ISRAEL JOURNAL OF MATHEMATICS, Vol. 50, Nos. 1-2, 1985

CUTTING AND STACKING, INTERVAL EXCHANGES AND GEOMETRIC MODELS

BY

PIERRE ARNOUX,[a] DONALD S. ORNSTEIN[b,†] AND BENJAMIN WEISS[c,†]

[a] *U.E.R. Sciences de Reims, 51062 Reims, France*;

[b] *Stanford University, Stanford, CA 94305, USA*;

and [c] *The Hebrew University of Jerusalem, Jerusalem, Israel.*

ABSTRACT

Every aperiodic measure-preserving transformation can be obtained by a cutting and stacking construction. It follows that all such transformations are infinite interval exchanges. This in turn is used to represent any ergodic measure-preserving flow as a C^∞-flow on an open 2-manifold. Several additional applications of the basic theorems are also given.

The cutting and stacking construction as a basic tool in ergodic theory is well known (cf., for example, [2] Ch. 6). It is less widely known that every aperiodic measure-preserving transformation (a.m.p.t.) can be obtained in this way. By using this basic fact one can realize an arbitrary ergodic transformation as an infinite interval exchange with the same cluster point in the domain and in the range. As a first application of this fact we show in §2 how any a.m.p.t. can be realized as the Poincaré return map for a nice cross-section of a C^∞-flow on an open 2-manifold. Combining this with a smoothing technique found in [4] we are also able to show that any ergodic flow can be represented as a smooth flow on such a manifold. This of course does not advance us in the study of the basic question of finding smooth *compact* models but does highlight the fact that in the absence of compactness the problem is quite easy.

Another application of the basic fact is the representation of an arbitrary ergodic transformation as the speed up $T_{(x)}^{p(x)}$ of any preassigned ergodic T. We should point out that just as requiring $p(x)$ to be integrable places severe

† Partial support for this work was given by the National Science Foundation under grant number MCS81-07092.

Received April 24, 1983 and in revised form September 2, 1984

restrictions on the nature of T^p (see below in §3), so too requiring that the partition of the interval in an infinite interval exchange S have finite entropy forces restrictions on S, such as forcing S to have zero entropy.

§1. Cutting and stacking and infinite interval exchange

We begin by reviewing briefly the cutting and stacking construction in order to establish notation. A *stack* $\mathcal{S}(h, w)$ of height h and width w is a collection of h intervals of length w thought of as being placed one above the other, with a transformation implicitly defined on the first $h-1$ levels as translating each level to the level above. The union of the levels of a stack $\mathcal{S}$ is denoted by $|\mathcal{S}|$. If w_i are widths that add up to w, $\Sigma_{i=1}^{b} w_i = w$, then to *cut* a stack (h, w) into b stacks $\mathcal{S}_i(h, w_i)$ is to divide each level of $\mathcal{S}$ into b intervals of length $w_1, w_2, \ldots, w_b$ (reading from left to right) and then forming a single stack $\mathcal{S}_i$ from the ith interval of each level, so that $|\mathcal{S}| = \bigcup_{i=1}^{b} |\mathcal{S}_i|$.

An abstract construction by cutting and stacking begins with a finite collection of stacks, cuts each one into smaller stacks, possibly adds some number of new levels to these (not necessarily the same on each) and then forms new stacks of greater height by placing stacks of the same width one on top of the other. This is repeated infinitely often and if the total length of intervals is finite and if the widths tend to zero, then this always defines an ergodic transformation of a finite interval preserving Lebesgue measure. To describe this is a precise way we denote by $\mathcal{S}_1(h_1, w) \cdot \mathcal{S}_2(h_2, w)$ the stack of height $h_1 + h_2$ and width w formed by placing the levels of $\mathcal{S}_2$ directly above the highest level of $\mathcal{S}_1$. Note that $\mathcal{S}_1\mathcal{S}_2$ is only defined if the width of $\mathcal{S}_1$ equals the width of $\mathcal{S}_2$. The basic step in the construction proceeds formally in a sequence of three steps beginning from $\mathcal{S}_i(h_i, w_i)$, $1 \leqq i \leqq a$:

(i) for w_{ij}'s satisfying $\Sigma_{j=1}^{b_i} w_{ij} = w_i$ cut $\mathcal{S}_i$ into $\mathcal{S}_{ij}(h_i, w_{ij})$;

(ii) for non-negative integers f_{ij}, g_{ij} take new levels in the form of stacks $\mathcal{S}_{ij}^{*}(f_{ij}, w_{ij})$, $\mathcal{S}_{ij}^{**}(g'_{ij}, w_{ij})$ and form $\mathcal{S}_{ij}^{*}\mathcal{S}_{ij}\mathcal{S}_{ij}^{**}$;

(iii) partition the indices (i,j), $1 \leqq i \leqq a$, $1 \leqq j \leqq b_i$ into ordered sets I_k, $1 \leqq k \leqq \bar{a}$ that have the property that the w_{ij} for $(i,j) \in I_k$ have a common value $\bar{w}_k$ (these need not be maximal with this property, and the ordering is arbitrary) and form $\bar{\mathcal{S}}_k$ by concatenating in order the $\mathcal{S}_{ij}$'s with $(i,j) \in I_k$. Clearly the height of $\bar{h}_k$ of $\bar{\mathcal{S}}_k$ is given by $\Sigma_{(i,j)\in I_k}(f_{ij} + h_i + g_{ij})$.

The parameters that determine the transition from the $\mathcal{S}_i$'s to the $\bar{\mathcal{S}}_k$'s are the *cutting widths* w_{ij}, the numbers of new levels added at the bottom and at the top f_{ij}, g_{ij}, and the way in which stacks are put together which is governed by the sets

I_k. A *cutting and stacking construction* consists of an infinite sequence of such basic steps and is completely determined by the initial stacks and the parameters described above for each step.

Turning now to an a.m.p.t. $(X, \mathcal{B}, \mu, T)$ we denote by $\mathcal{T}(B, n, \varepsilon)$ a Rohlin tower of height n and base B with residual set ε, so that the sets $B, TB, \ldots, T^{n-1}B$ are disjoint and $\mu(|\mathcal{T}|) = 1 - \varepsilon$, where $|\mathcal{T}(B, n, \varepsilon)| = \bigcup_{j=0}^{n-1} T^jB$. We will need the existence of a sequence of R-towers with $|\mathcal{T}_1| \subset |\mathcal{T}|_2 \subset \cdots$ and even more, that for each l, $|\mathcal{T}_{l+1}| \setminus (B_l \cup T^{n_l-1}B_l) \supset |\mathcal{T}_l|$, i.e., $\mathcal{T}_{l+1}$ contains $\mathcal{T}_l$ in its "interior." To achieve this take some sequence $\varepsilon_i \searrow 0$ and $n_1 < n_2 < \cdots$ tending to infinity rapidly enough so that

$$\sum_{i=l}^{\infty} \frac{n_i}{n_{i+1}} < \tfrac{1}{4}\varepsilon_l \qquad \text{for } 1 \leqq l < \infty.$$

Start with $\mathcal{T}_n^1(B_1^1, n_1, \frac{1}{2}\varepsilon_1)$, and $\mathcal{T}_2^1(B_2^1, n_2, \frac{1}{2}\varepsilon_2)$. If $B_2^1 \cup T^{n_2-1}B_2^1$ intersects $|\mathcal{T}_1^1|$ at all, change $\mathcal{T}_1^1$ to $\mathcal{T}_1^2$ by omitting whole columns (i.e., sets of form $\bigcup_0^{n_1-1} T^jA$ where $A \subset B_1^1$) so that $|\mathcal{T}_1^2(B_1^2, n_1, \varepsilon_1^2)|$ is now disjoint from $B_2^1 \cup T^{n_2-1}B_2^1$. Clearly we have omitted a set of measure at most $2n_1/n_2$. Choose now some $\mathcal{T}_3^1(B_3^1, n_3, \frac{1}{2}\varepsilon_3)$, and modify in a similar way $\mathcal{T}_2^1$ to $\mathcal{T}_2^2$, by omitting whole columns. This further reduces $\mathcal{T}_1^2$ to $\mathcal{T}_1^3$ by an amount at most $2n_2/n_3$, and reduces $\mathcal{T}_2^1$ by at most the same. We continue this infinitely many times; clearly for fixed l the towers $\mathcal{T}_l^m$ converge to a limit as $m \to \infty$, and we obtain towers $\mathcal{T}_l(B_l, n_l, \bar{\varepsilon}_l)$ in the limit with $\bar{\varepsilon}_l \leqq \varepsilon_l$, and the required inclusion relations. The main result of this section is

THEOREM 1. *Any a.m.p.t. $(X, \mathcal{B}, \mu, T)$ can be obtained from a cutting and stacking construction.*

PROOF. Let $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots$ be a sequence of finite partitions of X so that $\bigvee_1^\infty \mathcal{P}_l = \mathcal{B}$, and let $\mathcal{T}_i$ be a sequence of R-towers with each $|\mathcal{T}_l|$ contained in the "interior" of $|\mathcal{T}_{l+1}|$, $n_l \nearrow \infty$, $\varepsilon_l \searrow 0$. We will now describe a cutting and stacking construction. For the first collection of stacks take $\mathcal{T}_1$ and divide it into pure columns with respect to $\mathcal{P}_1$, that is to say, partition B_1 into maximal sets B_{1i}, $i \leqq j \leqq a_1$ so that for $0 \leqq n < n_1$ the sets T^nB_{1i} lie entirely in a single set of $\mathcal{P}_1$, and then take stacks $\mathcal{S}_i^1(n_1, \mu(B_{1i}))$, $1 \leqq i \leqq a_1$. These stacks correspond exactly to R-towers with height n_1 and have base B_{1i}, and we can also think of each level in $\mathcal{S}_i^1$ labeled by the index of $\mathcal{P}_1$ that indicates that atom of $\mathcal{P}_1$ to which the corresponding level in $\mathcal{S}_{1i}(B_{1i}, n_1, \varepsilon_{1i})$ belongs.

Take now $\mathcal{T}_2$ and divide it into maximal $\mathcal{P}_2$-pure columns, that arise by partitioning B_2, the base of $\mathcal{T}_2$, into sets B_{2k}. We want the stacks in the second step of our cutting and stacking construction to correspond to the towers $\mathcal{T}_{2k}(B_{2k}, n_k, \varepsilon_{2k})$. This dictates both the cutting widths, the number of new labels (there is some freedom in assigning new levels between successive $\mathcal{T}_1$-columns in the $\mathcal{T}_{2k}$ to the end of the lower or beginning of the upper $\mathcal{T}_1$-column), and the index sets. It is exactly the property that $|\mathcal{T}_1|$ be contained in the interior of $\mathcal{T}_2$ that makes this possible. Once again we label the levels of the abstract stacks using the $\mathcal{P}_2$-labels in $\mathcal{T}_2$. This procedure is continued infinitely often and concludes in a cutting and stacking construction defining a transformation S that also has partitions $\bar{\mathcal{P}}_1 \subset \bar{\mathcal{P}}_2 \subset \cdots$ so that the processes $(S, \bar{\mathcal{P}}_i)$ are isomorphic to $(T, \mathcal{P}_i)$ and so that $\bigvee_{i=1}^{\infty} \bigvee_{-\infty}^{\infty} S^i \bar{\mathcal{P}}_i$ generates the full σ-algebra of S. This shows that S is isomorphic to T and completes the proof. □

To realize an abstract cutting and stacking construction as an interval exchange one simply places the levels side by side from left to right and adds the new levels in a single step, in some order on the right. We illustrate in Fig. 1 a representation of a cutting and stacking construction that begins with a single stack $\mathcal{S}_1(2, \frac{1}{8})$, $\{w_{ij} : \frac{1}{32}, \frac{1}{32}, \frac{1}{16}\}$, $f_{1j} = g_{1j} = 1$, $1 \leqq j \leqq 3$, $I_1 = \{11, 12\}$, $I_2 = \{13\}$.

Initially we have two intervals, $[0, \frac{1}{8}]$, $[\frac{1}{8}, \frac{1}{4}]$, with the first interval in the domain of our function — but not in the range, and vice versa for the second. Adding the further intervals to the right puts both intervals in the domain and in the range — where to do the first requires that $[\frac{1}{8}, \frac{1}{4}]$ be divided into the intervals. It follows from the proof of Theorem 1 that any transformation has a cutting and stacking model where the f's and g's are always at least one, so that intervals need be subdivided at most once, and we have thus established

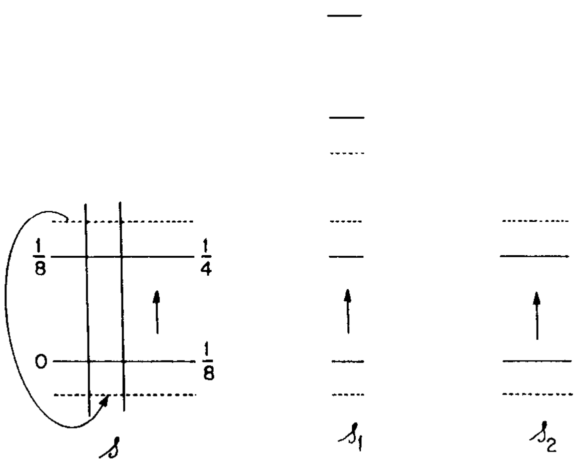

Fig. 1.

THEOREM 2. *Any a.m.p.t. is isomorphic to an interval exchange transformation* $T:[0,1]\to[0,1]$ *of the following type:*

(1) *sub-intervals* $I_1, I_2, I_3, \dots$ *are given by*

$$I_j = (t_{j-1}, t_j)$$

with $0 = t_0 < t_1 < t_2 < \cdots$, $\lim t_j = 1$;

(2) *there exist real constants* $\{a_j\}$, *so that for* $t \in I_j$, $T(t) = t + a_j$;

(3) *the only accumulation point of* $\{t_{j-1} + a_j\} \cup \{t_j + a_j\}$ *is* 1;

(4) T *is one-to-one.*

§2. Geometric models for ergodic measure-preserving transformations and flows

We now show how to extend the standard suspension of interval exchange maps (cf. [1]) to ergodic measure-preserving transformations, so as to obtain

THEOREM 3. *Given any a.m.p.t.* $(X, \mathcal{B}, \mu, T)$, *there exists a* C^∞*-flow* ϕ *on an open 2-manifold M, preserving a finite measure, and a transverse curve* γ *on M, such that T is isomorphic to the Poincaré return map of* ϕ *in* γ.

PROOF. From Theorem 2 it is sufficient to make this construction for an infinite interval exchange transformation, say f, defined on a set of intervals $\{I_i\}_{i\in N}$, of lengths $m(I_i)$.

We begin as for the standard suspension: take the square $]0,1[\times[0,1]$, with constant vector field pointing upwards $\partial/\partial y$, and glue to it strips $]0,m(I_j)[\times[0,1]$, with the same constant field, identifying:

$$]0,m(I_j)[\times\{0\} \qquad \text{with } I_j\times\{1\},$$

$$]0,m(I_j)[\times\{1\} \qquad \text{with } f(I_j)\times\{0\}.$$

The only problem is the existence of extremities for the interval $I_j\times\{1\}$ and $f(I_j)\times\{0\}$ which are boundary points for the surface and where the flow is not defined (Fig. 2). By slowing down the flow in a C^∞-way in a neighbourhood of

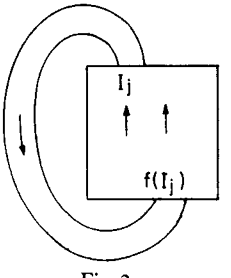

Fig. 2.

those points we can create singularities, and remove them; it is clear that we obtain in this way a C^∞-flow on an open manifold, satisfying the proposition, except maybe for the finiteness of invariant measure. We shall give a local model that will ensure this also.

We take the rectangle $[0,1[\times[-1,1]$, and define on it two vector fields:

$$X_1=\frac{\partial}{\partial Y}\quad\text{(original vector field)},$$

$$X_2(x,y)=x\frac{\partial}{\partial x}-y\frac{\partial}{\partial y}\quad\text{(model for perturbation)}.$$

Choosing $0<B<1$, we take any C^∞-function φ on the rectangle such that:

$$\varphi=1\quad\text{outside } [0,B]\times[-B,B],$$

$$\varphi>\frac{1}{4}\quad\text{outside }\left[0,\frac{3B}{4}\right]\times\left[\frac{-3B}{4},\frac{3B}{4}\right],$$

$$\varphi<\frac{1}{2}\quad\text{on }\left[0,\frac{3B}{4}\right]\times\left[\frac{-3B}{4},\frac{3B}{4}\right],$$

$$\varphi=0\quad\text{on }\left[0,\frac{B}{2}\right]\times\left[\frac{-B}{2},\frac{B}{2}\right],$$

and define: $X=\varphi X_1+(1-\varphi)X_2$.

Take as model those trajectories of X which lie under the separatrice leaving the singular point $(0,0)$. We can restrict our attention to the flow box $[0,B]\times[-B,B]$ (see Fig. 3), since nothing is changed outside; any trajectory of the model enters the box at a point $(x,-B)$ of the lower side, and leaves it by the upper side after a time t_x. Since one gets an invariant measure for the flow by

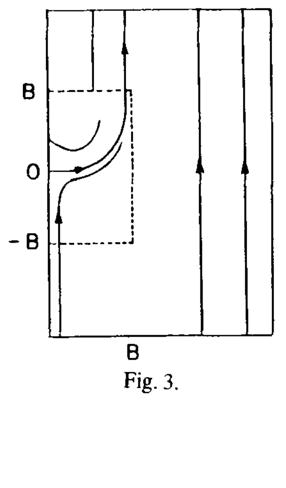

Fig. 3.

integrating the time along the trajectories with respect to the Lebesgue measure for a transverse curve, it is enough to bound $\int_0^B t_x dx$.

An elementary study shows that

$$t_x < 8B \qquad \text{if } x > \frac{3B}{4},$$

$$t_x < 8B + 2\log\left(\frac{3B}{4x}\right) \qquad \text{if } x < \frac{3B}{4};$$

hence

$$\int_0^B t_x dx < \frac{3B}{2} + 8B^2$$

so that the invariant measure is only increased $3B/2 + 6B^2$ by the perturbation. Hence, applying this perturbation near each boundary point in a neighbourhood of decreasing diameter, one gets finite invariant measure.

If we want to find a smooth model for a flow T_t we begin by representing the flow as a special flow built under a function as indicated by the Ambrose–Kakutani theorem. Then we apply Theorem 3 to the base transformation of the special flow. This gives us a smooth model of a flow which is obtained from T_t by a measurable time change. At this point we recall the technique of [4] which will enable us to smooth out this measurable time change without changing the isomorphism class. We assume that T_t is ergodic and assert the following proposition whose proof is a minor modification of the proof of Theorem 1 in [4].

PROPOSITION 4. *If ϕ_t is a smooth ergodic measure-preserving flow on a non-compact manifold M, then for any measurable time change $\tilde{\phi}_t$ there is a smooth time change $\bar{\phi}_t$ (so that $\bar{\phi}_t$ is also a C^∞-flow) so that $\bar{\phi}_t$ is isomorphic to $\tilde{\phi}_t$.*

Note that the non-compactness of M is essential — we have to be able to drive small measurable changes out to infinity so that they eventually disappear. The ergodicity is also used in an essential way — since if there would be compact invariant sets then once again the corrections would condense in a finite portion of the manifold. We also note that the smooth time change for $\bar{\phi}_t$ can be chosen to be bounded over the whole manifold. If we were to allow unbounded velocity changes, then the argument of [4] could be simplified.

Now combining Proposition 4 with Theorem 3 we obtain

THEOREM 5. *Given (X, μ, T_t), an ergodic measure-preserving flow, there is a C^∞-flow ϕ_t on an open 2-manifold M that preserves a smooth measure and is isomorphic to T_t.*

§3. Generalized powers

In [5] we make reference to the fact that if (X, T) and (Y, S) are arbitrary a.m.p.t. then there is a positive integer-valued function p on X so that $\bar{S}(x) = T^{p(x)}(x)$ is invertible and isomorphic to S. Note that in contrast to the situation in Dye's theorem, the orbits of $\bar{S}$ are strictly sub-orbits of the orbits of T. The novelty is that p is positive integer-valued. In case p is integrable J. Neveu [3] has shown that the entropy of T^p equals $(\int p) \cdot h(T)$. Nonetheless it will follow that we can have situations where both $\int p$ and $h(T)$ are infinite and yet $h(T^p) = 0$.

The key remark is that if $(X, \mathcal{B}, \mu, T)$ is e.m.p.t. and A and B are any two such sets of positive equal measure, then there is a positive integer-valued function p defined on A so that $T^p(A) = B$ up to a set of measure zero. Thus one can simply copy a cutting and stacking construction representation for S inside (X, T) using partial transformations of the type that we have just described. If one carries out the procedure that we have just described, then one gets only that T has a *factor* that is isomorphic to S, since we have not provided for the fact that in the cutting and stacking construction the levels generate the whole σ-algebra. In order to guarantee that we must exercise more care. The key is the fact that many steps in a cutting and stacking construction may be combined to a single step. This can easily be seen by reading off from the stacks at the end of a finite sequence of steps the necessary parameters just as we did in §1 when we read these parameters from the pure columns of a R-tower. Consequently, we can always assume that the cutting widths, w_{ij}, are as small as we please in any given stage.

THEOREM 4. *For any pair of a.m.p.t.'s $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ there is a $\mathcal{B}$-measurable function $p: X \to \mathbf{N}$ such that $\bar{S}(x) = T^{p(x)}(x)$ is invertible μ-a.e. and $(X, \mathcal{B}, \mu, \bar{S})$ is isomorphic to $(Y, \mathcal{C}, \nu, S)$.*

PROOF. Fix some cutting and stacking construction that represents S and a sequence of finite partitions $\mathcal{P}_i$ of X such that $\bigvee_1^\infty \mathcal{P}_i = \mathcal{B}$. To be definite, let us suppose that $\mathcal{P}_i$ has 2^i elements each of measure 2^{-i}. If necessary, amalgamate the first few steps in the cutting and stacking construction for S so that all levels in the initial stacks have measure less than 10^{-1}, while their total mass is at least $1 - 10^{-1}$. It is now clear that one can choose sets in X to represent these levels in such a way that they will approximate $\mathcal{P}_1$ with an error at most 2×10^{-1}. By our basic remark, one can now define p_1 on the sets in X that represent the non-final levels in the initial stacks, so that the partial transformation

$$S_1(x) = T^{p_1(x)}(x)$$

models the cutting and stacking transformation on the initial stacks.

Next one examines how $\mathscr{P}_2$ divides these sets in X and purifies these towers using S_1. More precisely, if $A_1, A_2, \ldots, A_n \subset X$ represented one of the initial stacks, so that $S_1 A_i = A_{i+1}$, $1 \leqq i < h$, we partition A_1 into maximal sets A_1^j of positive measure with the property that for each $0 \leqq n < h$, $S_1^n(A_1^j)$ lies entirely in one set of $\mathscr{P}_2$, and then consider $S_1^n(A_1^j)$, $0 \leqq n < h$. Let ε_2 denote the maximum value assumed by $\mu(S_1^n(A_1^j))$ over the various stacks, and let m_2 equal the number of these sets. Knowing ε_2 and m_2 we now amalgamate the next few steps in the cutting and stacking construction so that the cutting widths are small enough and the total measure large enough so that we can model them by sets in X which will approximate $\mathscr{P}_2$ by an error at most 10^{-2}. Having done so, we define p_2 as an extension of p_1, so that $S_2(x) = T^{p_2(x)}(x)$ extends S_1 and S_2 has R-towers that copy the above steps in the cutting and stacking construction.

This process is carried out infinitely many times and gives a model in (X, T) for the cutting and stacking construction of S with levels that approximate $\mathscr{P}_i$ better and better and hence generate $\mathscr{B}$. Thus the resulting $\bar{S}$, the union of $S_1 \subset S_2 \subset \cdots$, is isomorphic to S. □

References

1. P. Arnoux, *Exchanges d'intervalle et flots sur les surfaces*, in *Théorie Ergodique*, Monographie No. 29 de l'Enseignement Mathématiques, 1980, pp. 5–38.

2. N. A. Friedman, *Introduction to Ergodic Theory*, New York, 1970.

3. J. Neveu, *Une démonstration simplifiée et une extension de la formule d'Abramov sur l'entropie des transformations induites*, Z. Wahrscheinlichkeitstheor. Verw. Geb. **13** (1969), 135–140.

4. D. S. Ornstein and M. Smorodinsky, *Continuous speed changes for flows*, Isr. J. Math. **31** (1978), 161–168.

5. D. S. Ornstein and B. Weiss, *Any flow is the orbit factor of any other*, Ergodic Theory and Dynamical Systems **4** (1984), 105–116.