Ergod. Th. & Dynam. Sys. (1985), **5**, 59-69
Printed in Great Britain

On generic points in the Cartesian square of Chacón's transformation

A. DEL JUNCO AND M. KEANE

Mathematics Department, University of Toronto, Toronto M5S 1A1, Canada; Technische Hogeschool Delft, Afdeling Wiskunde, 2628 BL Delft, The Netherlands

(*Received* 13 *January* 1984)

Abstract. We give an example of a compact metric space X and a strictly ergodic homeomorphism T of X with invariant probability μ such that for every $x \in X$ the set $\{y \in X: (x, y)$ is not generic for $\mu \times \mu\}$ is countable.

1. *Introduction*

The purpose of this note is to establish the following result.

THEOREM 1. *There exists a compact metric space X and a strictly ergodic homeomorphism of X with invariant probability μ with the property that for every $x \in X$ the set $\{y \in X: (x, y)$ is not generic for $\mu \times \mu\}$ is countable.*

Here generic means right and left generic, i.e. for both $T \times T$ and $T^{-1} \times T^{-1}$.

The example we shall provide is a topological version of the classical weakly-mixing map of Chacón which is described below and also in [**4**] and [**5**]. We shall see that in this example there is a countable set of exceptional points, consisting of a pair of orbits, which may be described as those points which are asymptotic to some other point. We then have the following more precise statement.

THEOREM 1′. *For Chacón's example if x and y are on different orbits and at least one is non-exceptional then (x, y) is generic for product measure.*

Theorem 1 answers a question posed to us by S. Glasner. He had the following application in mind: if (X, T) satisfies the condition of theorem 1 and in addition the measure theoretic system $\mathscr{X} = (X, \mu, T)$ has minimal self-joinings (see [**10**]), then any other measure theoretic system $\mathscr{Y}$ is either disjoint from $\mathscr{X}$ or is an extension of a symmetric Cartesian power of $\mathscr{X}^{n\odot}$ of $\mathscr{X}$, that is the usual Cartesian power $\mathscr{X}^n$ restricted to the sigma-algebra of sets invariant under co-ordinate permutations. It was later discovered [**6**] that this 'universal disjointness' result holds for any $\mathscr{X}$ with MSJ. Glasner [**2**] has given another proof of this fact.

Glasner [**3**] has given another application of theorem 1. A topological system (Q, T) is called *affine* if Q is a compact convex subset of a locally convex space and T is an affine homeomorphism of Q. A map $\phi: (X, T) \to (Q, T)$ is called an *affine embedding* if (Q, T) is affine, ϕ is continuous, 1-1 and equivariant and co $(\phi(X)) = Q$. (X, T) is said to be *absolutely extremal* if for every $x \in X$ and every affine embedding $\phi : X \to Q$, $\phi(x)$ is an extreme point of Q. In [3] it is shown that any strictly ergodic (X, T) which is POD (see [1]) and satisfies theorem 1 is absolutely extremal, thereby providing weakly-mixing examples of absolutely extremal systems. Chacón's example is in fact strictly ergodic and POD. The POD property is observed in [4] and it also follows from theorem 1′ once one has a precise description of the exceptional points. [5] contains a simple direct proof.

Some further comments are in order. Theorem 1′ implies 2-fold minimal self-joinings and may be viewed as a strong version thereof. A similar strong version of k-fold minimal self-joinings follows easily from theorem 1′ and k-fold minimal self-joinings.

THEOREM 2. *For Chacón's example if at most one of* $x_1, x_2, \ldots, x_k$ *is exceptional and no two are on the same orbit then* $(x_1, x_2, \ldots, x_k)$ *is generic for* μ^k.

Theorem 1′ implies that if (x, y) are on different orbits and not both are exceptional then the past and future limit sets of (x, y) are both $X \times X$. This may be viewed as a definition of topological minimal self-joinings, modulo exceptional points (not to be confused with the weaker definition studied in [8]). This definition can be extended in a natural way to k-fold topological minimal self-joinings and turns out to be satisfied by Chacón's example and to have consequences quite analogous to (measure-theoretic) minimal self-joinings. This will be the subject of a future paper by the first author.

In both the topological and measure-theoretic settings one can ask for examples without exceptional points. For example, is there a uniquely ergodic topological system (X, T) with the property that, for any x, y on different orbits, (x, y) is generic for $\mu \times \mu$? has dense orbit? For discrete time symbolic systems such as Chacón's example one has an obstacle in the existence of forward asymptotic points so we certainly cannot ask for randomness (topological or measure-theoretic) in both past and future. Passing, however, to flows, it can be seen, using results of Ratner [9], that in any horocycle flow (X, T_t) defined by a discrete co-compact, maximal, non-arithmetic subgroup of SL $(2, \mathbb{R})$ when x and y are on different orbits the past and future limit sets of (x, y) are $X \times X$. One may then ask whether (x, y) is generic for product measure and also for k-fold statements. For discrete time systems we do not know whether it is possible to have past and future limit sets of (x, y) equal to $X \times X$ whenever x and y are on different orbits.

Most of the present work was done while A. del Junco was visiting the Technische Hogeschool Delft in the Netherlands. A del Junco was supported in part by NSF Grant MCS-8003038.

2. *Chacón's example*

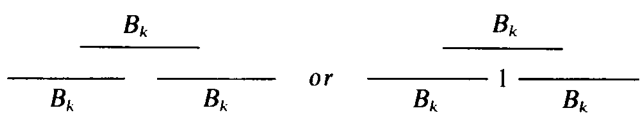
This will be defined as a subshift of $\{0, 1\}^{\mathbb{Z}}$ using a simple block structure. Define finite blocks $B_1, B_2, \ldots$ inductively by setting

$$B_0 = 0, \quad B_1 = 0\ 0\ 1\ 0, \quad B_{k+1} = B_k\ B_k\ 1\ B_k.$$

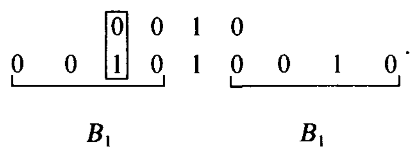
We refer to B_k as a *k-block* and denote its length by h_k. We let $X \subset \{0, 1\}^{\mathbb{Z}}$ consist of all those sequences x such that each finite segment of x is a segment of B_k for some k. X is evidently a closed shift-invariant subset of $\{0, 1\}^{\mathbb{Z}}$. T denotes the shift homeomorphism on X.

By a *spaced concatenation* of k-blocks we mean a concatenation of k-blocks with a single 1 interposed between some of the k-blocks. The following lemma tells us that in a spaced concatenation of k-blocks we see k-blocks only where we put them.

LEMMA 1. *In a spaced concatenation of k-blocks, k-blocks occur only at the natural positions, that is, one never sees*



Proof. We use induction starting with $k = 1$. If B_1 occurs in a spaced concatenation x of 1-blocks its single 1 must be either the 1 of the 'genuine' 1-blocks in x or else a 1-block spacer. In the first case we are done and the second is evidently impossible:



Now suppose B_{k+1} occurs in a spaced concatenation x of $k+1$-blocks. x is also in a natural way a spaced concatenation of k-blocks. By induction the component k-blocks of B_{k+1} must appear in natural positions in x and this can only happen if B_{k+1} itself occurs naturally. □

There is a natural way to produce sequences in x by a nesting procedure as follows. Choose a sequence $\xi \in \{1, 2, 3\}^{\mathbb{N}}$ which we may call the nesting instruction. Now start with a B_0, consider it as the $\xi(1)$th B_0 in a B_1, which in turn is considered as the $\xi(2)$th B_1 in a B_2 and so on. In this way the B_k's expand to define an infinite sequence ξ^*, which is well defined only up to a shift. ξ^* will be a doubly infinite sequence unless $\xi(i) = 3$ eventually in which case ξ^* is a left infinite sequence which we denote $B_{-\infty}$, or $\xi(i) = 1$ eventually which yields a right infinite sequence denoted B_∞. We can now give a precise description of X.

LEMMA 2. *X consists (up to shifts) of all the doubly infinite sequences ξ^*, $\xi \in \{1, 2, 3\}^{\mathbb{N}}$ and the sequences $B_{-\infty}B_\infty$* and *$B_{-\infty}1B_\infty$.*

Proof. In a doubly infinite ξ^* any finite segment is covered by a B_k, so $\xi^* \in X$. In $B_{-\infty}B_\infty$ any finite segment is covered by B_kB_k which appears in B_{k+1}. In $B_{-\infty}1B_\infty$ any finite segment is covered by B_k1B_k which also occurs in B_{k+1}.

Now suppose $x \in X$. Observe that for each k we can find in x a spaced concatenation of k-blocks covering any given finite segment η of x. To see this expand η on either side by h_k to get a larger segment η^*. Now η^* appears in some B_K and hence in all B_K for K sufficiently large. But for $K > k$, B_K is a spaced concatenation of

k-blocks. Hence in η^* we see a spaced concatenation of k-blocks which must cover η.

We have thus what might be called local k-block structure in x and lemma 1 tells us that this structure is unique. Now suppose we have any k-block η in x. Then there is a spaced concatenation of $k+1$-blocks in x covering η and by uniqueness η must be a component k-block of one of its $k+1$-blocks. In this way we obtain a nested sequence of k-blocks in x so x is either a ξ^* or x contains a B_∞ or a $B_{-\infty}$. If x contains a B_∞, this B_∞ is preceded either by a 0 or a 01 since no $x \in X$ contains two 1's in a row. In the first case this 0 is contained in a nested sequence of k-blocks, none of which can overlap the B_∞ by uniqueness, so $x = B_{-\infty}B_\infty$. In the second case we find similarly that $x = B_{-\infty}1B_\infty$. In the same way if x contains a $B_{-\infty}$, $x = B_{-\infty}B_\infty$ or $x = B_{-\infty}1B_\infty$. □

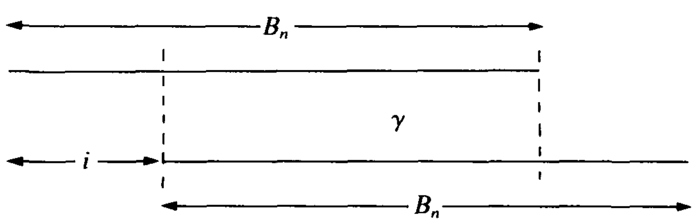
We call the countably many sequences of the form $B_{-\infty}B_\infty$ and $B_{-\infty}1B_\infty$ exceptional. For $x \in X$, $x \neq B_{-\infty}1B_\infty$ the 0th coordinate in x is contained for sufficiently large k in a unique k-block called the *time* 0 *k-block*. For $x = B_{-\infty}1B_\infty$ we will also speak of the time 0 k-block by first shifting if necessary, which will not affect any of our arguments.

Using lemma 2, each $x \in X$ is, for every k, an (infinite) spaced concatenation of k-blocks. It is then an easy exercise to establish the strict ergodicity of (X, T). All we shall need to know about the invariant measure μ is that it gives the same measure to each k-block level, a k-block level being a set consisting of all $x \in X$ whose 0th coordinate lies at some given position in a k-block.

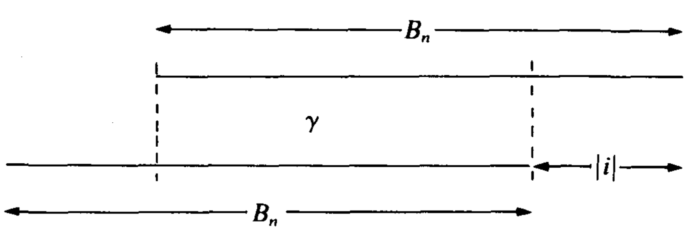
3. *Proof of theorem* 1′

We now fix once and for all an (x, y) satisfying the hypotheses of theorem 1′. We regard the pair (x, y) in $\{0, 1\}^{\mathbf{Z}} \times \{0, 1\}^{\mathbf{Z}}$ also as a single sequence in $(\{0, 1\}^2)^{\mathbf{Z}}$.

By an n-block overlap with shift i, $|i| < h_n$, we mean the finite $\gamma \in (\{0, 1\}^2)^{h_n - |i|}$ defined, if $i \geq 0$, by the picture



and, if $i < 0$, by



We call γ a right overlap when $i \geq 0$ and a left overlap when $i < 0$. By the complementary overlap of γ, denoted $\bar{\gamma}$, we mean the overlap with shift $i - h_n \operatorname{sgn}(i)$.

Given any finite segment γ of (x, y) the non-normalized distribution of right k-block overlaps, denoted $f(\gamma, k)$, is the measure on $[0, h_k - 1]$ which counts for each $i \in [0, h_k - 1]$ the number of k-block overlaps with shift i which occur in γ. If γ is the overlap of the time 0 n-blocks in x and y we denote by γ_n whichever of γ and $\bar{\gamma}$ is a right overlap and set

$$d(n, k) = \frac{f(\gamma_n, k)}{\|f(\gamma_n, k) + f(\bar{\gamma}_n, k)\|}, \qquad \bar{d}(n, k) = \frac{f(\bar{\gamma}_n, k)}{\|f(\gamma_n, k) + f(\bar{\gamma}_n, k)\|}.$$

We denote by $d_l(n, k)$ and $\bar{d}_l(n, k)$ the corresponding quantities for left k-block overlaps.

We denote the length of a finite sequence γ by $|\gamma|$. We note here that our assumptions on (x, y) imply that $|\gamma_n| \to \infty$ and $|\bar{\gamma}_n| \to \infty$. For if either were not the case we could then find a subsequence n_i and an l such that the time 0 n_i-block in x begins at coordinate $j(n_i)$ and some n_i-block in y begins at $j(n_i) + l$. When x is not exceptional its time 0 n-blocks grow to cover all of x so we can conclude that $x = T^l y$ contradicting our assumption on (x, y).

Let us say that a probability measure P on $\mathbb{Z}$ is ε, h-*locally flat* if it is within ε (distance measured in total variation norm) of an average of probabilities each of which is uniform on some interval I of length greater than h contained in the support of P. We denote the normalization of a measure ν by $\underline{\nu}$. The following lemma is an important reduction in our proof of theorem 1′.

LEMMA 3. *In order to prove theorem* 1′ *it suffices to establish*:

(*) *For all* ε, h, K *there exists* N *such that for* $n > N$ *there exists* $k > K$ *with* $h_n > \varepsilon^{-1} h_k$ *such that if* ν *is any of the measures* $d(n, k)$, $\bar{d}(n, k)$, $d_l(n, k)$ $\bar{d}_l(n, k)$ *then*:

(i) ν *is supported on* $J \subset [b, h_k - b]$ *with* $|J| < \varepsilon b$; *and*

(ii) *either* $\|\nu\| < \varepsilon$ *or* $\underline{\nu}$ *is* ε, h-*locally flat.*

Proof. Given a finite segment ξ of $(x; y)$, ξ induces an empirical joint distribution of l-block levels. To describe it, relabel the co-ordinates of (x, y) as follows: if a co-ordinate sits in an l-block label it $0, \ldots, h_l - 1$ according to its position in the l-block, otherwise label it h_l. Then as a finite sequence in $([0, h_l]^2)^{|\xi|}$, ξ induces an empirical distribution on $[0, h_l - 1]^2$, which we denote $\mu(\xi, l)$. To prove theorem 1′ our task is to show that given ε, l, for any sufficiently long segment ξ of (x, y) which contains time 0, $\mu(\xi, l)$ is within ε of uniform on $[0, h_l - 1]^2$. We first claim that in fact it suffices to establish:

(**) For all ε, l there exists N such that for $n > N$ $\mu(\gamma_n, l)$ is ε-close to uniform unless $|\gamma_n|/h_n < \varepsilon$, and the same for $\bar{\gamma}_n$.

To see the sufficiency of (**) let ξ be a large segment of (x, y) containing 0 and then choose h_n of order $\varepsilon|\xi|$ (this is possible since $h_{n+1} \sim 3h_n$) so that, except for a fraction about ε of its length, ξ is made up of a number j of the order of ε^{-1} right and left n-block overlaps $\{\alpha_i\}$ and $\{\beta_i\}$ respectively. Each α_i (resp. β_i) is a shift of γ_n (resp. $\bar{\gamma}_n$) by not more than j, since the shifts of adjacent right (resp. left) overlaps differ by at most 1. Since $|\gamma_n|\to\infty$ and $|\bar\gamma_n|\to\infty$, by making $|\xi|$ large enough we may assume that j is small compared to $|\gamma_n|$ and $|\bar\gamma_n|$, so that $|\alpha_i|$ and $|\beta_i|$ are essentially the same as $|\gamma_n|$ and $|\bar\gamma_n|$ respectively.

Now $\mu(\xi, l)$ is within ε of an average of $d(\alpha_i, l)$ and $d(\beta_i, l)$ and in this average the α_i have total weight about $|\gamma_n|/h_n$. If $|\gamma_n|/h_n<\varepsilon$ the α_i make only a negligible contribution. But assuming (**) if $|\gamma_n|/h_n>\varepsilon$, $\mu(\gamma_n, l)$ is ε-close to uniform, whence $\mu(\alpha_i, l)$ is also close to uniform since α_i is a shift of γ_n by at most j. Making the same remarks for β_i and $\bar\gamma_n$ we see that all the non-negligible $\mu(\alpha_i, l)$ and $\mu(\beta_i, l)$ are close to uniform so $\mu(\xi, l)$ is itself close to uniform. This establishes the sufficiency of (**).

To show that (*) implies (**) we proceed as follows. For notational convenience we replace ε in (**) by ε'. Given ε', l choose N so that (*) holds for an ε, h, K yet to be specified and then choose $n>N$. We show that (**) holds for γ_n, as the argument is the same for $\bar\gamma_n$. Suppose, then, that $|\gamma_n|/h_n\geq\varepsilon'$. Condition (i) of (*) ensures that each full k-block in the x part of γ_n gives rise to exactly one right and one left k-block overlap in γ_n. Thus,

$$\|d(n,k)\|>\frac{|\gamma_n|}{h_n}-\eta>\varepsilon'-\eta,$$

where the error η is of the order of $h_k/|\gamma_n|$ plus the frequency of k-block spacers. Since $|\gamma_n|>\varepsilon' h_n>\varepsilon'\varepsilon^{-1}h_k$, the first contribution to the error can be made small by making ε small while the second will be small if k is sufficiently large. Thus we can ensure that $\|d(n,k)\|>\varepsilon'/2>\varepsilon$ and the same for $d_l(n,k)$, so both $\underset{\sim}{d}(n,k)$ and $\underset{\sim}{d}_l(n,k)$ are ε, h-locally flat.

Condition (i) of (*) ensures that all the right k-block overlaps in γ_n have the same length up to an ε relative error and the same for the left overlaps. If ε is sufficiently small and k is sufficiently large, it follows that $\mu(\gamma_n, l)$ is as close as we like to an average of the measures

$$\sum_i \underset{\sim}{d}(n,k)(i)\mu(\alpha_i, l) \tag{1}$$

and

$$\sum_i \underset{\sim}{d}_l(n,k)(i)\mu(\beta_i, l) \tag{2}$$

where α_i and β_i denote respectively the right and left k-block overlaps with shift i.

So it suffices now to show that the measure (1) (the argument is the same for (2)) is close to uniform on $[0, h_l-1]^2$. Since $\underset{\sim}{d}(n,k)$ is ε, h-locally flat it is within ε of an average $A=\sum_j a_jp_j$ where each p_j is a uniform probability on an interval of length at least h. Thus (1) is within ε of

$$\sum_i A(i)\mu(\alpha_i, l)=\sum_j a_j\sum_i p_j(i)\mu(\alpha_i, l).$$

So finally it suffices to show that each $\nu_j=\sum_i p_j(i)\mu(\alpha_i, l)$ is close to uniform on $[0, h_l-1]^2$. Note first that both marginals of $\mu(\alpha_i, l)$ are close to uniform for (i) of (*) guarantees that $|\alpha_i|>b>\varepsilon^{-1}|J|>\varepsilon^{-1}$, so $|\alpha_i|$ can be made large compared to h_l

by making ε small. Thus also both marginals of ν_j are close to uniform. It remains to show that the conditionals of ν_j are close to uniform. For this we just have to require that h be large compared to h_l, so that each l-block position in y scans across a large number of l-blocks in x, thus seeing the different l-block positions with approximately equal probabilities. □

The remainder of our efforts are devoted to establishing condition (*). We will do this for $d(n, k)$ and $\bar{d}(n, k)$ as the argument is the same for left k-block overlaps. We will use the following lemma to obtain local flatness. Its proof is left as an exercise.

LEMMA 4. *Given* ε, h *there exists* $\eta > 0$ *such that if* P *is any probability distribution on* $\mathbb{Z}$ *which is an average of an increasing and a decreasing distribution with* $\max_i P(i) < \eta$ *then* P *is* ε, h *locally flat.*

We regard $d(n, k)$ and $\bar{d}(n, k)$ as measures on $\mathbb{Z}$, supported on $[0, h_k - 1]$ and denote their Fourier transforms by

$$\phi_n^k(z) = \sum_{i \in \mathbb{Z}} d(n, k)(i) z^i$$

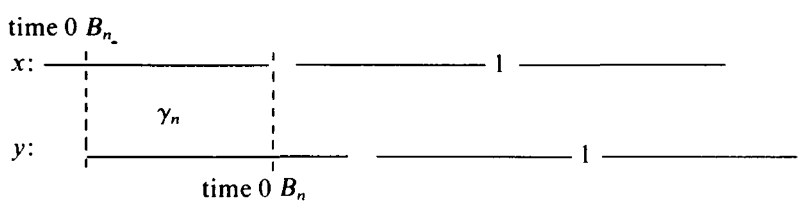
and

$$\bar{\phi}_n^k(z) = \sum_{i \in \mathbb{Z}} \bar{d}(n, k)(i) z^i.$$

We also set

$$\Phi_n^k(z) = \begin{pmatrix} \phi_n^k(z) \\ \bar{\phi}_n^k(z) \end{pmatrix}.$$

We now obtain a formula for Φ_{n+1}^k in term of Φ_n^k. Suppose, for example, that we see in x and y the following pattern of time 0 n-blocks and $n+1$-blocks.



Assuming that $f(\gamma_n, k)$ and $f(\gamma_n^*, k)$ have no mass at 0 or $h_k - 1$ and denoting by S the left shift on $\mathbb{Z}$ we see that

$$f(\gamma_{n+1}, k) = 3f(\gamma_n, k) + f(\bar{\gamma}_n, k) + Sf(\bar{\gamma}_n, k)$$
$$f(\bar{\gamma}_{n+1}, k) = f(\bar{\gamma}_n, k)$$

Thus

$$\begin{aligned}\|f(\gamma_{n+1}, k)\| &= \|3f(\gamma_n, k) + 2f(\bar{\gamma}_n, k) + Sf(\bar{\gamma}_n, k)\| \\ &= 3\|f(\gamma_n, k) + f(\bar{\gamma}_n, k)\|,\end{aligned}$$

so

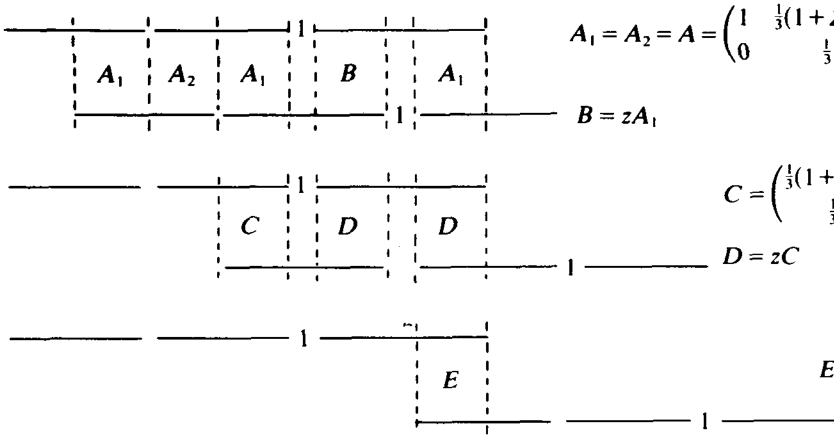
$$d(n+1, k) = d(n, k) + \tfrac{1}{3}(\bar{d}(n, k) + S\bar{d}(n, k))$$
$$\bar{d}(n+1, k) = \tfrac{1}{3}\bar{d}(n, k)$$

https://doi.org/10.1017/S0143385700002753 Published online by Cambridge University Press

and

$$\Phi_{n+1}^k(z)=\begin{pmatrix}1 & \frac{1}{3}(1+z^{-1})\\ 0 & \frac{1}{3}\end{pmatrix}\Phi_n^k(z).$$

We now summarize the matrices governing the transition from Φ_n^k to Φ_{n+1}^k corresponding to all possible pictures of time 0 n-blocks and $n+1$-blocks, assuming that $d(n,k)$ and $\bar{d}(n,k)$ have no mass at 0 or h_k-1. By abuse of notation the symbols A_1, A_2 etc. represent situations as well as the corresponding matrices even though the matrices corresponding to different situations may be equal. (We will need to distinguish situations A_1 and A_2 but not the two types of D situations.)



$$A_1=A_2=A=\begin{pmatrix}1 & \frac{1}{3}(1+Z^{-1})\\ 0 & \frac{1}{3}\end{pmatrix}$$

$$B=zA_1$$

$$C=\begin{pmatrix}\frac{1}{3}(1+z^{-1}) & \frac{1}{3}z^{-1}\\ \frac{1}{3} & \frac{1}{3}(1+z^{-1})\end{pmatrix}$$

$$D=zC$$

$$E=\begin{pmatrix}\frac{1}{3} & 0\\ \frac{1}{3}(1+z) & 1\end{pmatrix}$$

There are six more situations A_1', A_2', B', C', D' and E' corresponding to reversing the roles of x and y in A_1, A_2, B, C, D and E. This has the effect of replacing $M(z)$ by $fM(z^{-1})f$ where

$$f=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

Thus we have

$$A_1'=A_2'=E;$$
$$B'(z)=z^{-1}E(z);$$
$$C'=D,\quad D'=C;$$
$$E'=A.$$

We denote by M_n the matrix governing the transition from Φ_n^k to Φ_{n+1}^k. We will say that n is *k-admissible* if each $d(n',k)$, $k\le n'\le n$ has no mass at 0 or h_{k-1}, so that $\Phi_n^k=M_{n-1}\dots M_k\Phi_n^k$. For each fixed value of z we shall regard these matrices as operators on $\mathbb{C}^2$ and use their L_1-norm. Recall that

$$\left\|\begin{matrix}a & b\\ c & d\end{matrix}\right\|_1=\max\left(|a|+|c|,|b|+|d|\right).$$

Note that all matrices have norm bounded by 1 and that C and D have norm bounded away from 1 when z is bounded away from 1. Also AE and EA have norms similarily bounded away from 1. We'll refer to C, D, AE and EA as flattening products, since they decrease $\|\Phi_n^k(z)\|_1$.

In order to apply lemma 4 we will need some monotonicity of $d(n, k)$ which is a consequence of the following lemma. We will say a distribution on $\mathbb{Z}$ is *of type* $\wedge$(i) if it is symmetric about i, increasing on $(-\infty, i]$ and decreasing on $[i, \infty)$. We will say it is *of type* $\wedge(i+\frac{1}{2})$ if it is symmetric about $i+\frac{1}{2}$, increasing on $(-\infty, i]$ and decreasing on $[i+1, \infty)$.

LEMMA 5. *If n is k-admissible then*

$$\begin{pmatrix} d(n,k) \\ \bar{d}(n,k) \end{pmatrix}$$

always has one of the following symmetry configurations for some i:

$$\begin{pmatrix} \wedge(i) \\ \wedge(i+\frac{1}{2}) \end{pmatrix}, \quad \begin{pmatrix} \wedge(i-\frac{1}{2}) \\ \wedge(i) \end{pmatrix}.$$

Proof. This holds trivially when $n = k$ because $d(k, k)$ is a point mass and $\bar{d}(k, k) = 0$. We then simply check that it persists through application of any of the transition matrices. □

Remark. If we are working with $d_l(n, k)$ and $\bar{d}_l(n, k)$ then of course $\bar{d}_l(k, k)$ is a point measure while $d_l(k, k) = 0$.

In view of lemmas 4 and 5, in order to establish (*) it suffices to show that

(‡) given ε, K if n is sufficiently large there is a $k > K$ such that n is k-admissible, $h_n > \varepsilon^{-1} h_k$, $d(n, k)$ and $\bar{d}(n, k)$ are supported on $J \subset [b, h_k - b]$ with $|J| < \varepsilon b$, $\max d(n, k) < \varepsilon$ and $\max \bar{d}(n, k) < \varepsilon$.

Recall that there are only six possibilities for the matrices M_n: A, zA, C, D, E and $z^{-1}E$. We group the sequence $M_1, M_2, \ldots$ into maximal runs of one of the following types:

A's and zA's, (situations A_1, A_2, B, E');
E's and $z^{-1}E$'s, (situations A_1', A_2', B', E);
a single C;
a single D.

We now claim that:

(1) there are infinitely many runs;
(2) each pair of consecutive runs contains a flattening product;
(3) the product of each run contains entries which are polynomials in z and z^{-1} of degree 1.

(1) follows from our assumptions on (x, y) as follows. Suppose to the contrary that the sequence M_n eventually consisted only of A's and B's (the 'dual' case of E's, B''s is handled identically). Observe that within a run of situations A_1, A_2, B and

E' the only possible transitions are given by

$$\begin{tikzcd} & A_2 \arrow[dr] & \\ E' \arrow[loop left] \arrow[ur] \arrow[dr] & & A_1 \arrow[loop right] \\ & B \arrow[ur] & \end{tikzcd}$$

Thus we would in fact have that either E' holds eventually, which is impossible, since it means that $|\bar{\gamma}_n|$ is eventually constant, or A_1 holds eventually which would also mean that $|\bar{\gamma}_n|$ is eventually constant. (Recall that $|\bar{\gamma}_n| \to \infty$.)

(2) is obvious: if neither of the runs in a pair is a C or D then one must be a run of A's and zA's, the other of E's and zE's so we get a flattening product at the junction.

To see (3) observe that any power of A has entries which are of degree 1 in z^{-1} and that any run of A's and zA's can have at most one zA in it.

We have seen that for all $\varepsilon > 0$ there exists $C(\varepsilon) < 1$ such that if M is a flattening product and $|\arg z| > \varepsilon$ then $\|M\|_1 < C(\varepsilon)$. (We choose $-\pi < \arg z \leq \pi$.) To prove (‡), given ε, K choose M so that

$$C(\varepsilon/10)^M < \varepsilon/10, \qquad 3^M > \varepsilon^{-1},$$

and then choose $K' > K$ so that for $k \geq K'$

$$\min(|\gamma_k|, |\bar{\gamma}_k|) > 2M + 2.$$

Finally, take N so large that the interval $[K', N-1]$ contains at least $2M$ full runs.

Now fix $n > N$ and choose k as large as possible so that $[k, n-1]$ contains $2M$ full runs, so certainly $k \geq K$ and $n \geq k + M$ so $h_n \geq 3^M h_k \geq \varepsilon^{-1} h_k$. The product $M_{n-1} \cdots M_k$ is contained in $2M+2$ full runs so by (3) its entries have degree at most $2M+2$ in z, z^{-1}. Since $k \geq K'$ $d(k, k)$ and $\bar{d}(k, k)$ have no mass in the left or right $2M+2$ places of $[0, h_k - 1]$ so it follows that n is k-admissible. Moreover for $|\arg z| > \varepsilon/10$ we have

$$\|M_{n-1} \cdots M_k\|_1 < C(\varepsilon/10)^M < \varepsilon/10,$$

since the sequence $M_k, \ldots, M_{n-1}$ contains at least M consecutive pairs of full runs, each of which contains a flattening product. Thus for $|\arg z| > \varepsilon/10$

$$\begin{aligned} \|\Phi_n^k(z)\|_1 &= \|M_{n-1}(z) \cdots M_k(z) \Phi_k^k(z)\|_1 \\ &\leq \frac{\varepsilon}{10} \|\Phi_k^k(z)\| = \frac{\varepsilon}{10}. \end{aligned}$$

Finally, denoting normalized Lebesgue measure on $\{|z| = 1\}$ by dz we have for any i

$$d(n, k)(i) = \int_{|z|=1} |z^{-i} \phi_n^k(z)\, dz| \leq \int_{|z|=1} |\phi_n^k(z)|\, dz$$

$$\leq \int_{|z|=1} \|\Phi_n^k(z)\|_1 \, dz \leq \int_{|\arg z| > \varepsilon/10} (\varepsilon/10) \, dz + \int_{|\arg z| \leq \varepsilon/10} 1 \, dz$$

$$< \frac{\varepsilon}{10} + \frac{2\varepsilon}{10} \frac{1}{2\pi} < \varepsilon.$$

Similarly $\bar{d}(n, k)(i) < \varepsilon$. This completes the proof of (‡) and hence of theorem 1′. □

REFERENCES

[1] H. Furstenberg, H. Keynes & L. Shapiro. Prime flows in topological dynamics. *Israel J. Math.* **14** (1973), 26–38.
[2] S. Glasner. Quasifactors in ergodic theory. *Israel J. Math.* **45** (1983), 198–208.
[3] S. Glasner. Absolutely extremal points in minimal flows. Preprint.
[4] A. del Junco. A family of counterexamples in ergodic theory. *Israel J. Math.* **44** (1983) 160–188.
[5] A. del Junco. Prime systems in ergodic theory and topological dynamics. Lecture notes, Technische Hogeschool Delft.
[6] A. del Junco and D. Rudolph. To appear.
[7] A. del Junco, M. Rahe & L. Swanson. Chacón's automorphism has minimal self-joinings. *J. d'Analyse Math.* **37** (1980), 276–284.
[8] N. Markley. Topological minimal self-joinings. *Ergod. Th. & Dynam. Sys.* **3** (1983), 579–599.
[9] M. Ratner. Joinings of horocycle flows. Preprint.
[10] D. Rudolph. An example of a measure-preserving map with minimal self-joinings and applications. *J. d'Analyse Math.* **35** (1979), 97–122.