization is the extent to which it has been re-discovered and re-invented in different settings. For example in 1951, Blackwell [5, 6] considered majorization for general probability measures on the interval, as a means of comparing statistical experiments. Independently, it was introduced by Choquet [13] in greater generality, as a tool for representing members of compact convex sets as barycentres of measures supported on the extreme boundary (see e.g. [42] for background).

For  $\mu, \nu \in \mathcal{N}$ , the generalisation of the doubly stochastic matrix characterisation mentioned in §2 is that  $\mu \prec \nu$  if and only if there exists a coupling<sup>5</sup>  $P$  of the ordered pair  $(\mu, \nu)$  such that

$$(3.1) \quad \int_{A \times X} y dP(x, y) = \int_A x d\mu(x)$$

for all Borel sets  $A \subset X$  (see [5, 12]). Equivalently,  $\mu$  is a *fusion* of  $\nu$ , in the sense of Elton & Hill [18]; an *elementary fusion* of  $\nu$  is the probability measure obtained by taking some proportion  $p$  of the  $\nu$ -mass of a Borel set  $A \subset X$  and replacing it by the appropriately scaled Dirac mass  $p\delta_{b(A, \nu)}$  concentrated at the barycentre  $b(A, \nu) := \nu(A)^{-1} \int_A x d\nu(x)$ , while a *fusion* of  $\nu$  is by definition any weak\* accumulation point of finite compositions of elementary fusions. Another interpretation, due to Strassen [50, Thm. 8], is that  $\mu \prec \nu$  if and only if  $(\mu, \nu)$  is a *martingale pair*, i.e. there exist  $X$ -valued random variables  $\varphi$  and  $\psi$  on some probability space  $(\Omega, \mathcal{B}, P)$ , whose laws are  $\mu$  and  $\nu$  respectively, such that  $\varphi = E(\psi | \varphi)$  a.e. ( $P$ ), where  $E(\cdot | \cdot)$  denotes conditional expectation.

In view of these equivalent formulations of majorization, Theorem 1.1 implies the following characterisations of Sturmian measures  $S_\rho$ :

**Corollary 3.1.** *For every  $\times 2$ -invariant probability measure  $\mu$ :*

- (i) *The Sturmian measure  $S_{b(\mu)}$  is a fusion of  $\mu$ ,*
- (ii)  *$(S_{b(\mu)}, \mu)$  is a martingale pair,*

<sup>5</sup>By a coupling (see e.g. [34]) of  $(\mu, \nu)$  we mean a finite measure  $P$  on  $X \times X$  whose first and second marginals are  $\mu$  and  $\nu$  respectively:  $\mu = P\pi_1^{-1}$  and  $\nu = P\pi_2^{-1}$ , where  $\pi_1(x, y) = x$  and  $\pi_2(x, y) = y$ .

(iii) *There exists a coupling  $P$  of  $(S_{b(\mu)}, \mu)$  such that, for all Borel sets  $A \subset X$ ,*

$$\int_{A \times X} y dP(x, y) = \int_A x dS_{b(\mu)}(x).$$

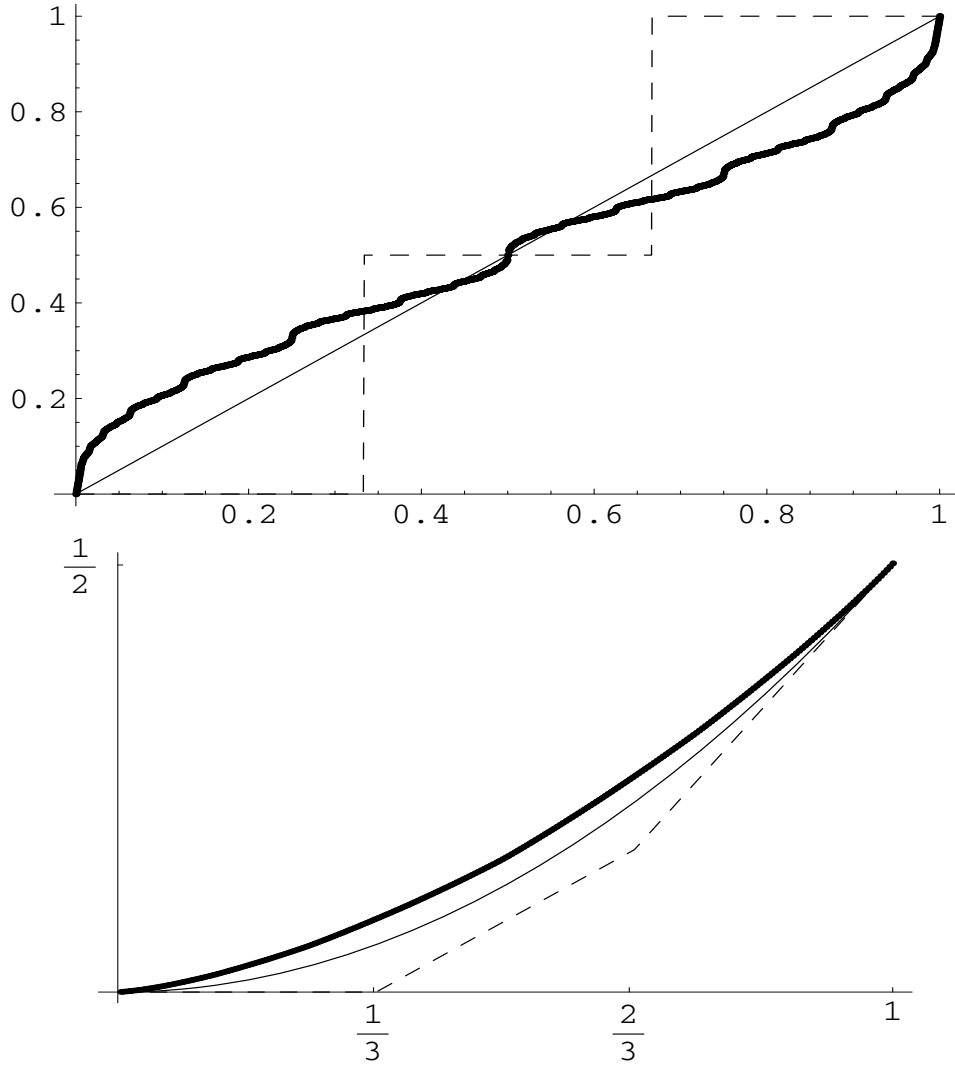


FIGURE 3. Cumulative distribution function  $x \mapsto \mu[0, x]$ , and its primitive  $t \mapsto \int_0^t \mu[0, x] dx$ , for  $\mu$  equal to Sturmiian measure  $01 = (\delta_{1/3} + \delta_{2/3})/2$  (dashed), Lebesgue measure (solid), and Gibbs measure (bold).

The generalisation of the Hardy-Littlewood-Pólya criterion (2.1) to arbitrary  $\mu, \nu \in \mathcal{N}$  is (see [29]) that  $\mu \prec \nu$  if and only if  $b(\mu) = b(\nu)$  and

$$(3.2) \quad \int_0^t \mu[0, x] dx \leq \int_0^t \nu[0, x] dx \quad \text{for all } t \in [0, 1].$$

The criterion (3.2) was used to locate the Morse, Lebesgue and Gibbs measures in the hierarchy of  $(\mathcal{M}_{1/2}, \prec)$  (see Figures 1 and 3); here the Gibbs measure corresponds to the potential function<sup>6</sup>  $x \mapsto \log(\frac{1}{4} + |x - \frac{1}{2}|)$ .

Combining (3.2) and Theorem 1.1 gives:

**Corollary 3.2.** *If  $\varrho \in [0, 1]$  and  $\mu \in \mathcal{M}_\varrho$  then*

$$\int_0^t S_\varrho[0, x] dx \leq \int_0^t \mu[0, x] dx \quad \text{for all } t \in [0, 1].$$

Following yet another re-invention of majorization, due to Rothschild & Stiglitz [45], the criterion (3.2) is widely used in economics, finance, and related areas (see e.g. [2]), signifying that the measure  $\mu$  is “riskier” than  $\nu$ . In the context of stopping times in Brownian motion (see e.g. [38, 52]) it is common to re-formulate (3.2) in terms of the left continuous inverse  $I_\mu(x) := \min\{s \in [0, 1] : \mu[0, s] \geq x\}$  of the cumulative distribution function: if  $b(\mu) = b(\nu)$  then  $\mu \prec \nu$  if and only if  $\int_0^t I_\mu(x) dx \geq \int_0^t I_\nu(x) dx$  for all  $0 \leq t \leq 1$  (cf. [28, 29]). Equivalently,  $\int_t^1 I_\mu(x) dx \leq \int_t^1 I_\nu(x) dx$  for all  $t \in [0, 1]$ , or indeed  $H_\mu \leq H_\nu$ , where  $H_\mu(t) := \frac{1}{1-t} \int_t^1 I_\mu(x) dx$  is the *Hardy-Littlewood maximal function* of  $\mu$  (cf. [18, 21, 33, 52]). So Theorem 1.1 implies:

**Corollary 3.3.** *For  $\varrho \in [0, 1]$ , the Hardy-Littlewood maximal function  $H_{S_\varrho}$  of the Sturmian measure  $S_\varrho$  is pointwise smaller than that of any other measure in  $\mathcal{M}_\varrho$ .*

#### REFERENCES

1. J.-P. Allouche & J. Shallit, *Automatic sequences. Theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003. MR1997038 (2004k:11028)
2. V. S. Bawa, Stochastic dominance: a research bibliography, *Management Science*, **28** (1982), 698–712. MR668272 (83i:90011)
3. A. F. Beardon, *Iteration of rational functions*, Springer-Verlag, 1991. MR1128089 (92j:30026)
4. C. Bernhardt, A class of endomorphisms of the circle, *Proc. London Math. Soc.*, **45** (1982), 258–280. MR670037 (84f:58065)
5. D. Blackwell, Comparison of experiments, *Proc. Second Berkeley Symp. Math. Statist. Prob.*, pp. 93–102, Univ. of California Press, Berkeley, 1951. MR0046002 (13:667f)
6. D. Blackwell, Equivalent comparisons of experiments, *Ann. Math. Statist.*, **24** (1953), 265–272. MR0056251 (15:47b)
7. T. Bousch, Le poisson n’a pas d’arêtes, *Ann. Inst. Henri Poincaré (Proba. et Stat.)*, **36** (2000), 489–508. MR1785392 (2001i:37005)
8. T. Bousch, La condition de Walters, *Ann. Sci. ENS*, **34** (2001), 287–311. MR1841880 (2002h:37051)
9. T. Bousch & O. Jenkinson, Cohomology classes of dynamically non-negative  $C^k$  functions, *Invent. Math.*, **148** (2002), 207–217. MR1892849 (2003c:37048)
10. R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer LNM, **470**, Berlin-Heidelberg-New York, 1975. MR0442989 (56:1364)
11. 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. MR1269932 (95j:58043)
12. P. Cartier, J. M. G. Fell, & P.-A. Meyer, Comparaison des mesures portées par un ensemble convexe compact, *Bull. Soc. Math. France*, **92** (1964), 435–445. MR0206193 (34:6014)
13. G. Choquet, Le théorème de représentation intégrale dans les ensembles convexes compacts, *Ann. Inst. Fourier (Grenoble)*, **10** (1960), 333–344. MR0126709 (23:A4003)
14. V. Chothi, G. Everest & T. Ward,  $S$ -integer dynamical systems: periodic points, *J. reine angew. Math.*, **489** (1997), 99–132. MR1461206 (99b:11089)

<sup>6</sup>See e.g. [10] for more on potential functions and their Gibbs measures (or *equilibrium states*, or *g-measures* [31]).

15. G. Contreras, A. Lopes, & Ph. Thieullen, Lyapunov minimizing measures for expanding maps of the circle, *Ergod. Th. & Dyn. Sys.*, **21** (2001), 1379–1409. MR1855838 (2002i:37038)
16. H. Dalton, The measurement of the inequality of incomes, *Econom. J.*, **30** (1920), 348–361.
17. A. Douady & J. H. Hubbard, Etude dynamique des polynômes complexes, *Publ. Math. Orsay* I (1984), II (1985). MR0762431 (87f:58072a); MR0762431 (87f:58072b)
18. J. Elton & T. P. Hill, Fusions of a probability measure, *Ann. Prob.*, **20** (1992), 421–454. MR1143430 (93a:60007)
19. J.-M. Gambaudo, O. Lanford III, & C. Tresser, Dynamique symbolique des rotations, *C. R. Acad. Sc. Paris*, **299** (1984), 823–826. MR772104 (86g:58086)
20. E. Harriss & O. Jenkinson, Flattening functions on flowers, *preprint*, 2006.
21. G. H. Hardy & J. E. Littlewood, A maximal theorem with function theoretic applications, *Acta Math.*, **54** (1930), 81–116.
22. G. H. Hardy, J. E. Littlewood, & G. Pólya, Some simple inequalities satisfied by convex functions, *Messenger Math.*, **58** (1929), 145–152.
23. G. H. Hardy, J. E. Littlewood, & G. Pólya, *Inequalities*, Cambridge University Press, Cambridge (1st ed., 1934; 2nd ed., 1952). MR0046395 (13:727e)
24. O. Jenkinson, Frequency locking on the boundary of the barycentre set, *Experimental Mathematics*, **9** (2000), 309–317. MR1780215 (2001g:37050)
25. O. Jenkinson, Ergodic optimization, *Discrete & Cont. Dyn. Sys.*, **15** (2006), 197–224. MR2191393
26. O. Jenkinson, A partial order on  $\times 2$ -invariant measures, *preprint*, 2006.
27. O. Jenkinson, Convex functions with unique Sturmian minimizing measure, *in preparation*.
28. J. Karamata, Sur une inégalité relative aux fonctions convexes, *Publ. Math. Univ. Belgrade*, **1** (1932), 145–148.
29. S. Karlin & A. Novikoff, Generalized convex inequalities, *Pacific J. Math.*, **13** (1963), 1251–1279. MR0156927 (28:170)
30. A. Katok & B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, 1995. MR1326374 (96c:58055)
31. M. Keane, Strongly mixing  $g$ -measures, *Invent. Math.*, **16** (1972), 309–324. MR0310193 (46:9295)
32. K. Keller, *Invariant factors, Julia equivalences and the (abstract) Mandelbrot set*, LNM 1732, Springer, 2000. MR1761576 (2001m:37090)
33. R. P. Kertz & U. Rösler, Stochastic and convex orders and lattices of probability measures, with a martingale interpretation, *Israel J. Math.*, **77** (1992), 129–164. MR1194790 (95d:60078)
34. T. Lindvall, *Lectures on the coupling method*, John Wiley & Sons, 1992. MR1180522 (94c:60002)
35. M. O. Lorenz, Methods of measuring concentration of wealth, *J. Amer. Statist. Assoc.*, **9** (1905), 209–219.
36. M. Lothaire, *Algebraic combinatorics on words*, Encyclopaedia of mathematics and its applications, vol. 90, Cambridge University Press, 2002. MR1905123 (2003i:68115)
37. A. W. Marshall & I. Olkin, *Inequalities: theory of majorization and its applications*, Mathematics in science and engineering, vol. 143, Academic Press, 1979. MR552278 (81b:00002)
38. I. Meilijson, Stochastic orders and stopping times in Brownian motion, in: From classical to modern probability, pp. 207–219, Progress in Probability, vol. 54, Birkhäuser, 2003. MR2045988 (2005a:60130)
39. J. Milnor, *Dynamics in one complex variable*, 2nd ed., Vieweg, 2000. MR1721240 (2002i:37057)
40. M. Morse & G. A. Hedlund, Symbolic Dynamics II. Sturmian Trajectories, *Amer. J. Math.*, **62** (1940), 1–42. MR0000745 (1:123d)
41. E. Ott, *Chaos in dynamical systems*, 2nd ed., Cambridge University Press, Cambridge, 2002. MR1924000 (2004a:37001)
42. R. R. Phelps, *Lectures on Choquet's theorem*, Math. Studies, no. 7, Van Nostrand, 1966. MR0193470 (33:1690)
43. A. C. Pigou, *Wealth and welfare*, Macmillan, New York, 1912.
44. N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Springer Lecture Notes in Mathematics, vol. 1794, 2002. MR1970385 (2004c:37005)
45. M. Rothschild & J. Stiglitz, Increasing risk I: A definition, *J. of Economic Theory*, **2** (1970), 225–243. MR0503565 (58:20284a)

46. K. Schmidt, *Dynamical systems of algebraic origin*, Birkhäuser, Basel, 1995. MR1345152 (97c:28041)
47. 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). MR0462893 (57:2858c)
48. H. G. Schuster, *Deterministic chaos*, 3rd ed., VCH, 1995. MR1393692 (97c:58103)
49. C. Series, The geometry of Markoff numbers, *Math. Intelligencer*, **7** (1985), 20–29. MR795536 (86j:11069)
50. V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Statist.*, **36** (1965), 423–439. MR0177430 (31:1693)
51. Y. L. Tong, *Probability inequalities in multivariate distributions*, Academic Press, 1980. MR572617 (82k:60038)
52. D. van der Vecht, *Inequalities for stopped Brownian motion*, CWI tract 21, Math. Centrum, Amsterdam, 1986. MR831940 (87g:60070)
53. P. Veerman, Symbolic dynamics of order-preserving orbits, *Physica D*, **29** (1987), 191–201. MR923891 (89h:58133)

SCHOOL OF MATHEMATICAL SCIENCES, QUEEN MARY, UNIVERSITY OF LONDON, MILE END ROAD, LONDON, E1 4NS, UK  
*E-mail address*: omj@maths.qmul.ac.uk