A Two Dimensional Steinhaus Theorem.

J. F. Geelen and R. J. Simpson

School of Mathematics and Statistics
Curtin University of Technology
Perth, Australia.

Abstract

We prove the following theorem: if α and β are real numbers, P and Q are positive integers then the set $\{\{i\alpha + j\beta\} : i = 0, \ldots, P-1; j = 0, \ldots, Q-1\}$ partitions the unit interval into subintervals having at most $P + 3$ distinct widths. This result has applications in the theory of Beatty sequences and implies a 2-dimensional version of Slater's Theorem.

1. Introduction

H. Steinhaus conjectured the following result.

The Three Gap Theorem Let α be real and N a positive integer. The points

$$\{\{i\alpha\} : i = 0, \ldots, N-1\}$$

(where, as usual $\{x\}$ denotes the fractional part of x) partition the unit interval into N subintervals which have at most 3 distinct widths.

In 1958 this was proved independently by S. Świerczkowski [15] and by P. Erdös and V.T. Sós [12],[13]. It is often called the Steinhaus theorem or Steinhaus Conjecture. A useful way of viewing the result is to think of a circle of unit circumference with points placed around the perimeter at distances $0, \alpha, \ldots, (N-1)\alpha$ from an arbitrary origin on the perimeter. Then the distances between adjacent points take at most 3 distinct values.

The result has close connections with the theory of continued fractions, Beatty sequences and Diophantine approximation. Other proofs have been published since 1958; see, for instance, [10] and its list of references.

In 1972 R.L. Graham [2],[7] conjectured the following variation of the Steinhaus problem.

The 3d Distance Theorem Let α and $\beta_1, \ldots, \beta_d$ be real numbers and $N_1, \ldots, N_d$ be positive integers. The points $\{\{i_k\alpha + \beta_k\} : i_k = 0, \ldots, N_k - 1;\ k = 1, \ldots, d\}$ partition the unit interval into subintervals having at most $3d$ distinct widths.

Australasian Journal of Combinatorics 8(1993), pp.169-197

A long and difficult proof of this result was obtained by F.R.K. Chung and R.L. Graham [1] in 1976. A very neat proof was found by F.M. Liang [8] three years later. In section 2 we will use ideas from Liang's proof to prove the following two dimensional version of the Steinhaus Theorem.

The P+3 Theorem Let α, β be real numbers and P, Q be positive integers. The points

$$\{\{i\alpha + j\beta\} : i = 0, \ldots, P-1,\ j = 0, \ldots, Q-1\}$$

partition the unit interval into $P\,Q$ subintervals having at most $P+3$ distinct widths.

Note that if either $P = 1$ or $Q = 1$ this theorem follows from the Three Gap Theorem. We will assume from now on that $P > 1$ and $Q > 1$.

As with the Three Gap and $3d$ Distance Theorems it is helpful to think of the points as being placed on a circle of unit circumference at a distance of $i\alpha + j\beta$ from some origin.

This result was conjectured by Ron Holzman and the second author in 1991 in connection with work on intersecting Beatty sequences. Holzman showed that the result held with $P+3$ replaced by $P+22$, Geelen and Simpson [5],[6] showed it held with $P+6$ and, later, with $P+4$. Holzman also showed that the bound $P+3$ cannot be improved for $P > 1$. We will demonstrate this in section 3.

The connection with Beatty sequences was investigated by Fraenkel and Holzman [4]. We briefly describe some of the results of their paper.

A Beatty sequence $S(\alpha, \beta)$ is a sequence $\{\lfloor n\alpha + \beta \rfloor : n \in \mathrm{Z}_0\}$. Such sequences have a large literature (see for instance [14] and [3] and their lists of references) and are connected with interesting unsolved problems. Their characteristic sequences are called Sturmian sequences or Sturmian words and have applications in computer science [9]. Fraenkel and Holzman showed that the following result would follow from the $P+3$ Theorem.

Intersecting Beatty Sequences Theorem Let $S(\alpha, \beta)$ and $S(R/P, \gamma)$ be two beatty sequences with R, P being positive integers, $(R, P) = 1$. Let $n_0 < n_1 < n_2, \ldots$ be the sequence of integers in the intersection of the two sequences. Then the set of values $n_{i+1} - n_i$ has cardinality at most $P+3$.

Since we will prove the $P+3$ theorem this result has the status of a theorem.

Fraenkel and Holzman also considered the following problem. Let α be real and r, s be positive integers with $(r, s) = 1$, and consider the sequence $\Theta(\alpha, r/s) = \left\{\left(\{n\alpha\}, \left\{n\frac{r}{s}\right\}\right) : n \in \mathrm{Z}_0\right\}$. This is a sequence of points in the unit square. Suppose that $0 \le x_1 < x_2 \le 1$, $0 \le y_1 < y_2 \le 1$ and let $R = R(x_1, x_2, y_1, y_2)$ be the rectangle in the unit square defined by

$$R(x_1, x_2, y_1, y_2) = \{(x,y) : x_1 < x \leq x_2; y_1 < y \leq y_2\}.$$

Then let $n_1 < n_2 < \ldots$ be the indices of those members of $\Theta(\alpha, r/s)$ which lie in R. Finally let

$$P(\Theta, R) = \left|\left\{i \in \mathrm{Z} : y_1 < \frac{i}{s} \leq y_2\right\}\right|.$$

Fraenkel and Holzman showed that the $P+3$ Theorem implies the following.

The Two-dimensional Slater Problem With the notation just described the set of values $n_{i+1} - n_i$ has cardinality at most $P(\Theta, R) + 3$.

The original (one-dimensional) Slater Theorem [11] concerned the sequence $\{\{i\alpha\} : n \in \mathrm{Z}_0\}$ and a subinterval of the unit interval.

2. The P+3 Problem

Notation: Define the set

$$S = \{\{i\alpha + j\beta\} : i = 0, \ldots, P-1, j = 0, \ldots, Q-1\}.$$

Let the **point** (i,j), where $0 \leq i < P$ and $0 \leq j < Q$, label the element $\{i\alpha + j\beta\}$ from S. We often think of points as positioned on the perimeter of a circle of unit circumference, at a clockwise distance of $\{i\alpha + j\beta\}$ from an arbitrary starting point.

We assume that there are no coincident points in S. This assumption is made with no loss of generality, since if S contained coincident points an arbitrarily small change in α and β would separate the points and not reduce the number of gap sizes.

We define **vectors** (δ, γ), where $-P < \delta < P$ and $-Q < \gamma < Q$, to be the vector differences between points (e.g. the vector distance between the points (i,j) and (m,n) is $(m-i, n-j)$). Although vectors and points share common notation, the distinction will usually be clear from the context. We allow points and vectors to be used together under rules of vector addition:

$$(m,n) - (i,j) = (m-i, n-j).$$

We define the **length** of a vector (δ, γ) to be

$$|(\delta, \gamma)| = \{\delta\alpha + \gamma\beta\}.$$

The **distance** between points (i,j) and (m,n) is defined

$$|(m,n) - (i,j)|.$$

This distance is in fact the clockwise distance along the perimeter from (i, j) to (m, n).

Note that if (i, j) is an element of S then so is

$$(P-1-i, Q-1-j);$$

we write

$$\mathcal{R}(i, j) = (P-1-i, Q-1-j),$$

and say $\mathcal{R}(i, j)$ is the **reflection** of (i, j). It is easy to see that for any point (i, j) we have,

$$\mathcal{R}(\mathcal{R}(i, j)) = (i, j),$$

hence reflection induces a pairing of the points in S. Now consider the positions of a point (i, j) and its reflection on the perimeter, the distance from $\mathcal{R}(i, j)$ to $(P-1, Q-1)$ is

$$\begin{aligned} |(P-1, Q-1) - \mathcal{R}(i, j)| &= |(P-1, Q-1) - (P-1-i, Q-1-j)| \\ &= |(i, j)| \\ &= |(i, j) - (0, 0)|, \end{aligned}$$

which is clearly the same as the distance from $(0, 0)$ to (i, j), as shown by figure 1. Hence the points are symmetrically disposed about the perimeter of the circle.

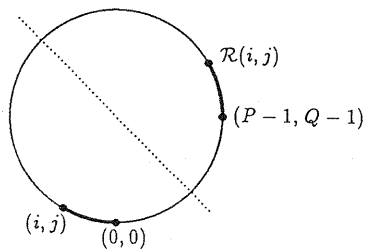

Figure 1: Reflections of points

For each point (i, j) let $\mathcal{T}(i, j)$ to be the element of $S \setminus \{(i, j)\}$ which minimises $|\mathcal{T}(i, j) - (i, j)|$. Thus $\mathcal{T}(i, j)$ is the point immediately clockwise of (i, j) on the perimeter. The set of values $\mathcal{T}(i, j)$ form a matrix which we call the **successor table** of S. $\mathcal{T}$ has **rows**

$$[\mathcal{T}(i, 0), \mathcal{T}(i, 1), \ldots, \mathcal{T}(i, Q-1)]$$

and **columns**

$$\begin{bmatrix} \mathcal{T}(0, j) \\ \mathcal{T}(1, j) \\ \vdots \\ \mathcal{T}(P-1, j) \end{bmatrix}.$$

We write $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ to denote $\mathcal{T}(i,j) = (m,n)$, $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ is called a **link**, and the vector $(m,n)-(i,j)$ is called a **gap vector,** the **width** or **gap size** of a link is the length of its gap vector. If $\mathcal{L}$ is a set of links then we define $\|\mathcal{L}\|$ as the number of distinct gap sizes in $\mathcal{L}$. We let T be the set of all links, and hence are required to prove that $\|T\| \leq P+3$.

The next lemma says that if two points are adjacent then their reflections are adjacent; this is intuitively obvious.

Lemma 1 if $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ then $\mathcal{R}(m,n) \xrightarrow{\mathcal{T}} \mathcal{R}(i,j)$.

PROOF: This follows easily from the symmetry in the distribution of the points in S and their reflections. □

Clearly a link and its reflection have the same gap vector (and hence the same width).

Notation: We define a **chain** to be a sequence of consecutive links in a row of $\mathcal{T}$ sharing a common gap vector,

$$(i,j) \xrightarrow{\mathcal{T}} (m,n), (i,j+1) \xrightarrow{\mathcal{T}} (m,n+1), \ldots, (i,j+k) \xrightarrow{\mathcal{T}} (m,n+k). \quad (1)$$

Clearly every link is in some chain, so the set of chains forms a partition of the links.

If $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ but it is not the case that $(i,j-1) \xrightarrow{\mathcal{T}} (m,n-1)$ then the link $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ is called the **start** of a chain. If $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ is the start of a chain then either:

- $j=0$ or $n=0$, hence we cannot have $(i,j-1) \xrightarrow{\mathcal{T}} (m,n-1)$. Let S_N be the set of chains with this type of start, such chains are said to have **natural** starts.
- $j>0$ and $n>0$ but there exists a point $(p,q) \neq (m,n-1)$ such that $(i,j-1) \xrightarrow{\mathcal{T}} (p,q)$. Let S_T denote the set of chains with this type of start, such chains are said to have **terminated** starts.

Similarly, if $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ but it is not the case that $(i,j+1) \xrightarrow{\mathcal{T}} (m,n+1)$ then we call $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ the **end of a chain.** If $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ is the end of a chain then either:

- $j=Q-1$ or $n=Q-1$. Let E_N be the set of chains with this type of end, such chains are said to have **natural** ends.
- $j<Q-1$ and $n<Q-1$ but there exists a point $(p,q) \neq (m,n+1)$ such that $(i,j+1) \xrightarrow{\mathcal{T}} (p,q)$. Let E_T be the set of chains with this type of end, such chains are said to have terminated ends.

We define the **reflection** of the chain (1) to be

$$\mathcal{R}(m,n+k) \xrightarrow{\mathcal{I}} \mathcal{R}(i,j+k), \mathcal{R}(m,n+k-1) \xrightarrow{\mathcal{I}} \mathcal{R}(i,j+k-1), \ldots, \mathcal{R}(m,n) \xrightarrow{\mathcal{I}} \mathcal{R}(i,j);$$

clearly the reflection of a chain is also a chain.

Lemma 2 Reflection induces a bijection from S_N onto E_N and from S_T onto E_T.

PROOF: Suppose $(i,j) \xrightarrow{\mathcal{I}} (m,n)$ is the start of a chain, its reflection is

$$(P-1-m, Q-1-n) \xrightarrow{\mathcal{I}} (P-1-i, Q-1-j),$$

clearly $j = 0$ if and only if $Q-1-j = Q-1$ and $n = 0$ if and only if $Q-1-n = Q-1$; So $(i,j) \xrightarrow{\mathcal{I}} (m,n)$ is the start of a chain in S_N if and only if its reflection is the end of a chain in E_N.

The reflection of any chain in S_T cannot be a member of E_N, and hence is a member of E_T.

□

Given a set $\mathcal{C}$ of chains let $\mathcal{R}(\mathcal{C})$ be the set of chain reflections; then lemma 2 implies that

- $\mathcal{R}(S_N \cap E_N) = S_N \cap E_N$,
- $\mathcal{R}(S_N \cap E_T) = S_T \cap E_N$,
- $\mathcal{R}(S_T \cap E_N) = S_N \cap E_T$ and
- $\mathcal{R}(S_T \cap E_T) = S_T \cap E_T$.

Note that the second and third items are saying the same thing since $\mathcal{R}(\mathcal{R}(\chi)) = \chi$, where χ is a point, link or chain.

Since $\mathcal{R}(S_N \cap E_N) = S_N \cap E_N$ and $\mathcal{R}(S_T \cap E_T) = S_T \cap E_T$ then it is possible for a chain from $S_N \cap E_N$ or $S_T \cap E_T$ to be the reflection of itself Such chains are said to be **symmetric.**

Lemma 3 There are at most 4 symmetric chains.

PROOF: We say the **midpoint** of a link is the midpoint of the arc it represents; so the midpoint of $(i,j) \xrightarrow{\mathcal{I}} (m,n)$ is either the mean of the ends, ie

$$\left\{\frac{(i\alpha+j\beta)+(m\alpha+n\beta)}{2}\right\} = \left\{\frac{(i+m)\alpha}{2} + \frac{(j+n)\beta}{2}\right\},$$

or diametrically opposite the mean

$$\left\{\frac{(i+m)\alpha}{2}+\frac{(j+n)\beta}{2}+\frac{1}{2}\right\}.$$

If we have a symmetric chain

$$(i,j)\xrightarrow{\mathcal{T}}(m,n),(i,j+1)\xrightarrow{\mathcal{T}}(m,n+1),\ldots,(i,j+k)\xrightarrow{\mathcal{T}}(m,n+k)$$

then, clearly, $i=P-1-m$ and $j=Q-1-n-k$. Thus each link in the chain has the form

$$(i,j+\ell)\xrightarrow{\mathcal{T}}(P-1-i,Q-1-j-k+\ell),$$

where $\ell\in[0,\ldots,k]$; when $\ell=\left\lfloor\frac{k}{2}\right\rfloor$ the midpoint of the link is either

$$\left\{\frac{(P-1)\alpha}{2}+\frac{(Q-1-\delta)\beta}{2}\right\}$$

or

$$\left\{\frac{(P-1)\alpha}{2}+\frac{(Q-1-\delta)\beta}{2}+\frac{1}{2}\right\},$$

where $\delta=k-2\ell$, thus δ is either 0 or 1 if k is even or odd respectively. Therefore there are 4 values which are candidates to be the midpoint of such a link; since no two links can have the same midpoint there are at most 4 such links and hence at most 4 symmetric chains. □

Every row of $\mathcal{T}$ is a sequence of chains, we call the first chain in a row a **left justified** chain, the start of a left justified chain clearly has the form $(i,0)\xrightarrow{\mathcal{T}}(m,n)$, thus all left justified chains are members of S_N. Similarly we call the last chain in a row **right justified**, the end of a right justified chain has the form $(i,Q-1)\xrightarrow{\mathcal{T}}(m,n)$, hence such chains are members of E_N.

Lemma 4 Every chain from $S_N\cap E_N$ is either left or right justified.

PROOF: Consider a chain

$$(i,j)\xrightarrow{\mathcal{T}}(m,n),(i,j+1)\xrightarrow{\mathcal{T}}(m,n+1)\ldots(i,j+k)\xrightarrow{\mathcal{T}}(m,n+k),$$

from $S_N\cap E_N$, either $j=0$ or $n=0$, and either $j+k=Q-1$ or $n+k=Q-1$. If the chain is not left justified then we have $n=0$ and $j>n$, this implies that $j+k>n+k$ and hence $j+k=Q-1$; thus the chain is right justified. □

Lemma 5

(a) If $\chi\in S_T$ is in a row r of $\mathcal{T}$ then every chain preceding χ in r is either from $S_T\cap E_N$ or is a left justified from $S_N\cap E_N$.

(b) If $\chi \in E_T$ is in a row r of $\mathcal{T}$ then every chain following χ in r is either from $S_N \cap E_T$ or is a right justified from $S_N \cap E_N$.

PROOF: (a) Let $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ be the start of χ, and let $(i,j-1) \xrightarrow{\mathcal{T}} (p,q-1)$ be the end of the chain preceding χ in r. Since $\chi \in S_T$ then $(m,n-1) \in S$, and hence $(i,j-1)$ must be closer to $(p,q-1)$ than $(m,n-1)$ which implies that, if $(p,q) \in S$, (i,j) is closer to (p,q) than (m,n), contradicting $(i,j) \xrightarrow{\mathcal{T}} (m,n)$. So $(p,q) \notin S$, which implies that $q = Q$, and hence the chain preceding χ in r is from E_N.

If this chain is in $S_N \cap E_N$ then, by lemma 4, it is left justified and we are done. Otherwise it is from $S_T \cap E_N$, in which case we can repeat the argument until we get back to the first chain.

Part (b) is proved similarly. □

Notation: Let B be the set of all chains of the following types:

- chains in $S_T \cap E_T$,
- left justified chains in $S_N \cap E_T$ and
- right justified chains in $S_T \cap E_N$.

Lemma 6 Each row of the successor table contains at most one chain from B, and hence $\|B\| \leq P$.

PROOF: By lemma 5(a) if $\chi \in S_T$ is in a row r of $\mathcal{T}$ then every chain preceding χ in r is either from $S_T \cap E_N$, but not right justified and hence not from B, or from $S_N \cap E_N$, and hence not from B. Similarly if $\chi \in E_T$ is in a row r of $\mathcal{T}$ then no chain following χ in r is from B.

The result is immediate upon consideration of the 3 different types of chains in B. □

Notation: We call a consecutive sequence of links in a column of $\mathcal{T}$ with a common gap vector a **cross chain**,

$$\begin{array}{c} (i,j) \xrightarrow{\mathcal{T}} (m,n) \\ (i+1,j) \xrightarrow{\mathcal{T}} (m+1,n) \\ \vdots \\ (i+k,j) \xrightarrow{\mathcal{T}} (m+k,n). \end{array}$$

Most of the concepts which apply to chains also apply to cross chains.

A **mesh** is a sequence of chains in consecutive rows of $\mathcal{T}$ with the following properties:

(a) for any two consecutive chains in the mesh there must exist a cross chain which intersects both and

(b) the sequence of chains cannot be extended while maintaining property (a).

Note that every link in the mesh has the same gap vector.

The **reflection** of a mesh is the sequence of the reflections of chains in the mesh, clearly, the reflection of a mesh is also a mesh. We say a mesh is **symmetric** if it is its own reflection.

Lemma 7 Every chain in $S_N \cap E_N$ is in a symmetric mesh.

PROOF: Suppose we have a chain from $S_N \cap E_N$

$$(i,j) \xrightarrow{\mathcal{T}} (m,n), (i,j+1) \xrightarrow{\mathcal{T}} (m,n+1), \ldots (i,j+k) \xrightarrow{\mathcal{T}} (m,n+k),$$

either $j = 0$ or $n = 0$ and either $j+k = Q-1$ or $n+k = Q-1$. Suppose $j = 0$ then $j \leq n$, so $j+k \leq n+k$; then $n+k = Q-1$ and our chain is

$$(i,0) \xrightarrow{\mathcal{T}} (m,n), (i,1) \xrightarrow{\mathcal{T}} (m,n+1), \ldots, (i, Q-1-n) \xrightarrow{\mathcal{T}} (m, Q-1). \quad (2)$$

Every link in this chain has a gap vector $(m-i, n)$.

Consider any link $(a,b) \xrightarrow{\mathcal{T}} ((a,b)+(m-i,n))$, clearly $0 \leq b \leq Q-1-n$, and hence there is a link $(i,b) \xrightarrow{\mathcal{T}} (m, b+n)$ from (2) which is in the same column as $(a,b) \xrightarrow{\mathcal{T}} (a+m-i, b+n)$. We wish to show that $(a,b) \xrightarrow{\mathcal{T}} (a+m-i, b+n)$ and $(i,b) \xrightarrow{\mathcal{T}} (m, b+n)$ are in the same cross chain. Suppose not, and assume, without loss of generality, that $a < i$. Clearly the following are all pairs of points in S

$$(\delta, b), (\delta+m-i, b+n), \quad a \leq \delta \leq i.$$

Since $(a,b) \xrightarrow{\mathcal{T}} (a+m-i, b+n)$ and $(i,b) \xrightarrow{\mathcal{T}} (m, b+n)$ are not in the same cross chain then for some $\gamma \in [a+1, \ldots, i-1]$ there must exist a point (p,q) which lies between (γ, b) and $(\gamma+m-i, b+n)$. Now unless

$$p - (\gamma - a) < 0 \quad (3)$$

we will have $(p-\gamma+a, q) \in S$, so $(p-\gamma+a, q)$ would lie between (a,b) and $(a+m-i, b+n)$, contradicting $(a,b) \xrightarrow{\mathcal{T}} (a+m-i, b+n)$. Also unless

$$p + (i-\gamma) > P-1 \quad (4)$$

we will have $(p+i-\gamma, q) \in S$, then $(p+i-\gamma, q)$ would lie between (i, b) and $(m, b+n)$, contradicting $(i, b) \xrightarrow{\mathcal{T}} (m, b+n)$. Therefore we can assume that (3) and (4) hold. Combining (3) and (4) yields

$$p + (i - \gamma) > P - 1 + p - (\gamma - a),$$

this simplifies to

$$i - a > P - 1,$$

which is not possible. Hence $(a, b) \xrightarrow{\mathcal{T}} (a + m - i, b + n)$ and $(i, b) \xrightarrow{\mathcal{T}} (m, b + n)$ are in the same cross chain.

This proves that any link whose gap vector is $(m - i, n)$ is in the same cross chain as some link from (2). So there is only 1 mesh which contains links whose gap vector is $(m - i, n)$, and hence this mesh is symmetric.

If $j \neq 0$ then $n = 0$ and the proof follows similarly. □

Notation: Let A be the set of meshes containing only chains from $S_N \cap E_N$; by lemma 7 these meshes are symmetric.

Lemma 8 The number of gap sizes induced by S is at most

$$\|T\| \leq \|A\| + \|B\|.$$

PROOF: The chains are partitioned into 4 sets

$$S_N \cap E_N, S_N \cap E_T, S_T \cap E_N \text{ and } S_T \cap E_T.$$

We will show that any chain which is neither a chain in B nor a chain from a mesh in A has the same gap size as some chain in B. This is vacuously satisfied for chains in $S_T \cap E_T$ since $S_T \cap E_T \subseteq B$.

If χ is a chain from $S_N \cap E_T$ which is not in B then χ is not left justified, so let $(i, j) \xrightarrow{\mathcal{T}} (m, 0)$ be the start of χ. The end of $\mathcal{R}(\chi)$ is clearly the reflection of the start of χ, which is

$$(P - 1 - m, Q - 1) \xrightarrow{\mathcal{T}} (P - 1 - i, Q - 1 - j),$$

hence $\mathcal{R}(\chi)$ is right justified. Furthermore, by lemma 2, $\mathcal{R}(\chi) \in S_T \cap E_N$, thus $\mathcal{R}(\chi) \in B$.

Similarly we can show that the reflection of a chain from $(S_T \cap E_N) \backslash B$ is a left justified chain from $S_N \cap E_T$ (and is hence a member of B).

Now consider any chain $\chi \in S_N \cap E_N$ which is not from a mesh in A. Then, by the definition of A, χ belongs to a mesh which contains some chain not in $S_N \cap E_N$ (ie. a chain from $S_N \cap E_T$, $S_T \cap E_N$ or $S_T \cap E_T$); hence χ is the same width as some chain in B. □

Notation: Let V be the set of all vectors, that is,

$$V = \{(\delta, \gamma) : -P < \delta < P, -Q < \gamma < Q\}.$$

Then define the following subsets

$$\begin{aligned} V^{++} &= \{(\delta,\gamma) : 0 \leq \delta < P, 0 \leq \gamma < Q\}, \\ V^{+-} &= \{(\delta,\gamma) : 0 \leq \delta < P, -Q < \gamma \leq 0\}, \\ V^{-+} &= \{(\delta,\gamma) : -P < \delta \leq 0, 0 \leq \gamma < Q\} \text{ and} \\ V^{--} &= \{(\delta,\gamma) : -P < \delta \leq 0, -Q < \gamma \leq 0\}. \end{aligned}$$

The union of these sets is V. Now let $u^{++} \in V^{++}$ be a vector defined so that

$$\left|u^{++}\right| = \min_{v \in V^{++} \setminus \{(0,0)\}} |v|,$$

we define u^{+-}, u^{-+} and u^{--} similarly.

Lemma 9 For all $v \in V$ there exist nonnegative integers c_1, c_2, c_3, c_4 such that

$$v = c_1 u^{++} + c_2 u^{-+} + c_3 u^{+-} + c_4 u^{--},$$

and

$$|v| = c_1 \left|u^{++}\right| + c_2 \left|u^{-+}\right| + c_3 \left|u^{+-}\right| + c_4 \left|u^{--}\right|.$$

PROOF: By induction on $|v|$.

The initial case where $v = (0,0)$ is trivial. Assume that the statement holds for all u such that $|u| < |v|$. Now assume, without loss of generality, that $v \in V^{++}$.

Now, clearly, $(v - u^{++}) \in V$ and $|(v - u^{++})| < |v|$ so, by the induction hypothesis, there exist nonnegative integers a_1, a_2, a_3, a_4 such that

$$v - u^{++} = a_1 u^{++} + a_2 u^{-+} + a_3 u^{+-} + a_4 u^{--}.$$

Hence

$$v = (a_1 + 1) u^{++} + a_2 u^{-+} + a_3 u^{+-} + a_4 u^{--}.$$

Also, by the induction hypothesis, we have

$$\left|v - u^{++}\right| = a_1 \left|u^{++}\right| + a_2 \left|u^{-+}\right| + a_3 \left|u^{+-}\right| + a_4 \left|u^{--}\right|,$$

and, since $|v| > |v - u^{++}|$ then $|v - u^{++}| = |v| - |u^{++}|$, so

$$|v| = (a_1 + 1) \left|u^{++}\right| + a_2 \left|u^{-+}\right| + a_3 \left|u^{+-}\right| + a_4 \left|u^{--}\right|,$$

as required. □

Notation: For each $v \in V$ we define the subset

$$S(v) = \{(i,j) : (i,j) \in S, (i,j) + v \in S\},$$

that is $S(v)$ is the set of points (i,j) for which there exists another point $(i,j)+v$ at a distance $|v|$ in the clockwise direction around the perimeter.

Note for all $v \in V^{++}$ that $(0,0) \in S(v)$, and $\mathcal{T}(0,0)-(0,0) \in V^{++}$ so by the definition of u^{++} we get $\mathcal{T}(0,0) = (0,0) + u^{++}$. Similarly it can be shown that

$$\begin{aligned} \mathcal{T}(P-1,0) &= (P-1,0) + u^{-+}, \\ \mathcal{T}(0,Q-1) &= (0,Q-1) + u^{+-} \text{ and} \\ \mathcal{T}(P-1,Q-1) &= (P-1,Q-1) + u^{--}. \end{aligned}$$

Lemma 10

(a) $S(u^{++}) \cap S(u^{--}) = \emptyset$, and

(b) $S(u^{-+}) \cap S(u^{+-}) = \emptyset$.

PROOF: We assume, without loss of generality, that $|u^{++}| > |u^{--}|$. Now suppose that $(i,j) \in S(u^{++}) \cap S(u^{--})$. Hence $(i,j)+u^{--}$ and $(i,j)+u^{++}$ are both points, and the distance from $(i,j)+u^{--}$ to $(i,j)+u^{++}$ is

$$\begin{aligned} |(i,j) + u^{++} - ((i,j) + u^{--})| &= |u^{++} - u^{--}|, \\ &= |u^{++}| - |u^{--}| \\ &< |u^{++}|. \end{aligned}$$

Now $u^{++} - u^{--}$ is the vector difference between the points, so $u^{++} - u^{--} \in V$; further, by considering the signs of the components, we have $u^{++} - u^{--} \in V^{++}$. This is impossible since u^{++} was defined as having minimum length over all vectors in $V^{++} \setminus \{(0,0)\}$. This contradiction proves part (a) of the lemma. Part (b) is proved similarly. □

We now show that the set of gap sizes of S is unchanged if we replace α with $1-\alpha$. Consider the set

$$\begin{aligned} \{\{i(1-\alpha) + j\beta\} : \ & 0 \le i < P, 0 \le j < Q\} \\ &= \{\{(P-1-i')(1-\alpha) + j\beta\} : 0 \le i' < P, 0 \le j < Q\} \\ &= \{\{i'\alpha + j\beta - (P-1)\alpha\} : 0 \le i' < P, 0 \le j < Q\}, \end{aligned}$$

which is the set of points S rotated anticlockwise by a distance $(P-1)\alpha$ on the perimeter. Clearly this shifting will not affect gap sizes. However, if we set $\alpha' = 1-\alpha$, we find that the link

$$(i,j) \xrightarrow{\mathcal{T}} (m,n),$$

whose gap vector is $(m-i, n-j)$, changes under this transformation to

$$(P-1-i, j) \xrightarrow{\mathcal{T}} (P-1-m, n),$$

whose gap vector is $(-(m-i), n-j)$. This will mean that the vector lengths in V^{++} will interchange with those from V^{-+} and the vector lengths from V^{+-} will interchange with those from V^{--} when we use α' in place of α. This allows us to assume, without loss of generality, that

$$\left|u^{++}\right| + \left|u^{--}\right| \leq \left|u^{-+}\right| + \left|u^{+-}\right|, \tag{5}$$

since if this were not the case then we could replace α with $1-\alpha$, forming a new set S, with the same set of gap sizes, which satisfies (5).

By a similar analysis we can show that by replacing β with $1-\beta$ the vector lengths in V^{++} and V^{-+} are interchanged with those in V^{+-} and V^{--} respectively.

Now set $v_0 = u^{++} + u^{--}$.

Lemma 11

(a) $(P-1, Q-1) - u^{++} \in S(v_0) \cap S(u^{++})$, and

(b) $(0,0) - u^{--} \in S(v_0) \cap S(u^{--})$.

PROOF: (a) Clearly $(P-1, Q-1) - u^{++} \in S$. Further

$$\left((P-1, Q-1) - u^{++}\right) + v_0 = (P-1, Q-1) + u^{--},$$

which belongs to S, so $(P-1, Q-1) - u^{++} \in S(v_0)$. Also

$$\left((P-1, Q-1) - u^{++}\right) + u^{++} = (P-1, Q-1)$$

so $(P-1, Q-1) - u^{++} \in S(u^{++})$. Thus $(P-1, Q-1) - u^{++} \in S(u^{++}) \cap S(v_0)$ as required.

Part (b) is proved similarly. □

We will often think of a set of points $U \subseteq S$ as indices which mark positions in the successor table $\mathcal{T}$, we then define the **row span** of U, denoted $R_s(U)$, as the set of rows of $\mathcal{T}$ which contain 1 or more positions marked by U.

Lemma 12

$$\left|R_s\left(S(v_0) \cap S(u^{++})\right)\right| = \left|R_s\left(S(v_0) \cap S(u^{--})\right)\right|.$$

PROOF: We will prove the stronger result that

$$(P-1-i, Q-1-j) - u^{++} \in S(v_0) \cap S(u^{++})$$

if and only if

$$(i,j) - u^{--} \in S(v_0) \cap S(u^{--}).$$

If $(P-1-i, Q-1-j) - u^{++} \in S(v_0) \cap S(u^{++})$ then, by the definition of $S(v_0)$ and $S(u^{++})$, the following points are members of S

$$(P-1-i, Q-1-j) - u^{++}, \tag{6}$$
$$((P-1-i, Q-1-j) - u^{++}) + v_0 \quad \text{and} \tag{7}$$
$$((P-1-i, Q-1-j) - u^{++}) + u^{++}. \tag{8}$$

The reflections of (6), (7) and (8) are, respectively,

$$((i,j) - u^{--}) + v_0, \tag{9}$$
$$(i,j) - u^{--} \quad \text{and} \tag{10}$$
$$((i,j) - u^{--}) + u^{--}, \tag{11}$$

which are also members of S. The existence of points (10) and (11) imply that $(i,j) - u^{--} \in S(u^{--})$; similarly (10) and (9) imply $(i,j) - u^{--} \in S(v_0)$; which in turn implies $(i,j) - u^{--} \in S(v_0) \cap S(u^{--})$, as required.

The reverse implication is proved similarly. □

We say a nonzero vector v **clashes** with another vector u if $|v| < |u|$ and $S(u) \cap S(v) \neq \emptyset$, that is, there is some point (i,j) in $S(u)$ which is closer to the point $(i,j)+v$ than it is to $(i,j)+u$.

Lemma 13 There are no vectors which clash with u^{++} or u^{--}.

PROOF: Suppose v clashes with u^{++} then, for some point (i,j) we have $(i,j)+v$ and $(i,j)+u^{++}$ belonging to S. The vector distance from $(i,j)+v$ to $(i,j)+u^{++}$ is $u^{++} - v$. By lemma 9, there exist nonnegative integers a_1, a_2, a_3, a_4 and b_1, b_2, b_3, b_4 such that

$$v = a_1 u^{++} + a_2 u^{-+} + a_3 u^{+-} + a_4 u^{--} \tag{12}$$
$$|v| = a_1 \left|u^{++}\right| + a_2 \left|u^{-+}\right| + a_3 \left|u^{+-}\right| + a_4 \left|u^{--}\right| \tag{13}$$

and

$$u^{++} - v = b_1 u^{++} + b_2 u^{-+} + b_3 u^{+-} + b_4 u^{--} \tag{14}$$
$$\left|u^{++} - v\right| = b_1 \left|u^{++}\right| + b_2 \left|u^{-+}\right| + b_3 \left|u^{+-}\right| + b_4 \left|u^{--}\right|. \tag{15}$$

Note that $|u^{++} - v| = |u^{++}| - |v|$, so adding (13) and (15) yields

$$\left|u^{++}\right| = (a_1 + b_1)\left|u^{++}\right| + (a_2 + b_2)\left|u^{-+}\right| + (a_3 + b_3)\left|u^{+-}\right| + (a_4 + b_4)\left|u^{--}\right|. \quad (16)$$

Clearly we cannot have $a_1+b_1 > 1$. Suppose $a_1+b_1 = 1$. Then $a_2+b_2 = 0$, $a_3+b_3 = 0$ and $a_4 + b_4 = 0$ so that $v = u^{++}$ or $v = (0,0)$ in neither case does v clash with u^{++}. Therefore $a_1 + b_1 = 0$.

We cannot have both $a_3 + b_3 \geq 1$ and $a_2 + b_2 \geq 1$ otherwise, by the assumption (5), the right hand side of (16) would exceed the left hand side. We will consider the case $a_2 + b_2 = 0$, the case $a_3 + b_3 = 0$ is proved similarly.

Adding equations (12) and (14), with $a_1 + b_1 = 0$ and $a_2 + b_2 = 0$, gives

$$u^{++} = (a_3 + b_3)u^{+-} + (a_4 + b_4)u^{--}. \quad (17)$$

If we also have $a_3 + b_3 = 0$ it is easily seen that $u^{++} = (0,0)$, contradicting its definition. If $a_4 + b_4 = 0$ then, by equation (17), we arive at

$$u^{++} = (a_3 + b_3)u^{+-}$$

which, by the definitions of u^{++} and u^{+-}, implies that $a_3+b_3 = 1$. Therefore we have $a_1 = 0$, $a_2 = 0$, $a_3 = 0$ or 1 and $a_4 = 0$, hence $v = (0,0)$ or $v = u^{++}$, and again v does not clash with u^{++}. The final case is $a_3 + b_3 \geq 1$ and $a_4 + b_4 \geq 1$. The second component of the vector on the right hand side of (17) is nonpositive, while u^{++} has a nonnegative second component; then the second component of both vectors must be 0. This implies that u^{--} and u^{+-} must both have 0 as their second component. However, this means that no vector with a negative second component is expressible in the form given in lemma 9. This contradiction implies that no vectors clash with u^{++}.

We can prove that no vectors clash with u^{--} in a similar way. □

Lemma 14 If a vector v clashes with v_0 then $v \in \{u^{++}, u^{--}, u^{-+}, u^{+-}\}$.

PROOF: For any point $(i,j) \in S(v) \cap S(v_0)$, the vector distance from $(i,j) + v$ to $(i,j) + v_0$ is $v_0 - v$. Then, by lemma 9, there exist nonnegative integers a_1, a_2, a_3, a_4 and b_1, b_2, b_3, b_4 such that

$$v = a_1 u^{++} + a_2 u^{-+} + a_3 u^{+-} + a_4 u^{--} \quad (18)$$

$$|v| = a_1\left|u^{++}\right| + a_2\left|u^{-+}\right| + a_3\left|u^{+-}\right| + a_4\left|u^{--}\right| \quad (19)$$

and

$$v_0 - v = b_1 u^{++} + b_2 u^{-+} + b_3 u^{+-} + b_4 u^{--} \quad (20)$$

$$|v_0 - v| = b_1\left|u^{++}\right| + b_2\left|u^{-+}\right| + b_3\left|u^{+-}\right| + b_4\left|u^{--}\right|. \quad (21)$$

Clearly

$$\begin{aligned}|v_0 - v| + |v| &= |v_0| \\ &= |u^{++} + u^{--}| \\ &= |u^{++}| + |u^{--}|,\end{aligned}$$

so adding (19) to (21) we get

$$\left|u^{++}\right| + \left|u^{--}\right| = (a_1+b_1)\left|u^{++}\right| + (a_2+b_2)\left|u^{-+}\right| + (a_3+b_3)\left|u^{+-}\right| + (a_4+b_4)\left|u^{--}\right|. \quad (22)$$

Note that we must have

$$a_1 + a_2 + a_3 + a_4 \geq 1$$

and

$$b_1 + b_2 + b_3 + b_4 \geq 1,$$

so if

$$(a_1 + b_1) + (a_2 + b_2) + (a_3 + b_3) + (a_4 + b_4) = 2$$

then

$$a_1 + a_2 + a_3 + a_4 = 1$$

and hence, by (18), $v \in \{u^{++}, u^{-+}, u^{+-}, u^{--}\}$. This is what we wanted to prove so instead assume

$$(a_1 + b_1) + (a_2 + b_2) + (a_3 + b_3) + (a_4 + b_4) > 2. \quad (23)$$

We cannot have both $a_1 + b_1 \geq 1$ and $a_4 + b_4 \geq 1$ otherwise, by (23), the right hand side of (22) will become larger than the left hand side. Similarly, by assumption (5), we cannot have both $a_3 + b_3 \geq 1$ and $a_2 + b_2 \geq 1$. This gives rise to 4 cases, we consider only the case $a_4 + b_4 = 0$ and $a_2 + b_2 = 0$, the other cases are proved similarly.

Consider adding (18) and (20), when $a_4 + b_4 = 0$ and $a_2 + b_2 = 0$, we get

$$u^{++} + u^{--} = (a_1 + b_1)u^{++} + (a_3 + b_3)u^{+-}. \quad (24)$$

Now if $a_1 + b_1 \geq 1$ then (24) becomes

$$u^{--} = (a_1 + b_1 - 1)u^{++} + (a_3 + b_3)u^{+-}. \quad (25)$$

We cannot have either $a_1 + b_1 - 1 = 0$ or $a_3 + b_3 = 0$ as this would imply, respectively, that $a_3 + b_3 = 1$ or $a_1 + b_1 - 1 = 1$, yielding a contradiction to (23). Now the first component of the vector on the right hand side of (25) is nonnegative while u^{--} has a nonpositive first component, thus both first components must be 0, this implies that u^{+-} and u^{++} both have 0 as their first component. Then no vector with a positive first component is expressible in the form given in lemma 9, which is impossible. Hence we must have $a_1 + b_1 = 0$, so equation (24) simplifies to

$$u^{++} + u^{--} = (a_3 + b_3)u^{+-}, \quad (26)$$

and (22) becomes

$$\left|u^{++} + u^{--}\right| = (a_3 + b_3)\left|u^{+-}\right|, \tag{27}$$

also (23) implies that $a_3 + b_3 > 2$.

Note that $(0,0) - u^{--}$ and $(0,0) + u^{++}$ are both points in S, $(0,0) - u^{--}$ is positioned immediately anticlockwise of $(0,0)$ and $(0,0) + u^{++}$ is immediately clockwise of $(0,0)$. Using (26), we can have,

$$\begin{aligned}(0,0) + u^{++} &= ((0,0) - u^{--}) + u^{++} + u^{--} \\ &= ((0,0) - u^{--}) + (a_3 + b_3)u^{+-}.\end{aligned}$$

As $(0,0) - u^{--}$ and $((0,0) - u^{--}) + (a_3 + b_3)u^{+-}$ are both points in S it is not hard to see that $((0,0) - u^{--}) + u^{+-}$ and $((0,0) - u^{--}) + 2u^{+-}$ are also points in S. However, by (27), both of these points would lie on the arc between $(0,0) - u^{--}$ and $(0,0) + u^{++}$ which should only contain the point $(0,0)$. This contradiction completes the proof.

□

Lemma 15 One of the following cases must hold

(1) $v_0 = u^{+-}$,	(5) $u^{++} = u^{+-}$,
(2) $v_0 = u^{-+}$,	(6) $u^{--} = u^{-+}$,
(3) $v_0 = u^{-+} + u^{+-}$,	(7) $u^{--} = u^{+-}$ *or*
(4) neither u^{+-} nor u^{-+} clashes with v_0,	(8) $u^{++} = u^{-+}$.

PROOF: Suppose (4) does not hold, and suppose that u^{+-} clashes with v_0. Thus there exists a point (i,j) such that $(i,j) + u^{+-}$ and $(i,j) + v_0$ belong to S. The vector distance between these 2 points is $v_0 - u^{+-}$. Then, by lemma 9, there exist nonnegative integers a_1, a_2, a_3, a_4 such that

$$v_0 - u^{+-} = a_1 u^{++} + a_2 u^{-+} + a_3 u^{+-} + a_4 u^{--} \tag{28}$$

and

$$\left|v_0 - u^{+-}\right| = a_1\left|u^{++}\right| + a_2\left|u^{-+}\right| + a_3\left|u^{+-}\right| + a_4\left|u^{--}\right|. \tag{29}$$

We can rewrite (28) as

$$u^{++} + u^{--} = a_1 u^{++} + a_2 u^{-+} + (a_3 + 1)u^{+-} + a_4 u^{--}, \tag{30}$$

also, since $|v_0 - u^{+-}| = |u^{++}| + |u^{--}| - |u^{+-}|$, we can rewrite (29) as

$$\left|u^{++}\right| + \left|u^{--}\right| = a_1\left|u^{++}\right| + a_2\left|u^{-+}\right| + (a_3 + 1)\left|u^{+-}\right| + a_4\left|u^{--}\right|. \tag{31}$$

Suppose $a_2 \geq 1$. Then, by assumption (5), we must have $a_1 = 0$, $a_2 = 1$, $a_3 = 0$ and $a_4 = 0$. Substituting these values into (30) gives case (3). Then suppose $a_2 = 0$.

If $a_1 \geq 1$ and $a_4 \geq 1$ then the right hand side of (31) exceeds the left hand side, so we must have $a_1 = 0$ or $a_4 = 0$. Suppose $a_4 = 0$. Equation (30) then gives,

$$u^{++} + u^{--} = a_1 u^{++} + (a_3 + 1)u^{+-}. \tag{32}$$

If $a_1 = 1$ we get $u^{--} = u^{+-}$, which is case (7).

If $a_1 > 1$ equation (32) becomes

$$u^{--} = (a_1 - 1)u^{++} + (a_3 + 1)u^{+-}, \tag{33}$$

however the first component of the vector on the right hand side of (33) is nonnegative while u^{--} has a nonpositive first component, thus both first components must be 0, this implies that u^{+-} and u^{++} both have 0 as their first component. Then no vector with a positive first component is expressible in the form given in lemma 9. The remaining possibility is $a_1 = 0$.

With $a_1 = 0$, $a_4 = 0$ and $a_2 = 0$ equations (30) and (31) simplify to

$$u^{++} + u^{--} = (a_3 + 1)u^{+-} \tag{34}$$

and

$$\left|u^{++}\right| + \left|u^{--}\right| = (a_3 + 1)\left|u^{+-}\right|. \tag{35}$$

If $a_3 = 0$ we have case (1). Assume instead that $a_3 \geq 1$.

Note that $(0,0) - u^{--}$ and $(0,0) + u^{++}$ are, respectively, the points immediately preceding and following $(0,0)$ on the perimeter. Using (34), $(0,0) + u^{++}$ can be reexpressed as

$$\begin{aligned} (0,0) + u^{++} &= ((0,0) - u^{--}) + u^{++} + u^{--} \\ &= ((0,0) - u^{--}) + (a_3 + 1)u^{+-}. \end{aligned}$$

So $(0,0) - u^{--}$ and $((0,0) - u^{--}) + (a_3 + 1)u^{+-}$ are both points in S, as are the following:

$$((0,0) - u^{--}) + u^{+-}, ((0,0) - u^{--}) + 2u^{+-}, \ldots, ((0,0) - u^{--}) + a_3 u^{+-}.$$

Furthermore, by equation (35), these points lie on the arc between $(0,0) - u^{--}$ and $(0,0) + u^{++}$. The only point which lies on this arc is $(0,0)$, hence $a_3 = 1$ and

$$(0,0) - u^{--} + u^{+-} = (0,0),$$

ie. $u^{--} = u^{+-}$. Therefore (34) simplifies to

$$u^{++} = u^{--},$$

which implies that $u^{--} = (0,0)$ and $u^{++} = (0,0)$ which is impossible.

So far we have considered the possibilities u^{+-} clashes with v_0 (at the start of the proof) and $a_4 = 0$ (before equation (32)), and shown that these possibilities lead to cases (1),(3) or (7). The alternative possibilities lead to the other cases. □

Note that parts (1) and (2) are special cases of part (4) in lemma 15, however these cases will be treated separately in later proofs.

Lemma 16

(a) If $(i,j) \xrightarrow{\mathcal{T}} (i,j)+v$ is the start of a chain χ then $\chi \in S_N$ if and only if $(i,j-1) \notin S(v)$ (or equivalently $\chi \in S_T$ if and only if $(i,j-1) \in S(v)$).

(b) If $(i,j) \xrightarrow{\mathcal{T}} (i,j)+v$ is the end of a chain χ then $\chi \in E_N$ if and only if $(i,j+1) \notin S(v)$ (or equivalently $\chi \in E_T$ if and only if $(i,j+1) \in S(v)$).

PROOF: (a) Let $(m,n) = (i,j)+v$.

First, if $\chi \in S_N$ then $j = 0$ or $n = 0$, which implies that $(i,j-1) \notin S$ or $(m,n-1) \notin S$, and hence $(i,j-1) \notin S(v)$.

Now, if $(i,j-1) \notin S(v)$ then $(i,j-1) \notin S$ or $(m,n-1) \notin S$. However $(i,j) \xrightarrow{\mathcal{T}} (m,n)$ implies $(i,j) \in S$ and $(m,n) \in S$, so it must be the case that $j = 0$ or $n = 0$, which implies that $\chi \in S_N$.

Part (b) is proved similarly. □

We are now nearly ready to prove the $P+3$ theorem, to do this we use the structure in the successor table, for each of the cases described in lemma 15, to bound the right hand side of the inequality

$$\|T\| \leq \|A\| + \|B\|,$$

obtained in lemma 8.

In each of these 8 cases we will define a set $\varphi \subseteq V$. Its definition will change from case to case but it will always have the following properties.

1. For all $v \in \varphi$ and $u \in V \backslash \varphi$, u does not clash with v.
2. For all $i = 0, 1, \ldots, P-1$ there exists $j \in [0, \ldots, Q)$ and $v \in \varphi$ such that $\mathcal{T}(i,j) = (i,j)+v$.

Define

$$\Phi = \bigcup_{v \in \varphi} S(v),$$

then the above properties imply:

1'. $(i,j) \in \Phi$ if and only if there exists $v \in \varphi$ such that $\mathcal{T}(i,j) = (i,j) + v$.

2'. $R_s(\Phi)$ contains all rows of $\mathcal{T}$.

Define $\Psi = S \backslash \Phi$, then let $R_2 = R_s(\Psi)$, and let R_1 be those rows of $\mathcal{T}$ not in R_2 (thus $|R_1| + |R_2| = P$).

Let A_1 and A_2 be the sets of meshes from A which are in the regions Φ and Ψ respectively. Then let a_i $(i = 1, 2)$ be the number of meshes in A_i consisting of a single chain.

Now let B_i $(i = 1, 2)$ be the set of chains from B in the rows of R_i, and let f be the number of rows in R_2 which contain no chains from B. Then, by lemma 6, $|B_2| = |R_2| - f$, and hence

$$|B_2| = P - |R_1| - f. \tag{36}$$

Finally let t be the number of rows from R_2 whose justified chains are both from $S_N \cap E_N$ and in the region Φ.

Lemma 17 If $\varphi \subseteq V$ has properties 1 and 2 then (with a_1, R_1 and t defined, as above, with respect to φ)

$$\|T\| \leq P + 2 + |\varphi| - |R_1| - \left\lceil \frac{a_1 + t}{2} \right\rceil.$$

PROOF: Let $\varphi \subseteq V$ satisfy properties 1 and 2 (and define A_1, A_2, B_1, B_2, R_1, R_2, a_1, a_2, Φ, Ψ, f and t, as above, with respect to φ).

Note that $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$ so lemma 8 implies

$$\|T\| \leq \|A_1\| + \|B_1\| + \|A_2\| + \|B_2\|. \tag{37}$$

Now A_1 and B_1 are both in the region Φ, and, by property 1', there are $|\varphi|$ gap vectors for this region. Hence (37) becomes

$$\|T\| \leq |\varphi| + \|A_2\| + \|B_2\|. \tag{38}$$

We now use properties 1 and 2 of φ to prove the following facts.

1. Every row from R_2 contains a chain from $S_N \cap E_N$ in the region Φ.

2. Every chain from B_2 is in the region Ψ.

3. Every chain from Ψ is reflected into Ψ.

Consider any row r from R_2. By definition, r contains some link from the region Ψ, and, by property $2'$, r contains some link from the region Φ; hence r is partitioned into intervals which are alternately contained within Φ and Ψ. Let r' be one of these intervals in Φ. By lemma 16 and property $1'$ the first and last chains in r' are from S_N and E_N respectively. Then if the last chain in r' is not in $S_N \cap E_N$ it is from $S_T \cap E_N$, so, by lemma 5(a), the first chain in r' must be from $S_N \cap E_N$. Hence either the first or last chain from r' is from $S_N \cap E_N$ which implies fact 1.

To prove fact 2 we assume, without loss of generality, that the first chain in r' is in $S_N \cap E_N$. If r' contains only one chain it is in $S_N \cap E_N$, and hence not in B_2. So assume that r' contains at least 2 chains. Under these assumptions r' is, by lemma 4, left justified, also by lemma 4 the last chain cannot be from $S_N \cap E_N$, otherwise r' would be right justified implying $r = r'$ contradicting $r \in R_2$. Therefore the last chain in r' is from $S_T \cap E_N$, but is not right justified and hence not in B, all other chains in r' are, by lemma 5(a), either from $S_T \cap E_N$, which are not right justified and hence not chains from B, or from $S_N \cap E_N$, and hence not in B. Hence no chain from B_2 is in the region Φ, which implies fact 2.

Since a link and its reflection have a common gap vector then, by property 1 of φ, the reflection of a link, chain or mesh from Φ is also in the region Φ. Which, by contrapositive, implies fact 3.

There are a_2 meshes in A_2 which consist of a single chain from $S_N \cap E_N$, the other $|A_2| - a_2$ meshes from A_2 each contain at least 2 chains from $S_N \cap E_N$. Therefore there are at least

$$2(|A_2| - a_2) + a_2 = 2|A_2| - a_2$$

chains from $S_N \cap E_N$ in the region Ψ, these chains are, by lemma 4, justified chains. These chains come from rows in R_2, and, by fact 1, every such row also contains a justified chain from $S_N \cap E_N$ which is in the region Φ. Hence there are at least $2|A_2| - a_2 + t$ rows from R_2 whose justified chains are both from $S_N \cap E_N$, at most f of these rows contain no chain from B_2, the other $2|A_2| - a_2 + t - f$ or more rows must, by the definition of B, each contain a chain from $S_T \cap E_T$.

There are $a_1 + a_2$ meshes in A consisting of single chains, these chains are, by the definition of A and lemma 7, symmetric chains from $S_N \cap E_N$. Then, by lemma 3, the number of symmetric chains from $S_T \cap E_T$ is at most $4 - (a_1 + a_2)$.

Therefore there are at least

$$(2|A_2| - a_2 + t - f) - (4 - (a_1 + a_2)) = 2|A_2| - 4 + a_1 + t - f$$

chains from $S_T \cap E_T$ in B_2 which are not symmetric. Such chains are, by fact 2, in the region Ψ and hence, by fact 3, their reflections (which, by lemma 2, are also chains in $S_T \cap E_T$) are also in the region Ψ. Therefore the $2|A_2| - 4 + a_1 + t - f$ or more non-symmetric chains from B_2 are paired off, under reflection, into pairs of chains with the same gap vector, so by (36),

$$\begin{aligned} \|B_2\| &\leq |B_2| - \left\lceil \tfrac{2|A_2|-4+a_1+t-f}{2} \right\rceil \\ &= P - |R_1| - f - |A_2| + 2 - \left\lceil \tfrac{a_1+t-f}{2} \right\rceil \\ &= P + 2 - |R_1| - |A_2| - \left\lceil \tfrac{a_1+t+f}{2} \right\rceil \\ &\leq P + 2 - |R_1| - |A_2| - \left\lceil \tfrac{a_1+t}{2} \right\rceil . \end{aligned}$$

Which, with (38), gives

$$\begin{aligned} \|T\| &\leq P + 2 - |R_1| + |\varphi| - \left\lceil \tfrac{a_1+t}{2} \right\rceil + (\|A_2\| - |A_2|) \\ &\leq P + 2 - |R_1| + |\varphi| - \left\lceil \tfrac{a_1+t}{2} \right\rceil , \end{aligned}$$

as required. □

P+3 Theorem

Proof: Consider the 8 cases of lemma 15. It is sufficient to prove that the theorem holds for instances of cases 1,3,4,5 and 7, since, by replacing α with $1-\alpha$ and/or β with $1-\beta$, we can transform instances of cases 2, 6 and 8 into instances of cases 1, 5 and 7 respectively. For example if we have a set S, which is an instance of case 2, we can construct a new set S' by replacing α with $1-\alpha$ and replacing β with $1-\beta$ which has the same set of gap sizes. However the lengths of vectors in V^{++} and V^{+-} will be interchanged, by the transformation, with those from V^{--} and V^{-+} respectively. Hence set S' is an instance of case 1, and the assumption (5) will remain satisfied by S'. So if the theorem holds for instances of case 1 it also holds for instances of case 2.

Cases 1, 5 and 7: $v_0 = u^{+-}$, $u^{++} = u^{+-}$ or $u^{+-} = u^{--}$.

Set $\varphi = \{u^{++}, u^{--}, v_0\}$, then property 1 is implied by lemmas 10, 13 and 14, and property 2 is implied by lemma 11. Hence, by lemma 17, we have

$$\begin{aligned} \|T\| &\leq P + 2 - |R_1| + |\varphi| - \left\lceil \frac{a_1+t}{2} \right\rceil \\ &\leq P + 5 - |R_1| - \left\lceil \frac{a_1}{2} \right\rceil . \end{aligned} \tag{39}$$

The successor table for these cases is represented by figure 2.

If $|R_1| > 1$ then, by (39), we get $\|T\| \leq P + 3$. Then suppose $|R_1| = 1$. Clearly $R_s(S(u^{++})) \subseteq R_1$, and hence $|R_s(S(u^{++}))| = 1$ which implies that $S(u^{++})$ contains

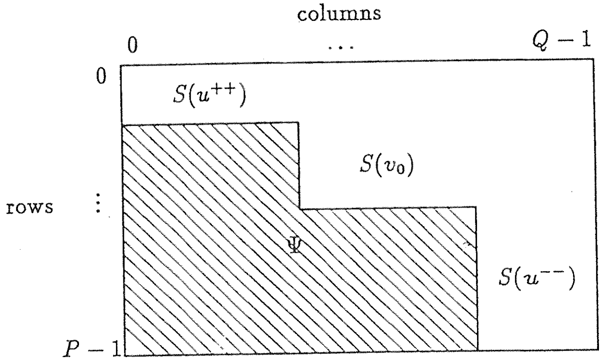

Figure 2: Successor table for cases (1), (5) and (7)

only 1 chain. Now, by lemmas 13 and 16, we see that $S(u^{++})$ is a mesh in A_1, and hence $a_1 \geq 1$. Therefore, by (39), $\|T\| \leq P+3$.

Case 3: $u^{++} + u^{--} = u^{-+} + u^{+-}$.

Note that replacing α with $1-\alpha$ and/or β with $1-\beta$ converts an instance of case 3 into another instance of case 3, while not affecting assumption (5). Using such transformations we can assume, without loss of generality, that $v_0 \in V^{++}$. Also, by applying such transformations to lemma 13, no vectors clash with u^{+-} or u^{-+}.

Define $\varphi = \{u^{++}, u^{+-}, u^{-+}, u^{--}, v_0\}$. Then lemmas 13 and 14 imply that φ has property 1, and lemma 11 implies that φ has property 2. Hence, by lemma 17, we get

$$\|T\| \leq P + 7 - |R_1| - \left\lceil \frac{a_1 + i}{2} \right\rceil \tag{40}$$

A block diagram of the successor table for this case is represented in figure 3 (where $\Psi = \Psi_1 \cup \Psi_2$).

When $|R_1| \geq 4$ (40) implies $\|T\| \leq P+3$, so we assume $|R_1| \leq 3$. Note that $R_s(S(u^{+-})) \subseteq R_1$ and $R_s(S(u^{--}) \cap S(v_0)) \subseteq R_1$; futhermore $R_s(S(u^{+-}))$ and $R_s(S(u^{--}) \cap S(v_0))$ are disjoint, so

$$|R_1| \geq \left|R_s\left(S(u^{+-})\right)\right| + \left|R_s\left(S(u^{--}) \cap S(v_0)\right)\right|.$$

Now, by lemma 12, $|R_s(S(u^{--}) \cap S(v_0))| = |R_s(S(u^{++}) \cap S(v_0))|$, and, since $v_0 \in V^{++}$, $|R_s(S(u^{++}) \cap S(v_0))| = |R_s(S(u^{++}))|$, which gives

$$|R_1| \geq \left|R_s\left(S(u^{++})\right)\right| + \left|R_s\left(S(u^{+-})\right)\right|. \tag{41}$$

Hence, by lemma 11, $|R_1| \geq 2$. Therefore there are 2 nontrivial cases to consider.

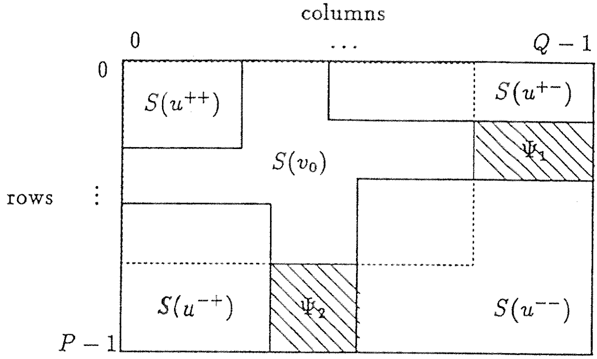

Figure 3: Successor table for case (3)

Case 3(a) $|R_1| = 3$.

By (41) either $|R_s(S(u^{++}))| = 1$ or $|R_s(S(u^{+-}))| = 1$, hence either $S(u^{++})$ or $S(u^{+-})$ contain only one chain. Now, by lemmas 13 and 16, $S(u^{++})$ and $S(u^{+-})$ are both meshes in A_1, so $a_1 \geq 1$, then, by (40),

$$\|T\| \leq P + 4 - \left\lceil \frac{1}{2} \right\rceil = P + 3.$$

Case 3(b) $|R_2| = 2$.

By (41) both $|R_s(S(u^{++}))| = 1$ and $|R_s(S(u^{+-}))| = 1$, then by the argument used in case 3(a) we see that $S(u^{++})$ and $S(u^{+-})$ are both meshes in A_1 which consist of a single chain, hence $a_1 \geq 2$. We assume, without loss of generality, that $t = 0$, since, by (40), if this were not the case then $\|T\| \leq P + 3$, as required. Note that any row in $R_s(\Psi_2)$ has both justified chains from $S_N \cap E_N$ and in the region Φ, so $t \geq |R_s(\Psi_2)|$, and hence $\Psi_2 = \emptyset$. Therefore $R_s(S(u^{--})) \subseteq R_1$, then, since $R_s(S(u^{--})) \cap R_s(S(u^{+-})) = \emptyset$, $|R_s(S(u^{--}))| = 1$. This implies that $S(u^{--})$ contains only 1 chain. By lemmas 13 and 16 $S(u^{--})$ is a mesh in A_1, as are $S(u^{++})$ and $S(u^{+-})$, so $a_1 \geq 3$. Thus, by (40), $\|T\| \leq P + 3$.

Case 4: Neither u^{+-} nor u^{-+} clash with v_0.

Set $\varphi = \{u^{++}, u^{--}, v_0\}$ (as in case 1), so that Φ has properties 1 and 2. Therefore, by lemma 17, we have

$$\begin{aligned} \|T\| &\leq P + 2 + |\varphi| - |R_1| - \left\lceil \frac{a_1 + t}{2} \right\rceil \\ &\leq P + 5 - |R_1| - \left\lceil \frac{a_1}{2} \right\rceil . \end{aligned} \tag{42}$$

We note that when $v_0 \in V^{+-}$ or $v_0 \in V^{-+}$ then we have an instance of case 1 and case 2 respectively. Furthermore, if $v_0 \in V^{--}$ then we can create a new problem by using $1-\alpha$ and $1-\beta$ in place of α and β respectively, this transformation will yield another instance of case 4 with the same set of gap sizes, but $v_0 \in V^{++}$. Therefore we can assume that $v_0 \in V^{++}$.

A block diagram of the successor table for this case is given by figure 4 (where $\Psi = \Psi_1 \cup \Psi_2$).

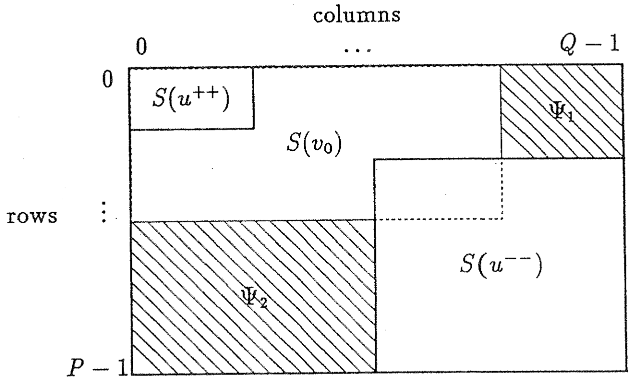

Figure 4: Successor table for case (4)

If $|R_1| \geq 2$ then, by (42), $\|T\| \leq P+3$. Then suppose $|R_1| \leq 1$. It is clear that $R_s(S(u^{--}) \cap S(v_0)) \subseteq R_1$, and hence $|R_s(S(u^{--}) \cap S(v_0))| = 1$. Therefore, by lemma 12, $|R_s(S(u^{++}) \cap S(v_0))| = 1$, and, since $v_0 \in V^{++}$, $|R_s(S(u^{++}))| = 1$. Thus $S(u^{++})$ contains only 1 chain. Now, by lemmas 13 and 16, $S(u^{++})$ is a mesh in A_1, so $a_1 \geq 1$. Therefore, by (42), $\|T\| \leq P+3$.

Hence, in general, $\|T\| \leq P+3$. □

3. Achievability

Clearly in the case $P=1$ (ie. the 3 gap problem) the bound $P+3$ is not achievable, however, the bound is achievable for $P>1$. In the following generalised example we show that the bound is achievable when $P \geq 3$. In this we use $Q = P+1$, it is not known if the $P+3$ is achievable when $P = Q$. The referee pointed out that this example does not work when $P=2$. He adds that "it can be shown that 5 gaps are not achievable when $P=2$, $Q=3$ (to have 5 gaps in this case, there would have to be 4 symmetric chains consisting each of a single link, but according to the proof of Lemma 3, there can only be 2 symmetric chains length of given parity). It is easy to construct an example with $P=2$, $Q=4$ having 5 gaps."

Example For $P \geq 3$ set $Q = P+1$, $\beta = 1$ and $\alpha = 1+\epsilon$, where $0 < P\epsilon < 1$.

We use a circle of circumference

$$(P-1)(1+\epsilon)+P-1-\delta,$$

where $\delta < \epsilon/2$. Naturally this can be normalised to a circle of unit circumference, however it is more convenient to use the given circumference.

Firstly we will investigate the lengths of the subintervals induced by partitioning the interval

$$[0,(P-1)(1+\epsilon)+P]$$

with the set

$$S' = \{i\alpha + j\beta : 0 \leq i < P, 0 \leq j < Q\}.$$

Then consider the gap sizes achieved by wrapping this interval onto the perimeter of a circle of circumference

$$(P-1)(1+\epsilon)+P-1-\delta.$$

S' simplifies as follows

$$\begin{aligned} S' = & \{i(1+\epsilon)+j : 0 \leq i < P,\ 0 \leq j < P+1\} \\ = & \{(i+j)+i\epsilon : 0 \leq i < P,\ 0 \leq j < P+1\} \\ = & \{k+i\epsilon : 0 \leq k < 2P,\ \max\{0,k-P\} \leq i \leq \min\{P-1,k\}\} \\ = & \{k+i\epsilon : 0 \leq k < P,\ 0 \leq i \leq k\} \\ & \cup \{k+i\epsilon : P \leq k < 2P,\ k-P \leq i \leq P-1\}. \end{aligned}$$

Define

$$S_1' = \{k+i\epsilon : 0 \leq k < P, 0 \leq i \leq k\}$$

and

$$S_2' = \{k+i\epsilon : P \leq k < 2P, k-P \leq i \leq P-1\},$$

since each element of S_2' is greater than or equal to P then the set of subintervals induced by partitioning $[0,P]$ with the set S_1' is a subset of the set of subintervals obtained by partitioning the interval

$$[0,(P-1)(1+\epsilon)+P]$$

with the set S'. The lengths of the subintervals in the partitioning of $[0,P]$ with S_1' are ϵ and

$$\{(\ell+1)-(\ell+\ell\epsilon) : 0 \leq \ell < P\} = \{1-\ell\epsilon : 0 \leq \ell < P\}.$$

Hence S' partitions the interval

$$[0,(P-1)(1+\epsilon)+P]$$

into subintervals of at least $P+1$ distinct lengths.

We now consider wrapping the interval

$$[0,(P-1)(1+\epsilon)+P]$$

around a circle of circumference

$$(P-1)(1+\epsilon)+P-1-\delta,$$

in doing so the ends of the interval will be overlapped by a length of $1+\delta$. There are only 2 elements in the set S_2' which are in this overlapped section, these are

$$2P-2+(P-1)\epsilon \text{ and } 2P-1+(P-1)\epsilon,$$

due to the modulo equivalence of distances around the perimeter these points in the overlapped section are equivalent to

$$\delta \text{ and } 1+\delta.$$

δ lies in the subinterval $[0,1]$, creating 2 new subintervals of length δ and $1-\delta$, but in doing so we lose the subinterval of length 1. $1+\delta$ lies in the subinterval $[1,1+\epsilon]$, creating 2 new subintervals of length $\epsilon-\delta$ and δ, and although the subinterval of length ϵ is destroyed there are other subintervals of this length, eg. $[2,2+\epsilon]$. So the following are all subinterval lengths (given in ascending order)

$$\delta, \epsilon-\delta, \epsilon, 1-(P-1)\epsilon, 1-(P-2)\epsilon, \ldots, 1-2\epsilon, 1-\epsilon \text{ and } 1-\delta.$$

Hence there are $P+3$ gap sizes.

4. Unsolved Problems

Notation: In this section we assume that all sets contain real numbers in the interval $[0,1)$, and that every set contains 0. We think of the elements of the set as points on the perimeter of a circle of unit circumference. The **gaps** of a set S are the arcs between adjacent points of S on the perimeter, and we define $\|S\|$ as the number of distinct gap sizes of S.

Given sets S and T we define the new set

$$S \vee T = \{\{s+t\} : s \in S, t \in T\},$$

$S \vee T$ is the wedge of S and T. Geometrically the set $S \vee T$ is obtained by extending the sequence T from each point in S on the perimeter.

The 3 gap theorem can be stated as

$$\|\{\{i\alpha\} : 0 \le i < N\}\| \le 3,$$

and the $P+3$ theorem implies that

$$\|\{\{i\alpha\} : 0 \le i < P\} \vee \{\{i\beta\} : 0 \le i < Q\}\| \le P+3.$$

An obvious generalisation of these problems is to find an upper bound for

$$\|\{\{i\alpha_1\} : 0 \le i < N_1\} \vee \{\{i\alpha_2\} : 0 \le i < N_2\} \vee \ldots \vee \{\{i\alpha_K\} : 0 \le i < N_K\}\| . \quad (43)$$

It has been shown that (43) is no greater than

$$\frac{3}{2} \prod_{i=1}^{K-1} N_i + 3,$$

but is conjectured to be

$$\prod_{i=1}^{K-1} N_i + c_K,$$

where c_K is independent of $N_1, \ldots, N_K$.

The following problem, posed by Erdös, is to find values of α and β for which the limit

$$\limsup_{N \to \infty} \|\{\{i\alpha\} : 0 \le i < N\} \vee \{\{i\beta\} : 0 \le i < N\}\| \quad (44)$$

is finite. Ron Holzman has shown that if α, β and 1 are linearly dependent over the rationals then (44) is finite, it is conjectured that this is a necessary condition. However there are no known instances for which (44) has been proven to be infinity.

For the final problem let

$$S = \{0, \alpha_1, \alpha_2, \ldots, \alpha_k\},$$

where α_1, $\alpha_2, \ldots,$ α_k and 1 are linearly independent over the rationals, then define

$$S^n = \overbrace{S \vee S \vee \ldots \vee S}^{n \text{ sets}}.$$

For some values of n and k (e.g. when $n = 1$) we can find values α_1, $\alpha_2, \ldots,$ α_k for which each of the $|S^n|$ arcs on the perimeter has a unique gap size, and hence $\|S^n\| = |S^n|$. However for other values of n and k $\|S^n\| < |S^n|$ regardless of the choice of α_1, $\alpha_2, \ldots,$ α_k; for example when $k = 1$ and $n > 2$ we have an instance of the 3 gap theorem for which this is true. We can show that when $n > 2k$

$$\|S^n\| < |S^n|,$$

for all possible values $\alpha_1, \alpha_2, \ldots, \alpha_k$. However we are unsure if $\|S^n\| = |S^n|$ when $n \le 2k$.

Acknowledgements

The authors thank Paul Erdös, Aviezri Fraenkel, Ron Holzman Jane Pitman, and Alison Wolfe for useful discussions. We particularly thank the referee for his very careful reading of this tedious document.

References

[1] F. R. K. Chung and R. L. Graham, *On the set of distances determined by the union of arithmetic progressions,* Ars Combinatoria 1 (1976) 57–76.

[2] V. Chvátal, D. A. Klarner and D. E. Knuth, *Selected combinatorial research problems.* Stanford CS Report 292 (1972).

[3] A. S. Fraenkel, *Complementing and exactly covering sequences.* J. Combin. Theory 14 (1973), 8–20.

[4] A. S. Fraenkel and R. Holzman, *Gap problems for integer part and fractional part sequences.* To be published.

[5] J. Geelen and R. J. Simpson, *Gap Theorems.* Technical Report 3/91, Curtin University of Technology, Perth, Australia (1991).

[6] J. Geelen, *3d or not 3d, and other questions generalizing the three gap theorem.* Dissertation, Curtin University of Technology, Perth, Australia (1991).

[7] D. E. Knuth, *The Art of Computer Programming.* Vol. 3, Ex. 6.4.10, (Addison-Wesley, 1968).

[8] F. M. Liang, *A short proof of the 3d distance theorem.* Discrete Mathematics 28 (1979), 325–326.

[9] F. Mignosi, *On the number of factors of Sturmian words.* Theoretical Computer Science 82 (1991), 71–84.

[10] T. van Ravenstein, *The three gap theorem (Steinhaus Conjecture),* J. Austral. Math. Soc. (A), 45 (1988), 360–370.

[11] N. B. Slater, *Gaps and steps for the sequence $n\theta$ mod 1.* Proc. Cambridge Phil. Soc. 63 (1967), 1115–1123.

[12] V. T. Sós, *On the theory of diophantine approximations I.* Acta Math. Acad. Sci. Hungar. 8 (1957), 461-472.

[13] V. T. Sós, *On the distribution Mod 1 of the sequence $n\alpha$.* Ann. University of science, Budapest, Eötvös Sect. Math. 1(1958), 127–134.

[14] K. B. Stolarsky, *Beatty sequences, continued fractions, and certain shift operators.* Can. Math. Bull. 19 (1976), 473–482.

[15] S. Świerczkowski, *On successive settings of an arc on the circumference of a circle.* Fund. Math. 46 (1958), 187–189.

(Received 14/8/92; revised 3/2/93)