

CHARACTERIZATION OF TWO-DISTANCE SEQUENCES

W. F. LUNNON and P. A. B. PLEASANTS

(Received 22 December 1989; revised 5 August 1991)

Communicated by T. E. Hall

Abstract

Three differently defined classes of two-symbol sequences, which we call the *two-distance* sequences, the *linear* sequences and the *characteristic* sequences, have been discussed by a number of authors and some equivalences between them are known. We present a self-contained proof that the three classes are the same (when ambiguous cases of linear sequences are suitably interpreted). Associated with each sequence is a real invariant (having a different appropriate definition for each of the three classes). We give results on the relation between sequences with the same invariant and on the symmetry of the sequences. The sequences are closely related to Beatty sequences and occur as digitized straight lines and quasicrystals. They also provide examples of minimal word proliferation in formal languages.

1991 *Mathematics subject classification* (*Amer. Math. Soc.*): 11 B 99, 05 B 99, 68 R 05, 68 U 05.

Keywords and phrases: Beatty sequences, Fibonacci sequences, symbolic dynamics, chain codes, digitized straight lines, quasicrystals.

1. Introduction and preliminaries

The sequences discussed in this paper are formed from two symbols, which we take to be a and b , and are infinite in both directions. Throughout the paper we shall regard a sequence as being a linear ordering of symbols with no particular origin. In other words, if $t_n = s_{n+k}$ then $\{s_n\}$ and $\{t_n\}$, although translates of each other, are the same sequence.

DEFINITIONS. A *word* of a sequence is a finite string of consecutive symbols from the sequence. The *length* of a word is the total number of symbols it contains. The *weight* (or *b-weight*) of a word is the number of b 's it contains.

© 1992 Australian Mathematical Society 0263-6115/92 \$A2.00 + 0.00

A sequence is *two-distance* if for every length l at most two weights occur for words of length l .

A trivial example of a two-distance sequence is the sequence $(ab)^\infty$ of alternating a 's and b 's. The weight of a word of length n is $n/2$, if n is even, and $(n \pm 1)/2$, if n is odd. A non-trivial example can be derived from the Beatty sequence

$$\mathcal{B} = \{[n\tau]: n \in \mathbb{Z}\} \\ = \dots, -10, -9, -7, -5, -4, -2, 0, 1, 3, 4, 6, 8, 9, \dots,$$

where $\tau = (\sqrt{5} + 1)/2$ is the golden ratio. (A definition of Beatty sequence is given in Section 5.) The characteristic function of this sequence, defined by

$$s(n) = \begin{cases} b & \text{if } n \in \mathcal{B}, \\ a & \text{if } n \notin \mathcal{B}, \end{cases}$$

is the sequence

$$S = \dots bbababbabab\underline{babbababb} \dots,$$

where the underlined term is $s(0)$.

Equivalently, this sequence can be obtained from the straight line $y = \tau x$ by marking an a where the line meets vertical lines of the integer grid and a b where it meets horizontal lines of the grid (with an ab at the origin). This is because there is an a between the b 's on consecutive horizontal grid lines $y = n$ and $y = n + 1$ if and only if $[(n + 1)/\tau] = [n/\tau] + 1$ which, since $\tau = 1 + 1/\tau$, occurs if and only if $[(n + 1)\tau] = [n\tau] + 2$. The sequence S is thus exhibited as a "linear sequence", defined at the beginning of Section 3. Linear sequences also occur in the literature under the names "mechanical sequences" [16], "spectra" [2], [3] and "chain codes" [4], [5].

Table 1 lists the words in S of length up to 5, together with their weights.

TABLE 1

Length	Words	Weights
1	a, b	0, 1
2	ab, ba, bb	1, 2
3	aba, abb, bab, bba	1, 2
4	$abab, abba, baba, babb, bbab$	2, 3
5	$ababb, abbab, babab, babba, bbaba, bbabb$	3, 4

The complete sequence S has a proportion τ of b 's to a 's and a word of length l has weight either $[l/\tau]$ or $[l/\tau] + 1$ (facts which follow easily from the definition of S as a linear sequence). The sequence S , and related sequences with proportion τ of b 's to a 's, occur in Penrose tilings in a geometrical guise (see [17], [11] and [9, Section 10.6]) and are variously known as "worm sequences", "musical sequences", "Fibonacci sequences" or "Wythoff sequences". As well as being the characteristic function of \mathcal{B} , the sequence S can also be generated by iterating the substitution $a \rightarrow ba$, $b \rightarrow bba$, starting from ab . (Essentially because the linear transformation $x \mapsto x + y$, $y \mapsto x + 2y$ fixes the line $y = \tau x$.) The sequence is therefore invariant under this substitution and hence also under the reverse operation $ba \rightarrow a$, $bba \rightarrow b$. An operation of the latter kind is called a "derivation" in the literature. Since it does not change S it can be repeated indefinitely, and consequently S is what is known as a "characteristic" sequence, defined in detail at the beginning of Section 2.

In their Penrose tiling context, worm sequences arise as sequences of intervals of two different lengths in the ratio τ to 1. Intervals are one-dimensional vectors, so we define a *geometric sequence* to be a sequence of vectors (it corresponds to an infinite piecewise-linear path in the space the vectors belong to) and say that a geometric sequence is *two-distance* if for each l there are at most two values for the sum of l consecutive vectors in the sequence. (It was with this situation in mind that we chose the name "two-distance".) It is easily seen that the two-distance geometric sequences are precisely the sequences obtained by assigning distinct vectors to the two symbols of a two-distance sequence of symbols.

The main object of this paper is to give a self-contained, accessible proof that the three classes of sequences mentioned above, namely

- (i) two-distance sequences (defined *via* the spelling of words),
- (ii) linear sequences (defined *via* 2-dimensional geometry),
- (iii) characteristic sequences (defined *via* successive substitution),

are the same. In particular, knowing that the two-distance sequences are the characteristic ones gives a convenient test for the two-distance property.

Each of Sections 2–4 is a link in the triangle of implications that does this. We exploit the geometry as far as we can, particularly in the characteristic-to-linear implication of Section 4. The final five sections give the historical background to these results, describe the alternative "cut-and-project" construction for two-distance sequences (familiar from quasicrystal theory), classify the "species" of locally-identical two-distance sequences, investigate the symmetry of two-distance sequences and suggest further questions.

Although we have not seen the three-way equivalence described above

stated as a unified result, it cannot be said to be new: M. Morse and G. A. Hedlund [16] show that two-distance sequences and linear sequences (in a different guise) are the same, while C. Series [18] defines “cutting sequences” (essentially the same as our linear sequences) and shows they are characteristic.

We end this section with a simple lemma.

LEMMA 1. *Let S be an infinite sequence of a 's and b 's.*

(i) *If S is two-distance then for every length l there is a number $w(l)$ such that every word in S of length l has weight either $w(l)$ or $w(l) + 1$.*

(ii) *If S is not two-distance then there exist l and w such that there are words in S of length l with weights w and $w + 2$.*

PROOF. The weights of two words of the same length in S with adjacent starting points differ by at most one. Hence, for given l , the weights of words of length l in S form a convex set of non-negative integers (that is, the set contains all integers intermediate between any two in the set). Both parts of the lemma follow.

2. Two-distance sequences are characteristic

The a -extension, $S\alpha$, of a sequence S of a 's and b 's is the sequence obtained by replacing each b in S with ab . Only two sequences are invariant under α : the sequence with no b 's and the sequence with just one b . Similarly the b -extension of S , $S\beta$, is obtained by replacing each a in S with ba , and only two sequences are invariant under β . We are principally interested in the inverse operations α^{-1} and β^{-1} . For a sequence S without consecutive b 's the a -reduction of S , $S\alpha^{-1}$, is the sequence obtained by removing the a immediately preceding each b . For a sequence S without consecutive a 's the b -reduction of S , $S\beta^{-1}$, is defined similarly. The sequences invariant under α^{-1} and β^{-1} are precisely those invariant under α and β . Finally, a sequence S is *characteristic* if there is an infinite descending chain of reductions starting from S . Those characteristic sequences whose reductions eventually stabilize to one of the four reduction-invariant sequences form a sub-class whose behaviour is exceptional in many respects. The only sequence to contain neither consecutive a 's nor consecutive b 's is the sequence $(ab)^\infty$ of alternate a 's and b 's, which a -reduces to b^∞ and b -reduces to a^∞ . So the chain of reductions associated with a given characteristic sequence is uniquely determined except for those se-

quences (other than a^∞ and b^∞ themselves) whose reductions eventually stabilize to a^∞ or b^∞ . The exceptional chains bifurcate when the sequence $(ab)^\infty$ is reached, from which point the chain can continue either with the reductions $\alpha^{-1}\beta^{-1}\beta^{-1}\beta^{-1}\dots$ or with $\beta^{-1}\alpha^{-1}\alpha^{-1}\alpha^{-1}\dots$.

This definition of a characteristic sequence is presented slightly differently from that given by Series in [18], but the effect is the same. Suppose the chain of reductions associated with a characteristic sequence S starts with n β^{-1} 's followed by an α^{-1} (which we shall denote by $\beta^{-n}\alpha^{-1}$). Then the b 's in $S\beta^{-n}$ are isolated, and hence every block of consecutive b 's in S has length either n or $n + 1$. Such an S is described in [18] as "almost constant" with "value" n , and $S\beta^{-n}$ is called its "derived sequence" (it differs only marginally from the derived sequence as defined by Morse and Hedlund [16, Section 8]). Consequently, if we write the complete chain of operations associated with S as $\beta^{-n_0}\alpha^{-n_1}\beta^{-n_2}\dots$ (where all n_i are ≥ 1 , except that n_0 is allowed to be 0 when the first operator is α^{-1}) then we can read off the successive values of S and its derivatives as n_0, n_1, n_2, \dots . In fact (as noted in [18]) $[n_0, n_1, n_2, \dots]$ is the regular continued fraction expansion of the proportion of b 's to a 's in S . This sequence of successive values is either infinite or terminates with the value ∞ , according as the proportion of b 's to a 's in S is irrational or rational. When S is associated with two different chains of operations the corresponding sequences of values terminate with $\dots, n + 1, \infty$ and $\dots, n, 1, \infty$, which represent the two regular continued fraction expansions of the same rational number.

LEMMA 2. *Every two-distance sequence is characteristic.*

PROOF. A two-distance sequence S either has no consecutive a 's or no consecutive b 's. (This is the case $l = 2$ of Lemma 1(i).) So S can be reduced, and we can suppose, without loss of generality, that it can be a -reduced. Suppose also that $S\alpha^{-1}$ is not two-distance. Then by Lemma 1(ii) there are two words, W_1 and W_2 , of length l in $S\alpha^{-1}$ with weights w and $w + 2$. Corresponding to W_2 there is a word of length $l + w + 1$ in S with weight $w + 2$ (obtained by inserting an extra a between each pair of b 's in W_2) and corresponding to W_1 there is a word of the same length $l + w + 1$ in S with weight w (obtained by inserting an extra a between each pair of b 's and also to the left of the leftmost b and to the right of the rightmost b). By Lemma 1(i) this is inconsistent with S being two-distance. Hence every two-distance sequence S has a reduction which is also two-distance. So S can be reduced indefinitely and is characteristic.

3. Linear sequences are two-distance

For the classes of linear and characteristic sequences to coincide exactly some minor (and fairly natural) modifications are needed to the definition given in [18] of linear sequence in cases when the associated line passes through an integer point.

Consider the square grid of vertical and horizontal lines through integer points in the plane. On an arbitrary line L of non-negative slope mark the points where it crosses the grid lines and label them a for a vertical grid line and b for a horizontal one. This gives a two-way infinite sequence of a 's and b 's. A sequence that can be obtained from a line L in this way is called *linear*. We need to consider what sequences to assign to a line L that passes through an integer point, thus crossing two grid lines at once. We distinguish three cases.

Case 1: *the line L passes through only one integer point.* In this case we associate with L both the sequences that arise from the two interpretations ab or ba of the integer point.

Case 2: *the line L passes through two integer points but is not parallel to either axis.* In this case L has rational slope and passes through an evenly spaced infinite sequence of integer points. Clearly the finite sequence of a 's and b 's between each successive pair of integer points is the same. We associate three sequences with L .

(i) The sequence got by interpreting every integer point as ab . This is the same as the sequence we would get by displacing L downwards slightly without changing its slope. On displacing L further downwards the sequence of a 's and b 's does not change until the next line of integer points is encountered. Since this line of integer points is congruent to the original one but is approached from above, it follows that if we had interpreted each integer point as ba instead of ab we would have obtained the same sequence (up to translation). It is also the same as the sequence arising from any line parallel to L that avoids integer points. It is periodic.

(ii) The sequence got by interpreting all integer points south-west of a certain point as ab and all north-east of that point as ba . (Altering the change-over point between ab 's and ba 's merely translates the sequence.) In view of (i), this sequence is the result of joining two one-way infinite sequences with the same period but whose ends do not match. In fact (ii) is (i) with a finite block of terms removed. It is not itself periodic because no translation of a periodic sequence can leave all except finitely many terms unchanged (as translation by the period of (i) does for this sequence).

(iii) The sequence got by interpreting all integer points south-west of a certain point as ba and all north-east of that point as ab . The same comments

apply as to (ii), although (ii) and (iii) are different sequences (see below).

Case 3: *the line L passes through an integer point and is parallel to an axis.* If L is parallel to the x-axis we associate with it (i) the sequence consisting entirely of a's and (ii) the sequence with only one b. If it is parallel to the y-axis we associate with it (i) the sequence consisting entirely of b's and (ii) the sequence with only one a.

We have seen that (i) of Case 2 is the result of a small translation of L. The sequences (ii) and (iii) of Case 2 can be interpreted as the results of infinitesimal rotations of L (in opposite directions). (The rotations must be infinitesimal since a finite rotation, no matter how small, changes the sequences fundamentally outside a finite region.) Alternatively, (i), (ii) and (iii) are limits of linear sequences associated with lines that tend to L in three different ways while avoiding integer points. A uniform description of (i), (ii) and (iii) is that they are the "convex" interpretations of the integer points on L: whenever two integer points on L have the same interpretation all intermediate integer points have this interpretation too. One way of seeing that none of these three sequences is a translate of either of the others is to scan the sequences from left to right looking for the first failure of periodicity. With (i) there is no failure, with (ii) the first failure is a b (that would need to be an a to maintain periodicity) and with (iii) the first failure is an a. (Theorem 2(ii) of Section 7 provides another proof that they are different, based on the weights of words whose length is the period length.)

Likewise the sequences (i) and (ii) of Case 3 can be regarded as resulting from an infinitesimal translation or rotation, respectively, of L. They are also limits of linear sequences associated with lines that tend to L either by translation or rotation.

LEMMA 3. *Every linear sequence is two-distance.*

PROOF. Let S be a linear sequence associated with the line L in the plane. Let W be a word in S of length l and weight w and let M be the segment of the line L whose end-points are the points corresponding to the first and last symbols of W. If u is the length of the projection of M on the x-axis then

$$l - w - 1 \leq u \leq l - w + 1$$

with equality possible on the right only when both end-points of M are integer points with the first corresponding to an ab in M and the second to a ba, in which case S must be a sequence of type 2(ii). Similarly, if v is the length of the projection of m on the y-axis then

$$w - 1 \leq v \leq w + 1$$

with equality possible on the right only when S is a sequence of type 2(iii). Dividing these inequalities gives

$$\frac{w - 1}{l - w + 1} \leq \lambda,$$

where λ is the slope of L and equality is possible only when S is of type 2(ii), and

$$\frac{w + 1}{l - w - 1} \geq \lambda$$

provided $w \neq l$, with equality possible only when S is of type 2(iii). (The last inequality must be charitably interpreted when $w = l - 1$.)

Suppose now that S has words of weights w_1 and w_2 with the same length l , where $w_1 < w_2$. Then

$$\frac{w_2 - 1}{l - w_2 + 1} \leq \lambda \leq \frac{w_1 + 1}{l - w_1 - 1},$$

with strict inequality on the left unless S is of type 2(ii) and on the right unless it is of type 2(iii). Since S cannot be of both these types, we have

$$\frac{w_2 - 1}{l - w_2 + 1} < \frac{w_1 + 1}{l - w_1 - 1},$$

which simplifies to

$$w_2 < w_1 + 2.$$

Thus the conclusion of Lemma 1(ii) is impossible for S , so S is two-distance.

4. Characteristic sequences are linear

We start by observing that the concepts of extension and linearity can be defined for words as well as for infinite sequences.

The a -extension, $W\alpha$, of a word W is the word got by replacing each b in W by ab , and the b -extension, $W\beta$, is the word got by replacing each a by ba . (For words in which every b is preceded by an a a -reduction could be defined too, but we shall not need that concept.) If an infinite sequence S of a 's and b 's is decomposed into words, $S = \dots W_{i-1}W_iW_{i+1}\dots$, then $S\alpha = \dots (W_{i-1}\alpha)(W_i\alpha)(W_{i+1}\alpha)\dots$ (and similarly for β). A word is *linear* if it can be obtained from a finite line segment in the plane with non-negative slope by labelling with a 's and b 's the points where it meets vertical and horizontal integer grid lines. With words it is not necessary to use line segments that pass through integer points, because a small enough displacement of a finite

line segment does not change the order of its simple intersections with grid lines.

We shall show that every word of a characteristic sequence is linear, from which the linearity of the complete sequence will follow by a simple compactness argument.

LEMMA 4. *If the word W is linear then so are $W\alpha$ and $W\beta$.*

PROOF. Let M be the line segment that gives rise to W . Since a small displacement of M does not change W , we can suppose that the complete line of which M is a part has strictly positive slope and does not pass through any integer point. The linear mapping ϕ defined by $(x, y)\phi = (x, y + x)$ maps vertical lines to themselves but increases the slope of other lines by 1. So the word derived from the line segment $M\phi$ has the same number of a 's as W but an extra b between each pair of neighbouring a 's. Furthermore if W starts with an a then $M\phi$, extended slightly to the south-west if necessary, gives rise to a b before this a . Hence $W\beta$ is linear. A similar argument, with the rôle of the axes reversed, shows that $W\alpha$ is also linear.

The next lemma is mentioned in [18] without proof.

LEMMA 5. *Every word of a characteristic sequence is linear.*

PROOF. Let S be a characteristic sequence with descending chain of reductions $\gamma_1^{-1}, \gamma_2^{-1}, \gamma_3^{-1}, \dots$, where each γ_i is α or β , and let W be a word of S . Since each γ_i^{-1} can be regarded as an operation of deleting certain symbols, $S_1 = S\gamma_1^{-1}, S_2 = S_1\gamma_2^{-1}, S_3 = S_2\gamma_3^{-1}, \dots$ is a descending chain of subsequences of S . The corresponding chain of subsequences of W is $W_1 = S_1 \cap W, W_2 = S_2 \cap W, W_3 = S_3 \cap W, \dots$, and being a descending chain of finite sequences it must eventually stabilize to a subsequence W_n of W which is unchanged by γ_{n+1}^{-1} and later γ^{-1} 's. The possibilities for W_n (apart from being empty, which is the commonest occurrence) are $aa \dots a, baa \dots a$ and the complementary sequences obtained by interchanging a and b . (It cannot contain ab and ba , nor can it contain aab —making $\gamma_{n+1} = \alpha$ which would change W_n —or bba .) We denote by W_n^* the word of W_n consisting of W_n and its two neighbouring terms. Then the possibilities for W_n^* (apart from aa, ab, ba and bb) are $aa \dots a, abaa \dots a, baa \dots a$ and their complements. (The right-hand neighbour of W_n cannot be a b unless $W_n = a$ and the left-hand neighbour is also a b ; the left-hand neighbour cannot be a b when W_n starts with b .) Each of these possibilities is a subword of a word of the form $aa \dots abaa \dots a$ or its complement. But this word is linear,

since it can be obtained from a long line segment with small slope that meets one horizontal grid line. Repeated application of Lemma 4 now shows that W , which is a subword of $W_n^* \gamma_n \gamma_{n-1} \dots \gamma_1$, is also linear.

Note that W_n is empty unless S is one of the exceptional characteristic sequences whose reductions eventually stabilize.

LEMMA 6. *Every characteristic sequence is linear.*

PROOF. Let S be a characteristic sequence. Choose a term c of S to be the *central term*, and choose a sequence W_1, W_2, W_3, \dots of words with $c \in W_1 \subseteq W_2 \subseteq W_3 \subseteq \dots$ whose union is S . By Lemma 5 each W_i arises from some finite line segment M_i , and by translating M_i by an integer vector we can suppose that the point of M_i corresponding to the central term lies on the y -axis between $y = 0$ and $y = 1$. Let y_i be the y -intercept of M_i and λ_i its slope. Being bounded, the pairs (y_i, λ_i) have at least one point of accumulation (y, λ) . It remains to show that S is the linear sequence obtained from the line L through the point $(0, y)$ with slope λ (or one of the linear sequences obtained from L , in the ambiguous case when L passes through an integer point).

First suppose that L passes through no integer points, and let S' be the sequence associated with L . Consider any W_i , and suppose it has at most n_i symbols each side of the central term. There is some $j \geq i$ such that M_j is so close to the central segment of L that the central $2n_i + 1$ symbols of W_j are the same as the corresponding symbols of S' . It follows that S' contains W_i as a central segment. This is true for all i , and hence $S' = S$.

If L passes through an integer point but is not parallel to either axis then, similarly, for each W_i we can find an M_j (not passing through integer points) whose central segment is sufficiently close to the central segment of L to ensure that W_i arises from the central segment of L with a "convex" interpretation of the integer points in this segment. It follows that S is one of the linear sequences associated with L (since a choice of interpretations of integer points that is everywhere locally convex is globally convex).

Finally, suppose that L is an axis, say the x -axis. Then for each W_i there is an M_j close enough to the x -axis to ensure that W_i has at most one b . Hence S has at most one b . Similarly when L is the y -axis S has at most one a .

We have now proved

THEOREM 1. *For a sequence S on two symbols the following are equivalent:*

- (i) S is two-distance;
- (ii) S is linear;
- (iii) S is characteristic.

5. Relation to other work

A. Markoff. Characteristic sequences occur as continued fractions in the work of A. Markoff on Diophantine approximation [12, 13, 14]. The numbers below 3 in the Markoff spectrum have continued fraction expansions made up of pairs of 1's and pairs of 2's in the form of a periodic characteristic sequence (with 1,1 for a and 2,2 for b). For these periodic characteristic sequences the invariant λ , defined in the next section, is rational. See [7] for a good account of this work.

M. Morse and G. A. Hedlund. In [16] M. Morse and G. A. Hedlund study in depth two-distance sequences (which they call "Sturmian trajectories") and linear sequences (which they call "mechanical sequences" and define without reference to geometry). Their definition of mechanical sequence excludes the sequences with no a 's and with only one a (corresponding to vertical lines). They also define "Sturmian series" and prove ([16, Theorem 7.1]) that they are identical to Sturmian trajectories apart from excluding the same exceptional sequences. In Sections 4 and 5 they prove that mechanical sequences and Sturmian series are the same.

In Section 8 a "derivation" operation is defined that can be applied to any two-distance sequence to produce a new sequence, and the proof of Theorem 8.1 shows that the derived sequence is also two-distance. So Morse and Hedlund show that two-distance sequences can be derived arbitrarily many times and hence are characteristic.

C. Series. In [18] C. Series discusses sequences obtained by laying straight lines across tessellations of Euclidean or hyperbolic space, in particular linear sequences obtained from the square grid in \mathbb{R}^2 (which she calls "cutting sequences"). She defines derivation, in a way very similar to Morse and Hedlund, and calls a sequence "characteristic" if it can be derived arbitrarily many times. She shows that linear sequences are characteristic but notes some simple counter-examples to the converse. The purpose of our elaborate (but, we believe, natural) definition of linear sequence in Section 3, associating more than one sequence with a line through an integer point, was to be able to exhibit all characteristic sequences as linear and so remove these counter-examples.

Series uses the word “linear” for finite cutting sequences and says that characteristic sequences are limits of such finite linear sequences. Our proof of Lemma 6 makes use of this idea.

Beatty sequences. A *Beatty sequence* is a sequence of the form $\{[n\alpha + \beta]\}$ for fixed real numbers α and β . It is also known as a *spectrum* (*homogeneous* if $\beta = 0$ and *nonhomogeneous* if $\beta \neq 0$). Beatty sequences are intimately related to the sequences studied here, since the characteristic function of a Beatty sequence is a linear sequence, as described in Section 1. (For this correspondence to be exact in both directions small modifications would be needed to the definition of Beatty sequence when α and β are rational, paralleling the definition of linear sequence in Section 3 when the line passes through an integer point.) A good bibliography of Beatty sequences (up to the time it was written) can be found in [20]. M. Boshernitzan and A. S. Fraenkel in [3] give an $O(n)$ algorithm for deciding whether a sequence of n integers is part of a Beatty sequence and giving ranges for α and β if so. Their algorithm thus gives a quick way of checking a finite sequence for the two-distance property.

The question of covering the integers by a finite number of disjoint Beatty sequences has been considered by several authors; see [8], [1] and [19].

Computer graphics. Two-distance sequences are identical to the *chain-codes* of *digital straight lines*, a connexion that is highlighted by the strip construction described in Section 6. A. M. Bruckstein discusses the effect of linear transformations on chain-codes in [4] and [5], which contain many references to other work on the graphical representation of straight lines.

The number of words of length n . Non-periodic two-distance sequences are characterized by the property that for every $n > 0$ the words of length n come in precisely two different weights. (When all words of length n have the same weight, n is a period.) Another interesting minimal property they have is that for every $n \geq 0$ there are precisely $n + 1$ different words of length n . (The fact that $n + 1$ words of length n is minimal for non-periodic sequences follows from [15, Theorem 7.3]. For periodic sequences, by contrast, the number of words of length n is bounded above by the period length for all n .) The latter property does not quite characterize non-periodic two-distance sequences, however, as sequences with the property can be constructed by joining the left and right halves of certain pairs of distinct periodic two-distance sequences. E. M. Coven and G. A. Hedlund [6] have fully determined the class of sequences with the property. Within the class of one-directional sequences the property does characterize the non-periodic two-distance ones.

6. An alternative geometric construction

Another geometric construction for two-distance sequences, related to their construction as linear sequences, is as particular cases of the “prism patterns” defined in [17] and [11] (which are themselves particular cases of quasicrystals constructed by the cut-and-project method). Let \mathcal{S} be an infinite strip on the plane that meets the x -axis in an interval of length 1. Then for each integer n there is just one integer point \mathbf{x}_n in \mathcal{S} with y -coordinate n . There are just two possibilities for the difference $\mathbf{x}_{n+1} - \mathbf{x}_n$ as n varies, namely $\mathbf{a} = [1/\lambda]\mathbf{e}_1 + \mathbf{e}_2$ and $\mathbf{b} = ([1/\lambda] + 1)\mathbf{e}_1 + \mathbf{e}_2$, where $\mathbf{e}_1, \mathbf{e}_2$ are the unit coordinate vectors and λ is the slope of the strip. So $\{\mathbf{x}_{n+1} - \mathbf{x}_n\}$ can be regarded as a sequence of two symbols and is, in fact, a two-distance sequence. (This is because for fixed l there are only the two possibilities, $[l/\lambda]\mathbf{e}_1 + l\mathbf{e}_2$ and $([l/\lambda] + 1)\mathbf{e}_1 + l\mathbf{e}_2$, for $\mathbf{x}_{n+l} - \mathbf{x}_n$ as n varies, and when $\mathbf{x}_{n+l} - \mathbf{x}_n$ is known the numbers of \mathbf{a} 's and \mathbf{b} 's that comprise it are known, since \mathbf{a} and \mathbf{b} are linearly independent.) If an integer point lies on an edge of the strip then there is another integer point with the same y -coordinate on the other edge. We use only one of the integer points from each such pair, the permissible choices of an integer point from each pair being the “convex” choices analogous to the permissible interpretations of integer points on the line L described in Section 3. This construction, unlike the linear sequence one, can be used to obtain a one-dimensional geometric two-distance sequence directly, by projecting all the integer points in the strip in the same direction. Any value (including negative ones) for the ratio of the lengths assigned to the symbols a and b can be achieved by a suitable choice of the projection direction.

It is not hard to see that the relation between the sequence S' obtained from the strip \mathcal{S} and the linear sequence S obtained from an edge of \mathcal{S} is that S is obtained from S' by the “inflation” operation $a \rightarrow a^k b$, $b \rightarrow a^{k+1} b$, where $k = [1/\lambda]$. The reverse operation $a^k b \rightarrow a$, $a^{k+1} b \rightarrow b$ by which S' is obtained from S is very similar to the “derivations” of [16, Section 8] and [18]. As a consequence, all two-distance sequences except that consisting entirely of b 's and that with only one a (Morse and Hedlund's exceptional sequences) can be obtained from the strip construction.

7. Species

DEFINITION. Two sequences are of the same *species* if every word of either occurs in the other.

We introduced this term in [17] and [11] in the context of quasicrystals;

though other writers on quasicrystals call this relationship being *locally isomorphic*. In formal language theory a *language* is a set of words, so a sequence gives rise to the language consisting of all words that occur in it. In this setting, two sequences are of the same species when they have the same language: a species is a set of sequences with a common language. The concept extends to a wider geometric context too and is one of a number of parallels between geometric patterns and languages.

The simplest non-trivial example of a species is provided by the worm sequences, mentioned in Section 1, which are the two-distance sequences with invariant τ . Although there are uncountably many of them that are not translationally equivalent, Lemma 7(i), below, shows that they comprise a single species. The species of a periodic sequence, by contrast, consists of its translates only. As we are not distinguishing between translates we regard this as a *singleton* species. Since a sequence that is not two-distance has a word in which the two-distance property fails, all sequences in the species of a two-distance sequence are two-distance.

A two-distance *cum* linear *cum* characteristic sequence S has a non-negative extended real (∞ is allowed) invariant $\lambda = \lambda(S)$ attached to it which has a different appropriate definition for each of the three ways of regarding S . (The definitions are easily seen to be equivalent, of course.)

TWO-DISTANCE DEFINITION. $\lambda(S)$ is the limiting ratio of b 's to a 's in S . The limit exists and is uniform in the sense that the proportion of b 's to a 's in a word of length l tends to λ uniformly as $l \rightarrow \infty$.

LINEAR DEFINITION. $\lambda(S)$ is the slope of the line that gives rise to S .

CHARACTERISTIC DEFINITION. $\lambda(S)$ is the value of the continued fraction $[n_0, n_1, n_2, \dots]$ described in Section 2, where n_i represents the number of reductions of the same type performed on S between the i th and $(i + 1)$ th alternations of reduction type.

It is clear from the first of these definitions (and also from the last) that λ is a species invariant: sequences in the same species have the same value of λ . Conversely, irrational values of λ determine the species uniquely, as we shall see. The second definition of λ is the most convenient for proving results in this direction.

LEMMA 7. (i) *If λ is irrational then the linear sequences with invariant λ form a single species.*

(ii) *Let S_1 and S_2 be linear sequences that differ in a finite number of terms. Suppose they are given by lines L_1 and L_2 such that $\mathbf{x}_1 \in L_1$ and $\mathbf{x}_2 \in L_2$ are grid-crossing points that give rise to corresponding equal terms of S_1 and S_2 . If $\lambda(S_1)$ is irrational then $\mathbf{x}_1 - \mathbf{x}_2 \in \mathbb{Z}^2$.*

PROOF. (i) We have already remarked that every sequence in the species of a two-distance is itself two-distance (and therefore linear) and that linear sequences of the same species have the same invariant. Let S_1 and S_2 be two linear sequences arising from lines L_1 and L_2 of slope λ . We need to show that every word W in S_1 occurs in S_2 . By extending W in both directions, if necessary, there is no loss of generality in assuming that W begins and ends with an a . (There are infinitely many a 's in S_1 because, being irrational, $\lambda \neq \infty$, so L_1 meets infinitely many vertical grid lines.) Then the segment M_1 of L_1 that gives rise to W has both ends on vertical grid lines. There is a short interval I such that the translated line segment $M_1 + (0, t)$ gives rise to the word W for all $t \in I$. (When there is no integer point on M_1 we can take $0 \in I$ and choose I small enough so that the parallelogram $P = \{x + (0, t) : x \in M_1, t \in I\}$ contains no integer point. When there is an integer point on M_1 we can take 0 as the left or right end-point of I (depending on whether the integer point is interpreted as ba or ab in S_1) and choose I small enough so that P contains no other integer point. Since L_1 has irrational slope there cannot be more than one integer point on M_1 .) The word W now certainly occurs in S_2 if there is a segment M_2 of L_2 that is the translate by an integer vector of $M_1 + (0, t)$ for some $t \in I$. This last statement is equivalent to the statement that there is an integer n such that $n\lambda$ lies in a certain interval modulo 1, and this in turn is a consequence of the well known result in Diophantine approximation that the fractional parts $\{n\lambda\}$ are uniformly distributed modulo 1 when λ is irrational. (In fact Kronecker's existence result, see [10, Chapter XXIII, Theorem 438], is enough for our purposes here.)

(ii) If S_1 and S_2 differ in only finitely many terms they have the same invariant (by the two-distance definition) so L_1 and L_2 have the same slope (by the linear definition). Without loss of generality the equal terms of S_1 and S_2 that correspond to x_1 and x_2 are both a 's. (If not, reflect the L 's in the line $x = y$.) Then x_1 and x_2 lie on vertical grid lines. Since translating a line by an integer vector does not change the corresponding sequence, we can also suppose, by adding integer vectors to them, that they both lie on the y -axis between the points 0 (inclusive) and $(0, 1)$ (exclusive). We show that with this normalization $x_1 = x_2$. If not, say x_1 were below x_2 on the y -axis, then the Diophantine approximation argument used in (i) would show that infinitely many vertical grid lines had their intercepts with L_1 and L_2 separated by an integer point and infinitely many did not. Consequently there would be infinitely many integers $i > 0$ such that the intercepts of L_1 and L_2 with $x = i$ were separated by an integer point (i, j) but the intercepts with $x = i - 1$ were not. For each such i the term of S_1 that was $i + j - 1$ places to the right of the a arising from x_1 would be an a , but the

corresponding term of S_2 would be a b . These infinitely many mismatches between S_1 and S_2 would contradict the hypothesis of (ii).

We can now classify two-distance sequences into species.

THEOREM 2. (i) *For irrational λ the two-distance sequences with invariant λ form a single species.*

(ii) *For rational $\lambda \neq 0$ or ∞ each of the three two-distance sequences with invariant λ forms a singleton species.*

(iii) *For $\lambda = 0$ or ∞ each of the two two-distance sequences with invariant λ forms a singleton species.*

PROOF. Part (i) is the same as (i) of Lemma 7. When $\lambda = B/A$ with $A, B \neq 0$ the three two-distance sequences with invariant λ can be shown to belong to different species by considering the weights of words of length $A + B$. Every word of length $A + B$ in the periodic sequence is a complete period, so has weight B . Every word of length $A + B$ in the sequence of invariant λ described in Case 2(ii) of Section 3 also occurs in the periodic sequence of invariant λ (so has weight B) except for the word from the b of the last integer point ab before the change-over point to the b of the first integer point ba after it. This word has weight $B + 1$. Similarly, all words of length $A + B$ in the sequence of Case 2(iii) of Section 3 have weight B except for one with weight $B - 1$. Consequently the three sequences of (ii) belong to different species distinguished by whether words of length $A + B$ all have weight B , have weights B and $B + 1$, or have weights B and $B - 1$. Part (iii) is transparent.

8. Symmetry

This section identifies all symmetries and near-symmetries of two-distance sequences. We classify mirror symmetries as *a-centred*, *b-centred* or *gap-centred* (the latter being a symmetry whose centre falls between two terms of the sequence, which are therefore both *a*'s or both *b*'s). We also consider *near-symmetries*, where instead of every term in the sequence matching its mirror image a finite number of mismatches are allowed.

First we look at a two-distance sequence S with irrational invariant. Let its associated line be L . Clearly the mirror image of S is associated with $-L$, the reflexion of L in $\mathbf{0}$. If S is symmetric or near-symmetric the linear sequences derived from L and $-L$ differ in at most finitely many terms, so corresponding grid-crossing points of L and $-L$ differ by an integer vector

\mathbf{x} , by Lemma 7(ii). Hence $-L = L - \mathbf{x}$, so $(1/2)\mathbf{x} \in L$. It is also clear that $(1/2)\mathbf{x}$ is the centre of the symmetry on L and that the type of symmetry S has depends on which of the four possible residue classes modulo \mathbb{Z}^2 this half-integer point belongs to. Explicitly, if $(1/2)\mathbf{x} \equiv (0, 1/2) \pmod{\mathbb{Z}^2}$ lies on a vertical grid line then S is a -centred, if $(1/2)\mathbf{x} \equiv (1/2, 0)$ lies on a horizontal line then S is b -centred, if $(1/2)\mathbf{x} \equiv (1/2, 1/2)$ then S is gap-centred and finally if $(1/2)\mathbf{x} \equiv (0, 0)$ is itself an integer point then S has a gap-centred near-symmetry in which only the central pair of terms fail to match. Conversely, it is clear that a line with irrational slope through a half-integer point gives rise to a two-distance sequence of the corresponding symmetry type. Since lines through points congruent modulo \mathbb{Z}^2 give rise to identical sequences and lines of irrational slope cannot pass through two half-integer points, we have

THEOREM 3. *Each species of two-distance sequences with irrational invariant contains precisely three symmetric sequences, one of each symmetry type, and precisely two near-symmetric sequences. The near-symmetric sequences are gap-centred and differ only in the central pair of terms which is ab in one and ba in the other and hence is a mismatch in both sequences. This central pair is the only mismatch in either sequence. Each of these five sequences has only one centre of symmetry or near-symmetry.*

In [11, Convention, p. 222] we made a case for not distinguishing between the two near-symmetric sequences.

We now turn to two-distance sequences with rational invariant. Let L be the line through $\mathbf{0}$ with rational slope B/A in its lowest terms. The integer points on L are $\{(nA, nB) : n \in \mathbb{Z}\}$ and the half-integer points are $\{(nA/2, nB/2) : n \in \mathbb{Z}\}$. The pattern of grid-crossings between $\mathbf{0}$ and (A, B) repeats indefinitely and gives the period of the periodic two-distance sequence derived from L , so the period has length $A+B$ with A a 's and B b 's. Since A and B are coprime this is the shortest period. The half-integer point $(A/2, B/2)$ lies on a vertical grid line if and only if A is even and on a horizontal one if and only if B is even. Let \mathbf{h} be a half-integer point not congruent to $\mathbf{0}$ or $(A/2, B/2)$ modulo \mathbb{Z}^2 . Then $\mathbf{h} + (A/2, B/2)$ belongs to the remaining residue class of half-integer points modulo \mathbb{Z}^2 . The line L' through \mathbf{h} and $\mathbf{h} + (A/2, B/2)$ is parallel to L and contains no integer points. It gives rise to the same periodic two-distance sequence that L does (cf., the comments accompanying Case 2(i) in Section 3) and plainly has two centres of symmetry \mathbf{h} and $\mathbf{h} + (A/2, B/2)$ separated by exactly half a period. There are two corresponding centres of symmetry in every period, of course, and since the period we have found is minimal there can be no other centres

of symmetry. A periodic sequence cannot have a near-symmetry since one mismatch implies an infinity of them, with separations increasing in steps of twice the period. Since \mathbf{h} and $\mathbf{h} + (A/2, B/2)$ are distinct modulo \mathbb{Z}^2 from each other and from $(A/2, B/2)$, when A is odd one of them is a -centred, when B is odd one is b -centred and when AB is even one is gap-centred.

The non-periodic sequences derived from L are near-periodic in the sense that shifting by a period changes only four terms. If $\lambda \neq 0$ or ∞ and we arrange that the change in interpretation of integer points occurs between $\mathbf{0}$ and (A, B) then clearly $(A/2, B/2)$ is a centre of exact symmetry and its translates by multiples of (A, B) are centres of near-symmetry. The points $\mathbf{0}$ and (A, B) are special centres of gap-centred near-symmetry with only the central pair of terms not matching. The other integer points on the line are also centres of near-symmetry. Since the above centres of near-symmetry are separated by half a minimum period they comprise all centres of symmetry and near-symmetry. (The product of two near-symmetries is a near-translation by twice the distance between the centres.) The near-symmetry centres fall into two families. The centres in each family have period spacing (that is, they are $A + B$ terms apart) and are all of the same type. Notice that owing to the temporary break in periodicity in the vicinity of the exact symmetry centre the members of a family to the left and right of this centre nevertheless fall into different places in the period: they fall at symmetrically opposite places, in fact. The behaviour of the two non-periodic two-distance sequences with invariants 0 and ∞ is transparent. We have

THEOREM 4. *Let S be a two-distance sequence with rational invariant B/A (in its lowest terms).*

(i) *If S is periodic then it is symmetric and has two centres of symmetry per period. One of these symmetries is a -centred if and only if A is odd; one is b -centred if and only if B is odd.*

(ii) *If S is non-periodic and A and B are non-zero then S has one exact symmetry, which is a -centred if A is even, b -centred if B is even and gap-centred otherwise (that is, it has whichever of the symmetry types does not occur in (i)). The centre of this symmetry is one of a family of centres of near-symmetry of the same type spaced at intervals of $A+B$ terms throughout the sequence. Midway between these centres are the centres of a family of gap-centred near-symmetries. The near-symmetries whose centres are closest each side of the exact symmetry centre are of the type occurring in Theorem 3, with only the central pair of terms failing to match.*

(iii) *The sequence with one a and the sequence with one b have every term and every gap a centre of near-symmetry. The odd term out is the only centre of exact symmetry.*

COROLLARY. *If S is a two-distance sequence and W is a word of S then the reverse of W is also a word of S .*

Morse and Hedlund noted the symmetry of non-periodic two-distance sequences with rational invariant (called by them “skew” Sturmian trajectories) in [16, Theorem 3.6, Corollary].

9. What sequences are three-distance?

We have made little progress towards classifying f -distance sequences for any $f > 2$, but we offer the following observations, omitting proofs (which are straightforward).

When $f > 2$ there is a distinction between *combinatorial f -distance sequences* (sequences of symbols such that for each l there are at most f words of length l such that no two are anagrams) and *geometric f -distance sequences* (sequences of vectors or numbers such that for each l there are at most f possibilities for the sum of l consecutive terms). For $f = 2$ we noted in Section 1 that these concepts are equivalent provided we assign different values in the geometric sequence to the two symbols in the combinatorial one. (Otherwise the geometric sequence would degenerate to being one-distance, of course.) For $f > 2$, however, less trivial relations between values can cause a combinatorial sequence that is not f -distance to degenerate to a geometric one that is. Moreover, the f -distance geometric sequence may not be obtainable from any f -distance combinatorial one at all. There is an example in the next paragraph.

For $d > 2$, a line L in d -dimensional space not lying in any rational hyperplane gives a linear sequence on d symbols that is not d -distance. In fact it can be shown not to be f -distance for any $f < 2^{d-1}$. Probably it is 2^{d-1} -distance, though we cannot prove it. With the generalization of the alternative geometric construction described in Section 6, a prism whose cross-section on one coordinate hyperplane is a fundamental region for the integer lattice in that hyperplane gives rise to a sequence on 2^{d-1} vectors that is geometrically 2^{d-1} -distance but is not geometrically f -distance for any $f < 2^{d-1}$. Such a sequence is not combinatorially 2^{d-1} -distance, however. For example, in \mathbb{R}^3 if the prism has cross-section $\{(x, y) \mid 0 \leq x, y < 1\}$ on the (x, y) -plane and axis parallel to $(\lambda, \mu, 1)$, where $0 < \lambda, \mu < 1/2$, then the four vectors of which the sequence is composed are $\mathbf{a} = (0, 0, 1)$, $\mathbf{b} = (1, 0, 1)$, $\mathbf{c} = (0, 1, 1)$ and $\mathbf{d} = (1, 1, 1)$. However, the five pairs of

vectors \mathbf{aa} , \mathbf{ab} , \mathbf{ac} , \mathbf{ad} and \mathbf{bc} all occur as adjacent terms in the sequence. So the sequence is not combinatorially four-distance even though it is geometrically four-distance. (It is the relation $\mathbf{a} + \mathbf{d} = \mathbf{b} + \mathbf{c}$ that makes this possible.)

ACKNOWLEDGEMENT. We are grateful to one of the referees for many helpful comments that improved our presentation of this work and drew our attention to related work we were unaware of.

References

- [1] M. A. Berger, A. Felzenbaum and A. S. Fraenkel, 'Disjoint covering systems of rational Beatty sequences', *J. Combin. Theory Ser. A* **42** (1986), 150–153.
- [2] M. Boshernitzan and A. S. Fraenkel, 'Nonhomogeneous spectra of numbers', *Discrete Math.* **34** (1981), 325–327.
- [3] M. Boshernitzan and A. S. Fraenkel, 'A linear algorithm for nonhomogeneous spectra of numbers', *J. Algorithms* **5** (1984), 187–198.
- [4] A. M. Bruckstein, 'The self-similarity of digital straight lines', Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, 1990.
- [5] A. M. Bruckstein, 'Self-similarity properties of digitized straight lines', Technical Report no. 616, Dept. of Computer Science, Technion, Israel Institute of Technology, 1990.
- [6] E. M. Coven and G. A. Hedlund, 'Sequences with minimal block growth', *Math. Systems Theory* **7** (1973), 138–153.
- [7] T. W. Cusick and M. E. Flahive, *The Markoff and Lagrange spectra* (Mathematical surveys and monographs no. 30, Amer. Math. Soc., Providence, Rhode Island, 1989).
- [8] R. L. Graham, 'Covering the positive integers by sets of the form $\{n\alpha + \beta\}$: $n = 1, 2, \dots$ ', *J. Combin. Theory Ser. A* **15** (1973), 354–358.
- [9] B. Grünbaum and G. C. Shephard, *Tilings and patterns* (Freeman, San Francisco, 1986).
- [10] G. H. Hardy and E. M. Wright, 'An introduction to the theory of numbers', (4th ed., Clarendon Press, Oxford, 1960).
- [11] W. F. Lunnon and P. A. B. Pleasants, 'Quasicrystallographic tilings', *J. Math. Pures Appl.* **66** (1987), 217–263.
- [12] A. Markoff, 'Sur les formes binaires indéfinies', *Math. Ann.* **15** (1879), 381–406.
- [13] A. Markoff, 'Sur les formes binaires indéfinies', *Math. Ann.* **17** (1880), 379–400.
- [14] A. Markoff, 'Sur une question de Jean Bernoulli', *Math. Ann.* **19** (1882), 27–36.
- [15] M. Morse and G. A. Hedlund, 'Symbolic dynamics', *Amer. J. Math.* **60** (1938), 815–866.
- [16] M. Morse and G. A. Hedlund, 'Symbolic dynamics II. Sturmian trajectories', *Amer. J. Math.* **62** (1940), 1–42.
- [17] P. A. B. Pleasants, 'Quasicrystallography: some interesting new patterns' in *Elementary and analytic theory of numbers*, (Banach Center Publications, Vol. 17, PWN-Polish Scientific Publishers, Warsaw, 1985), 439–461.
- [18] C. Series, 'The geometry of Markoff numbers', *Math. Intelligencer* **7** (3) (1985), 20–29.
- [19] R. J. Simpson, 'Disjoint covering systems of rational Beatty sequences', *Discrete Math.* (to appear).
- [20] K. B. Stolarsky, 'Beatty sequences, continued fractions and certain shift operators', *Canad. Math. Bull.* **19** (1976), 473–482.

References added in proof

- [21] F. Mignosi, 'On the number of factors of Sturmian words', *Theoretical Computer Science* **82** (1991), 71–84.
- [22] P. Séébold, 'Fibonacci morphism and Sturmian words', *Theoretical Computer Science* **88** (1991), 365–384.

Department of Computer Science
St. Patrick's College
Maynooth
County Kildare
Eire

School of Mathematics, Physics
Computing and Electronics
MacQuarie University
NSW 2109
Australia