TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 257, Number 1, January 1980

CONSTRUCTING SMALE DIFFEOMORPHISMS ON COMPACT SURFACES

BY

STEVE BATTERSON

ABSTRACT. A necessary condition for an isotopy class on a compact surface to admit a Smale diffeomorphism whose dynamics are a specified set of subshifts of finite type is that the Euler characteristic of the manifold be equal to a sum and difference of certain numbers obtained from the matrices representing the subshifts. In this paper it is shown that this condition is sufficient up to a finite power of the subshifts.

A diffeomorphism f on a compact manifold is a *Smale diffeomorphism* if and only if f satisfies Axiom A, the strong transversality condition, and has only zero dimensional basic sets (see **[3]** or **[10]** for definitions). Smale diffeomorphisms are structurally stable, dense in the C^0 topology on $\mathrm{Diff}(M)$ and appear in every isotopy class **[9]**, **[10]**, **[12]**.

For a Smale diffeomorphism the restriction of the map to a basic set is topologically conjugate to a subshift of finite type and thus can be represented by a nonnegative integer matrix. The *reduced degree* of an integer matrix A is the degree of the mod 2 reduction of the polynomial $\det(I - At)$.

In **[3]** Franks presented results which for a large class of manifolds of dimension greater than 2, related the topology of the manifold to the dynamical behavior of Smale diffeomorphisms on the manifold. In this paper we consider the problem for the 2-dimensional case. More specifically, given irreducible matrices corresponding to subshifts of finite type, we present techniques for isotoping the identity map to a Smale diffeomorphism with powers of the subshifts occurring as basic sets. Combining this with the generalized Morse inequalities yields:

THEOREM. *Let A_1 be a direct sum of irreducible matrices and let A_0 and A_2 be permutation matrices. On any compact surface M a necessary and sufficient condition for the existence of a Smale diffeomorphism isotopic to the identity and with basic sets of index j ($j = 0, 1, 2$) corresponding to A_j^k, for some integer $k \geqslant 1$, is that $d_2 - d_1 + d_0 = x(M)$ where d_j is the reduced degree of A_j and $x(M)$ is the Euler characteristic of M.*

In **[3]** Franks gave an example of a subshift of finite type which does not correspond to a basic set of any Smale diffeomorphism of S^2.

Received by the editors March 2, 1978 and, in revised form, July 3, 1978.
AMS (MOS) subject classifications (1970). Primary 58F20; Secondary 57A05.

COROLLARY 1. *For any isotopy class of diffeomorphisms on a two dimensional manifold and any subshift of finite type, there is a power of the subshift that can be realized as a basic set on a Smale diffeomorphism which is in a power of the isotopy class.*

COROLLARY 2. *Any subshift of finite type has a power whose suspension can be realized as a basic set of a structurally stable flow on S^3 and a flow on R^3 with everything moving in at infinity.*

I would like to thank John Franks for several valuable conversations and suggestions.

Preliminaries. In this section we present the background which is required to prove the theorem. Further definitions are available in **[3]**, **[10]**, and **[11]**.

Let A be an $n \times n$ matrix with nonnegative integer elements. To obtain the subshift of finite type associated with A, consider the graph with vertices $\{1, \ldots, n\}$ and A_{ij} unit length, oriented edges from vertex i to j. Let S_A be the set of maps from the reals to the graph which map integers to vertices and preserve both orientation and arclength. Define $\sigma: S_A \to S_A$ by

$$\sigma(g)(t) = g(t+1)$$

and give S_A the compact open topology. The map σ is called a *subshift of finite type*. Subshifts of finite type can also be defined via 0-1 matrices and/or bisequences (instead of graphs).

A nonnegative integer matrix A is *irreducible* if and only if for each index pair i and j, there exists a positive integer k such that $(A^k)_{ij} > 0$. Bowen in **[1]** proves that the action of a Smale diffeomorphism on a basic set is topologically conjugate to a subshift of finite type. In **[2]** Bowen and Lanford show that for the subshift of finite type $\sigma: S_A \to S_A$, the zeta function is

$$\zeta_\sigma(t) = \frac{1}{\det(I - At)}.$$

Reducing by $Z/2Z$ the coefficients of the rational function ζ_σ yields the *reduced zeta function*. Thus the study of reduced degrees is motivated by its equivalence with the degree of the reciprocal of the reduced zeta function representing a subshift of finite type.

We now review the definitions and lemma from the section in **[3]** on handle sets. We shall only be concerned with the case where $m = 2$.

If f is a Smale diffeomorphism on an m-dimensional manifold M, then a *filtration* for f is a collection $\{M_i\}_{i=0}^m$ of m-dimensional submanifolds with boundary such that

(1) $M_0 \subset M_1 \subset \cdots \subset M_m = M$,

(2) $M_{j-1} \cup \overline{f(M_j)} \subset \operatorname{int} M_j$,

(3) if $Y_j = \overline{M_j - M_{j-1}}$ then $\bigcap_{n \in Z} f^n(Y_j)$ is the union of all basic sets of index j.

A *handle set* H for the jth level of a filtration is a finite union of handles $\{H_i = D_i^j \times D_i^{m-j}\}$ attached to the boundary of M_{j-1} by disjoint embeddings of the $S_i^{j-1} \times D_i^{m-j}$ and embedded disjointly in int M_j.

For $x \in D_i^j \times p$ let $W_i^u(x) = D_i^j \times p \subset H_i$ and if $x \in q \times D_i^{m-j}$ let $W_i^s(x) = q \times D_i^{m-j}$. In this paper we are primarily concerned with handle sets for the 1st level of a filtration on a two dimensional manifold (i.e. $j = 1$ and $m = 2$).

The handle set $H = \bigcup H_i$ is said to be *compatible with f* if

(1) If x, $f(x) \in H$ then $\inf f(W^u(x)) \supset W^u(f(x))$ and $f(W^s(x)) \subset$ int $W^s(f(x))$,

(2) If $x, f(x) \in$ int H and $v \in T_x(W^s(x))$, $w \in T_x(W^u(x))$ then $|df(v)| < \lambda|v|$ and $|df(w)| \geqslant \lambda^{-1}|w|$ for some $\lambda \in (0, 1)$ and $|\ |$ a Riemannian metric on M.

For a compatible handle set the geometric intersection matrix G is defined by letting G_{ik} be the number of components of $f(H_i) \cap H_k$. The following lemma is proved in [3] and relates G to subshifts of finite type. Let $\Lambda_j = \bigcap_{n \in Z} f^n(Y_j)$.

LEMMA. *If $\Lambda_j = \bigcap_{n \in Z} f^n(H)$ then $f|\Lambda_j$ is topologically conjugate to σ: $S_{G_j} \to S_{G_j}$ where G_j is the geometric intersection matrix. Furthermore if G_j is a direct sum of irreducible matrices for each j then f satisfies Axiom* A.

Combining results in [3] and [13] yields the following theorem for subshifts of finite type.

PROPOSITION. *If A is a direct sum of nonnegative irreducible matrices then there exists an integer $k > 0$ and a matrix B such that*

(1) *B is a direct sum of positive matrices,*

(2) *B is congruent* mod 2 *to a diagonal matrix,*

(3) *the subshifts corresponding to A^k and B are topologically conjugate.*

PROOF OF THE THEOREM. Necessity is proven in [3]. To show sufficiency we begin by constructing some local isotopies of sinks and sources in the plane. These isotopies will then be employed on S^2 to produce any dynamics satisfying the hypothesis of the theorem.

Let $D_s = \{x \in \mathbf{R}^2 |\, |x| \leqslant s\}$ and define g_0: $D_1 \to D_{3/2}$ by $g_0((r, \theta)) = (3r/2, \theta)$ and let g_1: $D_{3/2} \to D_1$ be defined by $g_1 = g_0^{-1}$. Then $\Omega(g_0)$ consists of 1 hyperbolic fixed source at the origin and $\Omega(g_1)$ has 1 hyperbolic fixed sink there. Let $C_0 = \{x | 2/3 \leqslant |x| \leqslant 1\}$ and $C_1 = \{x | 1 \leqslant |x| \leqslant 3/2\}$.

ISOTOPING LEMMA. *Let $B = (b_{ij})$ be an $(m + n) \times (m + n)$ positive integer matrix with even nondiagonal entries and m odd diagonal entries. Suppose m_1 and m_2 are nonnegative integers with $m = m_1 + m_2$. Then there exists a diffeomorphism g such that*:

(1) *g is either isotopic to g_0 relative to C_0 or g is isotopic to g_1 relative to C_1.*

(2) *The nonwandering set of g consists of m_1 hyperbolic fixed sources, m_2 hyperbolic fixed sinks, the subshift generated by B, and an additional hyperbolic fixed point* (*which is a source if g_0 is isotoped and a sink if g_1 is isotoped*).

(3) *There is a handle set which contains $\Omega(g)$ and is compatible with g.*

PROOF. We begin with the case in which $m_2 = 0$. An isotopy will be performed on g_0 in two steps. First m_1 sources and m_1 saddles will be added to the dynamics of g_0. Then the subshift will be obtained by weaving the images of handles.

Let s_i be the sector of D_1 defined by

$$s_i = \{(r, \theta) | 0 < r < 2/3 \text{ and } 2\pi(i-1)/(m+n) < \theta < 2\pi i/(m+n)\}.$$

For each i with b_{ii} odd we isotope g_0 inside s_i and relative to a neighborhood of the boundary of s_i. The isotopy creates a hyperbolic fixed saddle at $(\frac{1}{4}, \pi(2i-1)/(m+n))$ and a hyperbolic fixed source at $(\frac{1}{2}, \pi(2i-1)/(m+n))$. The new map is called g_2. We then place a handle H_i over the saddle in s_i. For each i with b_{ii} even, a nilpotent handle H_i is placed in s_i along a circle of radius $\frac{1}{4}$. In Figure 1 the handles and their images are shown for

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 1 \end{bmatrix}.$$

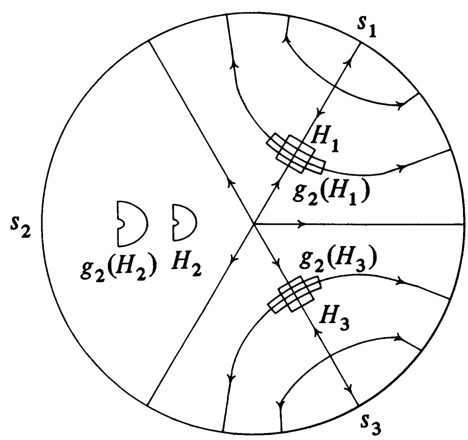

FIGURE 1

Define a matrix $G = (g_{ij})$ by $g_{ii} = 1$ if b_{ii} is odd and $g_{ij} = 2$ otherwise. The next step is begun by isotoping g_2 to a map g_3 such that the geometric intersection matrix of the handle set is the matrix G. To obtain the even diagonal elements the image of each nilpotent handle is isotoped so that it enters the handle on the left side and comes out on the right side (Figure 2). Then the image of each handle is isotoped in a weaving, counterclockwise manner so as to pass through each other handle and create no nonwandering points outside the handles. In Figure 3 the isotopy is shown for the matrix

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

with the handle images thinned to curves.

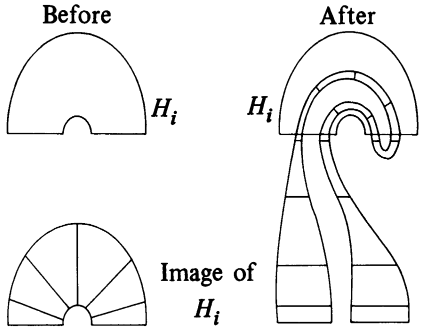

FIGURE 2

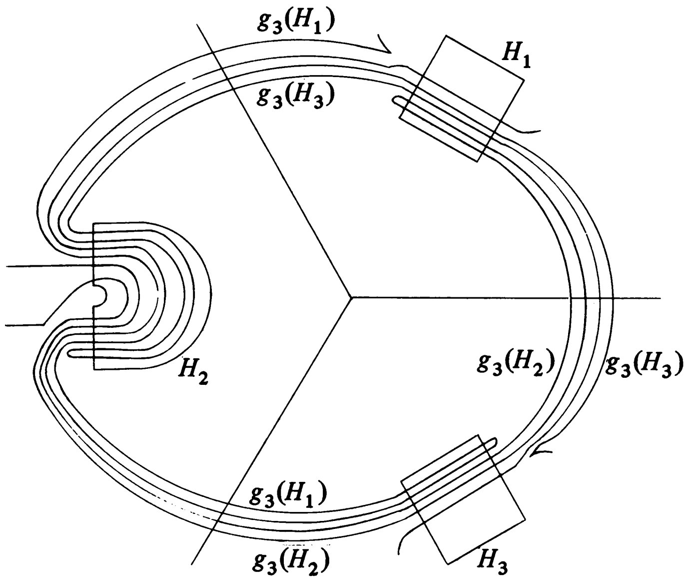

FIGURE 3

For each i and j, $b_{ij} - g_{ij}$ is a nonnegative even integer. To isotope g_3 to a map with geometric intersection matrix B, one performs a sufficient number of finger pushes on the image of H_i near H_j so that the number of components in H_j of the image of H_i is increased by $b_{ij} - g_{ij}$.

With care the isotopies can be performed so that the handle set is compatible with the resulting map g. The inverse of this map is a diffeomorphism from $D_{3/2}$ onto D_1 whose nonwandering set replaces hyperbolic sources with sinks and replaces the subshift corresponding to B by the subshift corresponding to the transpose of B. Thus the lemma is established for the cases in which $m_1 = 0$ or $m_2 = 0$.

Now suppose $m_1, m_2 > 0$, $m = m_1 + m_2$, and $m_1 - m_2 < 2$. We will outline the construction of an isotopy of g_1 to a diffeomorphism g with the desired properties. Once again an inverse argument completes the case for $m_1, m_2 > 0$ and thus the lemma will be proved.

The isotopy is performed in three steps. In the first step the sink at the origin is converted to a hyperbolic orientation preserving fixed source and $m_2 + 1$ hyperbolic fixed sinks and $m_2 + 1$ hyperbolic fixed saddles are added along $r = \frac{1}{2}$. The resulting dynamics are shown in Figure 4. On the annulus C_1 the map remains the same.

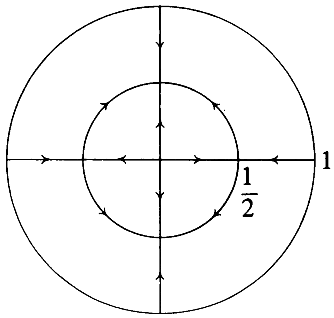

FIGURE 4

The next step is to add $m_1 - 1$ pairs of hyperbolic fixed points with each pair consisting of a saddle and a source. The isotopy creates each new pair along a segment from a sink to the origin with the saddle being placed on the circle $r = \frac{1}{4}$ and the source on the circle $r = \frac{1}{3}$. Handles are then placed over each saddle. Corresponding to each even diagonal entry of A a nilpotent handle is placed away from the fixed points and on the circle $r = \frac{1}{4}$. Finally a weaving isotopy yields the desired subshift. □

The isotopies in the lemma will now be employed to prove the theorem. Suppose A_1 is a direct sum of irreducible matrices and let A_0 and A_2 be permutation matrices. By the proposition of the previous section, let $B = \bigoplus_{i=1}^n B_i$ where each B_i is a positive integer matrix which is congruent mod 2 to a diagonal matrix and furthermore the subshifts corresponding to A_1^k and B are topologically conjugate. We may also assume that k has been chosen so that A_0^k and A_2^k are identity matrices.

For S^2 the condition that $d_2 - d_1 + d_0 = 2$ implies that the sum of the dimensions of A_0^k and A_2^k is two greater than the number of odd diagonal elements of B. Thus there exist two finite sequences of nonnegative integers $\{r_i\}_{i=1}^n$ and $\{t_i\}_{i=1}^n$ with

$$\sum r_i = \left(\dim A_0^k\right) - 1 \quad \text{and} \quad \sum t_i = \left(\dim A_2^k\right) - 1$$

and $r_i + t_i = \dim B_i$ for each i. By applying the previous isotopies to the northpole-southpole map, each B_i can be locally added along with r_i sources and t_i sinks. The result is an Axiom A, no-cycle diffeomorphism with the desired dynamics. A theorem of Dennis Pixton [6] allows the map to be further isotoped (without altering the geometric intersection matrices) so as to satisfy the strong transversality condition.

Let X_l be a connected sum of l projective planes. If a small open disc is removed from X_l, it is then diffeomorphic to a disc with l half-twisted attached strips [5]. There is a Morse-Smale diffeomorphism g_0 of X_l with a fixed sink at the center of the disc, a fixed saddle on each strip, and a fixed source at the center of the removed disc.

Suppose B is a positive integer matrix of reduced degree $m > l$ which is congruent mod 2 to a diagonal matrix. Let $m_1 + m_2 = m - l$ with $m_1, m_2 > 0$. Then we claim that either g_0 or g_0^{-1} can be isotoped, relative to the omitted disc, to a Smale diffeomorphism g whose nonwandering set consists of the subshift represented by B, $m_1 + 1$ hyperbolic fixed sources, and $m_2 + 1$ hyperbolic fixed sinks. This is a generalization of the isotoping lemma.

In Figure 5 the situation is partially illustrated for $m = l = 3$ with

$$B = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}.$$

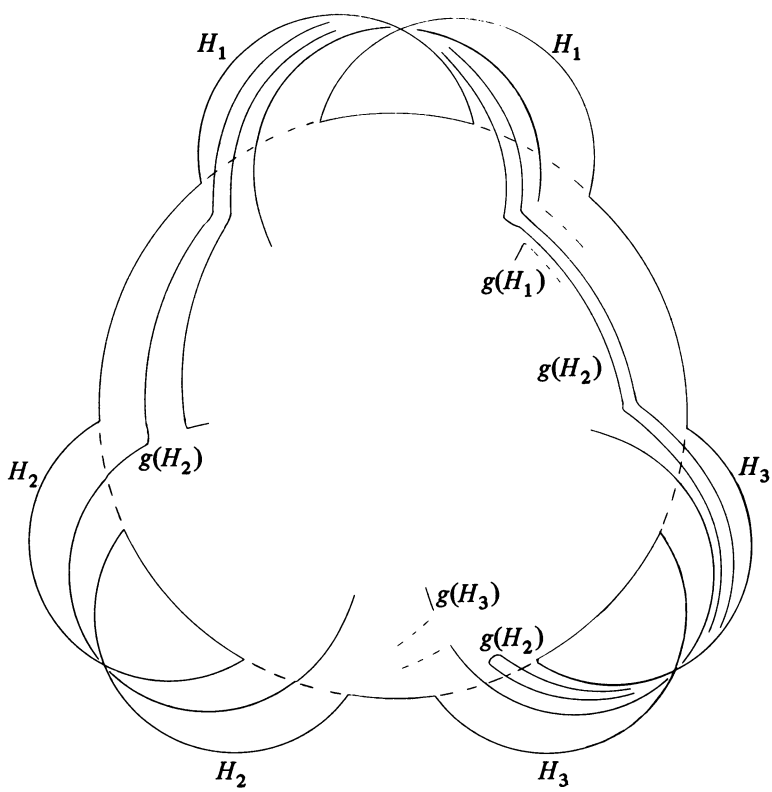

FIGURE 5

Each twisted strip acts as a handle H_i whose image is to be woven through the other strips in a clockwise fashion. Observe that as the image of H_2

emerges from H_1, it is inside the image of H_1. This is a result of the twist. To achieve the entry B_{23}, the image of H_2 is pushed along inside the image of H_1 which, although not shown, must also pass through H_3. After going through H_3, the isotoping of the image of H_2 is complete and the image of H_1 would continue through H_2 with the image of H_3 ahead of it.

For even diagonal elements nilpotent handles are introduced near the boundary of the disc and woven through as before. When $m_1 + m_2 > 0$ a preliminary isotopy is performed adding m_1 sources, m_2 sinks, and $m_1 + m_2$ saddles to the disc. This can be done in a manner similar to the Isotoping Lemma with the saddles occurring on the outside between twisted strips. Images of handles are then woven through the twisted strips, the nilpotent handles, and the outside portions of handles covering the new saddles.

Let h_i and h_j be diffeomorphisms of X_l and $X_{l'}$ which result from isotopies of the above type and yield the dynamics of B_i and B_j. From these maps and Pixton's result a Smale diffeomorphism G on $X_{l+l'}$ can be obtained by removing a neighborhood of a sink in X_l and a neighborhood of a source in $X_{l'}$, connecting the spaces and maps carefully, and isotoping. The proof of the theorem for a connected sum of projective planes is completed by producing a map h_i for each B_i and then connecting.

When a small disc is removed from a connected sum of l tori, the space obtained is diffeomorphic to a disc attached with l crossed pairs of strips [**5**]. To prove the theorem for a connected sum of tori, similar isotopies are performed on these spaces. $\square$

PROOF OF COROLLARY 2. By the Isotoping Lemma produce a diffeomorphism g: $D_{3/2} \to D_1$ whose nonwandering set contains a fixed sink at the origin and a power of the desired subshift. Suspending this map and rounding off the corners as in [**4**] yields a flow on the solid torus with everything on the boundary moving inward.

To obtain the structurally stable flow on S^3 attach another solid torus with one closed hyperbolic repellor orbit. The flow on R^3 is gotten by extending the flow on the first solid torus. This can be done by putting a rest point at the center of the torus and having everything flow in from infinity.

REFERENCES

1. R. Bowen, *Topological entropy and Axiom* A, Proc. Sympos. Pure Math., vol. 14, Amer. Math. Soc., Providence, R.I., 1970, pp. 23–41.

2. R. Bowen and O. Landord, *Zeta functions of restrictions of the shift transformation*, Proc. Sympos. Pure Math., vol. 14, Amer. Math. Soc., Providence, R.I., 1970, pp. 43–49.

3. J. Franks, *Constructing structurally stable diffeomorphisms*, Ann. of Math. **105** (1977), 343–359.

4. ______, *Non-singular flows on* S^3 *with hyperbolic chain-recurrent set*, Rocky Mountain J. Math. 7 (1977), 539–546.

5. F. Massey, *Algebraic topology*, Harcourt, Brace and World, New York, 1967.

6. D. Pixton, *Tameness and stability* (to appear).

7. J. Robbin, *A structural stability theorem*, Ann. of Math. **94** (1971), 447–493.

8. R. C. Robinson, *Structural stability of* C' *diffeomorphisms*, J. Differential Equations **22** (1976), 28–73.

9. M. Shub, *Structurally stable diffeomorphisms are dense*, Bull. Amer. Math. Soc. **78** (1972), 817–818.

10. M. Shub and D. Sullivan, *Homology theory and dynamical systems*, Topology **14** (1975), 109–132.

11. S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.

12. ______, *Stability and isotopy in discrete dynamical systems*, Proc. Sympos. on Dynamical Systems, Salvador, Brazil, Academic Press, New York, 1971.

13. R. F. Williams, *Classification of subshifts of finite type*, Ann. of Math. **98** (1973), 120–153.

DEPARTMENT OF MATHEMATICS, EMORY UNIVERSITY, ATLANTA, GEORGIA 30322