

ERGODIC GROUP AUTOMORPHISMS AND SPECIFICATION

D. A. Lind
University of Washington
Seattle, Washington 98195

Dedicated to the memory of Rufus Bowen

§1. Introduction.

I first want to discuss how I became interested in finding out which ergodic automorphisms of compact groups satisfy a property called specification, and then describe the answer for ergodic toral automorphisms. This settles a question raised by Sigmund [15, §2]. One consequence of the answer is that Markov partitions, such as those found by Adler and Weiss [1] for two-dimensional toral automorphisms and by Bowen [3] for hyperbolic toral automorphisms, cannot be constructed for nonhyperbolic automorphisms. There seem to be essential differences between the dynamic behavior of hyperbolic and nonhyperbolic automorphisms. Some questions that can be answered in the hyperbolic case using the Markov partition machinery, e.g. concerning the distribution of periodic orbits, remain open for nonhyperbolic toral automorphisms.

§2. Splitting skew products.

When dealing with the ergodic properties of group automorphisms, transformations of the following form often arise. Let $U: X \rightarrow X$ be an invertible measure-preserving transformation (hereafter shortened to "map") of a Lebesgue space (X, μ) , $S: G \rightarrow G$ be a (continuous, algebraic) automorphism of a compact metrizable group G written additively, and $\alpha: X \rightarrow G$ be a measurable function. Form the skew product $U \times_{\alpha} S: X \times G \rightarrow X \times G$ defined by $(U \times_{\alpha} S)(x, g) = (Ux, Sg + \alpha(x))$. Such skew products arise, for example, when there is a closed subgroup H of G that is invariant under S . For by taking a Borel cross-section

to the projection $G \rightarrow G/H$, the automorphism S can be written as the skew product of the factor automorphism $S_{G/H}$ with the restriction S_H of S to H (see [8, p. 209] for details).

While investigating the Bernoullicity of group automorphisms in [8], I noticed that by using Thouvenot's relative isomorphism theory one could show that skew products with ergodic group automorphisms are always isomorphic to direct products, via an isomorphism that preserves the group fibers. This means that there is a map $W: X \times G \rightarrow X \times G$ of the form $W(x, g) = (x, W_x(g))$ such that

$$(1) \quad (U \times_{\alpha} S)W = W(U \times S).$$

Unfortunately, Thouvenot's theory gives no information about the individual maps $W_x: G \rightarrow G$. Demanding that they be group translations, i.e. $W_x(g) = g + \beta(x)$, when substituted into (1), amounts to solving the functional equation

$$(2) \quad \alpha(x) = \beta(Ux) - S\beta(x),$$

where α , U , and S are known, and the measurable function $\beta: X \rightarrow G$ is to be found. It turns out that this is always possible.

Splitting Theorem [9]. If S is an ergodic group automorphism, then for arbitrary α and U the functional equation (2) can be solved for β .

One application of this result is the simplest proof so far of Katznelson's result [7] that ergodic toral automorphisms are Bernoulli (see [9, §5]), one that avoids the Diophantine approximation arguments used in previous proofs. Other applications are mentioned in [9, §1].

Bowen's property "specification" plays a key role in the proof of the Splitting Theorem given in [9], so let me first review this property, and then indicate its use.

Let (Y, d) be a compact metric space, and $f: Y \rightarrow Y$ be continuous. The transformation f obeys specification if for every $\epsilon > 0$ there is an $M(\epsilon)$ such that for every $r \geq 2$ and r points $y_1, \dots, y_r \in Y$, and for every set of

integers $a_1 \leq b_1 < a_2 \leq b_2 < \dots < a_r \leq b_r$ and p with $a_i - b_{i-1} \geq M(\epsilon)$ ($2 \leq i \leq r$) and $p \geq b_r - a_1 + M(\epsilon)$, there is a point $y \in Y$ with $d(f^n y, f^n y_i) < \epsilon$ for $a_i \leq n \leq b_i$, $1 \leq i \leq r$, and with $f^p(y) = y$. Basically this definition means that given specified pieces $\{f^n y_i : a_i \leq n \leq b_i\}$ of orbits of different points for disjoint blocks of times, if there is enough space between these blocks then these pieces can be well approximated during the same time blocks by the orbit of a single periodic point.

Several important dynamical systems obey specification, including hyperbolic toral automorphisms and subshifts of finite type (see [5, Ch. 21]). Bowen introduced this property to produce the unique equilibrium measure for an Axiom A diffeomorphism [3, Ch. 4] and to construct Markov partitions for such diffeomorphisms [3, Ch. 3]. Sigmund [12], [13], [14], [15] and Kamae [6] have used variants of specification to study orbits and generic properties of invariant measures to generalize number-theoretic facts about decimal expansions.

Actually, solving (2) involves only the orbit copying part of specification. Hence say that $f:Y \rightarrow Y$ obeys weak specification if it satisfies the definition of specification except for the periodic point condition $f^p(y) = y$. Weak specification was used by Ruelle [11] in studying the statistical mechanics of lattice actions.

Let me now sketch how to solve the functional equation (2) for those automorphisms S obeying weak specification. Let S be an ergodic automorphism of a compact abelian group G , and equip G with a translation invariant metric ρ . Suppose that S satisfies weak specification on (G, ρ) . Let $U:X \rightarrow X$ and $\alpha:X \rightarrow G$ be measurable. I will find β by constructing approximating solutions β_k defined on successively larger parts of X .

To begin, choose positive ϵ_k with $\sum \epsilon_k < \infty$. Let $M(\epsilon_1)$ be given by the weak specification property of S , and choose an integer h_1 so that $M(\epsilon_1)/h_1 < \epsilon_1$. Let $F_1 \subset X$ be a Rohlin base for U of height $h_1 + M(\epsilon_1)$, and let $E_1 = \cup \{U^j F_1 : 0 \leq j < h_1\}$. The base F_1 can be chosen so that $\mu(E_1) > 1 - 3\epsilon_1$. Define β_1 arbitrarily but measurably on F_1 . The functional equation (2) forces the definition of β_1 inductively up the stack. Specifically, for

$x \in F_1$,

$$\beta_1(Ux) = S\beta_1(x) + \alpha(x) ,$$

$$\begin{aligned} \beta_1(U^2x) &= S\beta_1(Ux) + \alpha(Ux) \\ &= S^2\beta_1(x) + S\alpha(x) + \alpha(Ux), \end{aligned}$$

and in general

$$\beta_1(U^jx) = S^j\beta_1(x) + \alpha_j(x) \quad (x \in F_1, 0 \leq j < h_1)$$

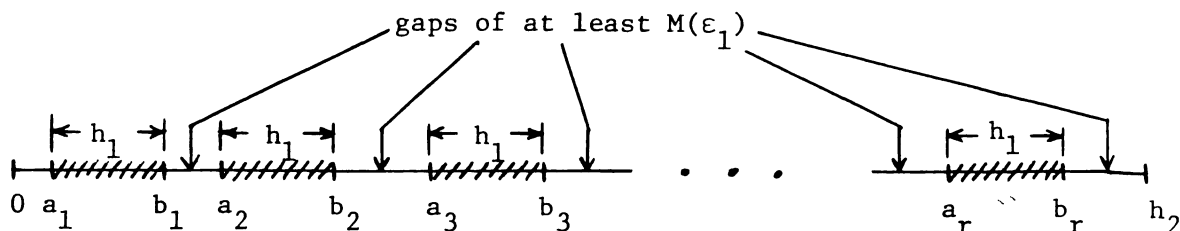
where

$$\alpha_j(x) = \sum_{m=0}^{j-1} S^{j-m-1} \alpha(U^m x).$$

This defines β_1 on the stack E_1 , and it satisfies (2) on all but the top level.

Similarly, for the given $\epsilon_2 > 0$, find $M(\epsilon_2)$, h_2, F_2, E_2 , where h_2 is much larger than h_1 . Now comes the essential point. Once β_2 is defined at $x \in F_2$, its definition on the orbit piece $\{U^jx: 0 \leq j < h_2\}$ is forced by (2). However, β_1 is already defined on certain subpieces of this piece, namely the blocks of time when the orbit of x is in E_1 . It may not be possible to select a value for $\beta_2(x)$ so that on the subpieces of the orbit where β_1 is already defined, β_2 will agree with, or even be close to, to previous function β_1 . The role of specification is to show that such a selection is possible.

Suppose I let $\beta_2'(x) = g_0$, and see what the trouble is. Now β_1 is already defined on certain subpieces, say $\{U^n x: a_i \leq n \leq b_i, 1 \leq i \leq r\}$, where $b_i = a_i + h_1$, and by construction $a_{i+1} - b_i \geq M(\epsilon_1)$. The definition of β_2' on



those pieces is forced:

$$\beta_2'(U^{a_i+j} x) = S^{a_i+j} g_0 + \alpha_{a_i+j}(x)$$

for $1 \leq i \leq r$, $0 \leq j < h_1$. Also,

$$\beta_1(U^{a_i+j} x) = S^j \beta_1(U^{a_i} x) + \alpha_j(U^{a_i} x).$$

Since

$$\alpha_{a_i+j}(x) - \alpha_j(U^{a_i} x) = S^j \alpha_{a_i}(x),$$

subtraction gives

$$\beta_2'(U^{a_i+j} x) - \beta_1(U^{a_i+j} x) = S^j [S^{a_i} g_0 - \beta_1(U^{a_i} x) + \alpha_{a_i}(x)].$$

Since the bracketed expression is independent of j , the error on the subpiece $\{U^n x: a_i \leq n \leq b_i\}$ is the orbit of the point in brackets, different points for different i . The time gaps between these subpieces are at least $M(\epsilon)$, so weak specification can be used to adjust our original choice of $\beta_2'(x)$ to decrease these errors to be uniformly less than ϵ_1 . For by weak specification, there is a $g_1 \in G$ such that

$$\rho(S^{a_i+j} g_1, S^j [S^{a_i} g_0 - \beta_1(U^{a_i} x) + \alpha_{a_i}(x)]) < \epsilon_1$$

for $0 \leq j < h_1$, $1 \leq i \leq r$. Define $\beta_2(x) = \beta_2'(x) - g_1$. Then it is easy to check using translation invariance of ρ that if β_2 is defined up the stack using (2), then $\rho(\beta_2(U^n x), \beta_1(U^n x)) < \epsilon_1$ whenever $U^n x \in E_1$, $0 \leq n < h_2$.

Thus β_2 is defined on more of X , solves (2) where defined, and is uniformly close to β_1 where the latter is defined.

Similarly, construct β_k defined on E_k with $\mu(E_k) \rightarrow 1$, such that β_k solves (2) where defined, and such that β_{k+1} and β_k are within ϵ_k on E_k . Since $\sum \epsilon_k < \infty$, $\{\beta_k\}$ is almost uniformly Cauchy, so converges to a measurable function β defined almost everywhere on X which solves (2). Details of this argument are in [9, §4].

§3. Specification.

If every ergodic group automorphism satisfied weak specification, then the simple argument in §2 would be all that is needed to solve (2). A general ergodic automorphism can be built up from certain basic automorphisms that do satisfy weak specification by using factors, products, inverse limits, and extensions by basic automorphisms (see [9, §7]). All but the last process preserve weak specification. Unfortunately, there are extensions of homeomorphisms obeying weak specification by a basic automorphism that do not obey weak specification. In fact, a toral automorphism S' is given below having an invariant subgroup H such that both $S'_{G/H}$ and S'_H obey weak specification, but S' does not.

Hence at least some of the complication in [9] of solving (2) for a general ergodic group automorphism seems intrinsic to the specification approach. Dan Rudolph [10] has found another way of solving (2) by using a relativised isomorphism theorem for measure-preserving actions of a skew product of the integers with a compact group.

Which group automorphisms obey specification?

The shift whose state space is a finite group certainly does, and I believe so do all ergodic automorphisms of totally disconnected compact abelian groups. This is true in many cases, but I have not yet found a general proof.

However, for ergodic toral automorphisms there is a complete answer. Such automorphisms come in three flavors, depending on the spectral properties of the associated linear map.

(1) Hyperbolic automorphisms have no eigenvalues on the unit circle (i.e. no unitary eigenvalues).

(2) Central spin automorphisms have some unitary eigenvalues, and on the eigenspace of the unitary eigenvalues the associated linear map is an isometry (with an appropriate metric); this means that the Jordan blocks for the unitary eigenvalues have no off-diagonal 1's.

(3) Central skew automorphisms are what's left, namely those with off-diagonal 1's in the Jordan block of some unitary eigenvalue.

Central spin and central skew automorphisms occur. Let

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -4 & 2 & -4 \end{pmatrix} .$$

The eigenvalues for S are $\sqrt{2} - 1 \pm i \sqrt{2\sqrt{2} - 2}$ (both unitary), and $-\sqrt{2} - 1 \pm \sqrt{2\sqrt{2} + 2}$ ($\approx -4.61, -0.21$). Hence S is a central spin automorphism. Dimension 4 is least possible for such an automorphism. Also, if I denotes the 4×4 identity matrix, then

$$S' = \begin{pmatrix} S & I \\ 0 & S \end{pmatrix}$$

is a central skew automorphism. Dimension 8 is least possible for such an automorphism.

The specification behavior of each of these classes is different.

Theorem. (i) Hyperbolic automorphisms obey specification.

(ii) Central spin automorphisms obey weak specification, but never obey specification.

(iii) Central skew automorphisms never obey even weak specification.

Part (i) is due to Bowen [2]. The first part of (ii) is proved in [9] as part of the general solution to (2). In the next two sections I briefly indicate the geometric ideas behind (iii) and the negative part of (ii).

§4. Central skew automorphisms.

Let S be a central skew automorphism of the d -dimensional torus $\Pi^d = \mathbb{R}^d / \mathbb{Z}^d$, and denote its lifting to a linear map of \mathbb{R}^d by the same symbol. Denote by $E^s, E^u, E^c \subset \mathbb{R}^d$ the stable, unstable, and central subspaces of S corresponding to the eigenvalues of S inside, outside, and on the unit circle, respectively. Then $\mathbb{R}^d = E^s \oplus E^c \oplus E^u$. It is not dangerous to identify these

subspaces with their projections to Π^d since by ergodicity of S they are disjoint from \mathbb{Z}^d .

Since S is central skew, there are disjoint 2-dimensional subspaces E_1^C and E_2^C of E^C such that the restriction of S to $E_1^C \oplus E_2^C$ has the form

$$\begin{pmatrix} Q & I \\ 0 & Q \end{pmatrix},$$

where Q is a rotation and I is the 2×2 identity matrix. For example, in the central skew automorphism S' given in §3,

$$\begin{pmatrix} \sqrt{2} - 1 & \sqrt{2} \sqrt{2 - 2} \\ -\sqrt{2} \sqrt{2 - 2} & \sqrt{2} - 1 \end{pmatrix}$$

with respect to an appropriate basis. E^C could be larger than $E_1^C \oplus E_2^C$. Note that the matrix of S^j on $E_1^C \oplus E_2^C$ is

$$(3) \quad \begin{pmatrix} Q^j & jQ^{j-1} \\ 0 & Q^j \end{pmatrix}$$

For convenience, give Π^d a translation-invariant product metric ρ inherited from the eigenspaces, and let $B^S(\epsilon)$ denote the ϵ -ball around 0 in the stable subspace (projected to Π^d), and similarly define $B^u(\epsilon)$, $B^c(\epsilon)$, $B(\epsilon)$ (for details, see [9, §6]).

Note that $u \in \Pi^d$ stays within ϵ of $t \in \Pi^d$ under the first n iterates of S precisely when $u-t$ stays within ϵ of 0. Hence it is important to analyse

$$\begin{aligned} A(n, \epsilon) &= \{t \in \Pi^d : \rho(S^j t, 0) < \epsilon \text{ for } 0 \leq j \leq n\} \\ &= \bigcap_{j=0}^n S^{-j} B(\epsilon) \\ &= A^S(\epsilon, n) \oplus A^c(\epsilon, n) \oplus A^u(\epsilon, n), \end{aligned}$$

where $A^S(\epsilon, n) = \bigcap_{j=0}^n S^{-j} B^S(\epsilon)$, etc.

Since S^{-1} expands on E^S , $A^S(\epsilon, n)$ is "essentially" $B^S(\epsilon)$, i.e. There are absolute constants K_1 and K_2 independent of ϵ and n such that $K_1 B^S(\epsilon) \subset$

$$A^S(\epsilon, n) \subset K_2 B^S(\epsilon).$$

Similarly, $A^u(\epsilon, n)$ is "essentially" $S^{-n} B^u(\epsilon)$.

For the central direction, let $t_i \in E_i^C$, so $t_1 \oplus t_2 \in E_1^C \oplus E_2^C$. Then from (3),

$$S^j(t_1 \oplus t_2) = (Q^j t_1 + jQ^{j-1} t_2) \oplus Q^j t_2.$$

Since this is to be within ϵ of 0 for $0 \leq j \leq n$, and Q is an isometry, clearly $\rho(t_1, 0) < \epsilon$, $\rho(t_2, 0) < \epsilon$, and $\rho(Q^j t_1 + jQ^{j-1} t_2, 0) < \epsilon$. The last inequality forces $\rho(t_2, 0) < 2\epsilon/j$, so in particular $\rho(t_2, 0) < 2\epsilon/n$. This thinness is one eigendirection, decreasing with n , is the key geometrical fact in the proof.

These observations are enough to show that S does not obey weak specification. Fix an $\epsilon > 0$ small enough so that the projection of $B(5\epsilon) \subset \mathbb{R}^d$ to Π^d is injective. Let $M > 0$ be given. I will find $t_1, t_2 \in \Pi^d$ and integers $0 = a_1 < b_1 < a_2 < b_2$ with $a_2 - b_1 = M$ such that no $t \in \Pi^d$ has $\rho(S^j t, S^j t_i) < \epsilon$ for $a_i \leq j \leq b_i$, $i = 1, 2$.

The intersection $S^M B^u(\epsilon) \cap [B^S(5\epsilon) \oplus B^C(5\epsilon)]$ is finite, say $\{s_1, \dots, s_r\}$. For n sufficiently large, the projection P of $\cup_{i=1}^r \{B_2^C(\epsilon/n) + s_i\}$ to $B_2^C(5\epsilon)$ cannot cover all of $B_2^C(5\epsilon)$. For such an n , choose an element $u \in B_2^C(5\epsilon)$ and an integer m such that $u + B_2^C(\epsilon/m)$ is disjoint from P .

Now let $t_1 = 0$, $t_2 = u$, $a_1 = 0$, $b_1 = n$, $a_2 = n + M$, and $b_2 = a_2 + m$. Suppose there were a $t \in \Pi^d$ with $\rho(S^j t, S^j t_i) < \epsilon$ for $a_i \leq j \leq b_i$. Then $t \in A(\epsilon, n)$ and $S^{n+M} t \in u + A(\epsilon, m)$. Hence

$$S^{n+M} t \in [S^{n+M} A(\epsilon, n)] \cap [u + A(\epsilon, m)].$$

Since the component of $A(\epsilon, n)$ in the E_2^C direction is $B_2^C(\epsilon/n)$, the projection to $B_2^C(5\epsilon)$ of the first term in this intersection is contained in P , while that of the second is by construction disjoint from P . This contradiction completes the argument.

§5. Central spin automorphisms.

Central spin automorphisms obey weak specification [9, §6], but the argument here shows that they never obey specification. In fact, if S is such an automorphism of Π^d , then all sufficiently small $\epsilon > 0$ will have the property that for every $M > 0$ there are $t_1 \in \Pi^d$ and $n > 0$ such that no $t \in \Pi^d$ has $\rho(S^j t, S^j t_1) < \epsilon$ for $-n \leq j \leq 0$ and $S^{n+M} t = t$. This will contradict the specification definition for $r = 1$, $a_1 = -n$, $b_1 = 0$, and $p = n + M$.

Since S has a central spin factor with irreducible characteristic polynomial, and specification is preserved under factors, assume that S has irreducible characteristic polynomial. The purpose of this is to guarantee that $(E^S \oplus E^u) \cap \mathbb{Z}^d = \{0\}$, which follows from irreducibility because then S cannot preserve a nontrivial lattice.

Choose $\epsilon > 0$ as small as in §4. Let $t_1 \in E^u$ such that $\rho(t_1, 0) = 2\epsilon$.

Since S is central spin, its restriction to the central subspace E^c is an isometry, say Q . Hence the identity on E^c can be arbitrarily well approximated by arbitrarily large powers of Q .

Now $\{S^M[B^u(\epsilon) + t_1]\} \cap \{B^S(5\epsilon) \oplus B^c(5\epsilon)\}$ is finite, say $\{s_1, \dots, s_r\}$. Since $0 \notin S^M[B^u(\epsilon) + t_1]$, and $(E^S \oplus E^u) \cap \mathbb{Z}^d = \{0\}$, the projection of s_1, \dots, s_r to $B^c(5\epsilon)$ are all displaced at least some quantity $\delta > 0$ from 0. Choose n so that Q^{n+M} is so close to the identity that for every $s \in B^c(5\epsilon)$ with $\rho(s, 0) \geq \delta$, the map $u \mapsto s + Q^{n+M} u$ has no fixed points $u \in B^c(5\epsilon)$.

Suppose there were a $t \in \Pi^d$ with $\rho(S^j t, S^j t_1) < \epsilon$ for $-n \leq j \leq 0$ and $S^{n+M} t = t$. If u denotes the projection of $S^{-n} t$ to $B^c(5\epsilon)$, then the projection of $S^M t = S^{n+M}(S^{-n} t)$ to $B^c(5\epsilon)$ has the form $s_i + Q^{n+M} u$ for some i . Since $S^{-n} t = S^M t$, their projections u and $s_i + Q^{n+M} u$ must agree. But $\rho(s_i, 0) \geq \delta$, so there are no fixed points of the map $u \mapsto s_i + Q^{n+M} u$ for $u \in B^c(5\epsilon)$, showing that such a t does not exist.

§6. Remarks.

Nonhyperbolic toral automorphisms seem to behave differently from the hyperbolic ones. For example, a modification of the geometric ideas here shows

that for nonhyperbolic automorphisms, every fine enough partition is not weak Bernoulli, although every partition is very weak Bernoulli since the automorphism is a Bernoulli shift. This should be contrasted with Bowen's result [4] that for hyperbolic automorphisms every smooth partition is weak Bernoulli. The geometry also shows clearly certain limits to independence that forced Katznelson [7] to introduce the intermediate idea of "almost weak Bernoulli" in the first proof that ergodic toral automorphisms are Bernoulli. Details concerning these remarks will appear elsewhere.

It follows from the theorem in §3 that Markov partitions in the sense of Bowen [3] do not exist for nonhyperbolic toral automorphisms. For the existence of a Markov partition would imply that the automorphism is a factor of a Markov shift. Such shifts obey specification, and specification is trivially preserved under factors.

Thus nonhyperbolic toral automorphisms are examples of smooth systems for which the usual machinery of Markov partitions is unavailable, but which still can be analysed in detail. Yet many questions about them, which can be answered in the hyperbolic case, remain unsettled. Sample: Are the periodic orbit measures weakly dense in the space of invariant measures? In particular, is there a sequence of periodic orbits that converges weakly to Lebesgue measure, i.e. is uniformly distributed? Nobody seems to know.

References

1. R. Adler and B. Weiss, Similarity of automorphisms of the torus, Mem. Amer. Math. Soc., 98 (1970).
2. Rufus Bowen, Periodic points and measures for Axiom A diffeomorphisms, Trans. Amer. Math. Soc., 154 (1971), 377-397.
3. _____, Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms, Springer Lecture Notes in Math. 470, Berlin, 1975.
4. _____, Smooth partitions of Anosov diffeomorphisms are weak Bernoulli, Israel J. Math. 21 (1975), 95-100.
5. Manfred Denker, Christian Grillenberger, and Karl Sigmund, Ergodic Theory on Compact Spaces, Springer Lecture Notes in Math. 527, Berlin, 1976.
6. Teturo Kamae, Normal numbers and ergodic theory, Proc. 3rd Japan-USSR Symp. Prob. Th., Springer Lecture Notes in Math. 550 (1976), 253-269.
7. Y. Katznelson, Ergodic automorphisms of T^n are Bernoulli shifts, Israel J. Math. 10 (1971), 186-195.
8. D. A. Lind, The structure of skew products with ergodic group automorphisms, Israel J. Math. 28 (1977), 205-248.
9. _____, Split skew products, a related functional equation, and specification, to appear, Israel J. Math.
10. Daniel J. Rudolph, An isomorphism theory for Bernoulli free Z -skew-compact group actions, to appear.
11. David Ruelle, Statistical mechanics on a compact set with Z^2 action satisfying expansiveness and specification, Trans. Amer. Math. Soc. 185, (1973), 237-251.
12. Karl Sigmund, Generic properties of invariant measures for Axiom A diffeomorphisms, Inventiones Math. 11 (1970), 99-109.
13. _____, Ergodic averages for Axiom A diffeomorphisms, Z. Wahrscheinlichkeits Theorie verw. Geb. 20 (1971), 319-324.
14. _____, Mixing measures for Axiom A diffeomorphisms, Proc. Amer. Math. Soc. 36 (1972), 497-504.
15. _____, On dynamical systems with the specification property, Trans. Amer. Math. Soc. 190 (1974), 285-299.