Journal of Statistical Physics, Vol. 40, Nos. 1/2, 1985

On Constructing Markov Partitions by Computer

Valter Franceschini[1] **and Fernando Zironi**[1]

Received December 4, 1984

Two methods to construct Markov partitions for two-dimensional systems are proposed. One is based on the existence of a known, or easily accessible by numerical analysis, hyperbolic fixed point; the other one, which is more general, is derived from Bowen's proof of the existence theorem of Markov partitions for hyperbolic systems. The methods are successfully implemented in two cases of hyperbolic systems: the linear automorphism $\left(\begin{smallmatrix}1 & 1\\ 1 & 2\end{smallmatrix}\right)$ of the 2-torus and a nonlinear perturbation of it. The methods are applied also to the Hénon mapping. In such nonhyperbolic case, however, they produce partitions of the Hénon attractor which lack some essential properties.

KEY WORDS: Symbolic dynamics; Markov partitions; hyperbolic systems; fixed point; stable and unstable manifolds; Hénon map.

1. INTRODUCTION

One approach to the study of a dynamical system is through symbolic dynamics.[1,2] In such context the key notion is that of Markov partitions. Bowen[3] showed that Axiom A diffeomorphisms admit Markov partitions of diameter less than any given δ. Bowen's constructive proof generalizes and at the same time simplifies a previous proof by Sinai.[4,5] The idea of Markov partitions as a way of codifying large classes of measures by a symbolic language was introduced by Adler and Weiss[6] and then extended by Sinai.[7]

The purpose of this paper is to describe two methods for the construction by computer of Markov partitions for two-dimensional systems. The first method, which is very simple, applies to transformations endowed with an *a priori* known hyperbolic fixed point and consists essentially in tracing

[1] Dipartimento di Matematica Pura ed Applicata, Universita' di Modena, via Campi 213/B, 41100 Modena, Italy.

two pieces of the stable and unstable manifolds through this point. The other method is inspired by the proof of Bowen's theorem and, more specifically, by its simplified version given in Ref. 2 in the two-dimensional case. Starting from a conveniently chosen initial covering, first one makes it acquire some desired properties by a convergent process of subsequent approximations, then a Markov partition can be easily derived. Both methods are proven to be rather efficient through a number of successful experiments on two different hyperbolic systems: the linear automorphism $\left(\begin{smallmatrix}1 & 1\\ 1 & 2\end{smallmatrix}\right)$ of the 2-torus and a nonlinear perturbation of it. Furthermore, the methods are applied in the case of a nonhyperbolic system: the Hénon mapping.[8] They still work, but produce partitions of the Hénon attractor which fail to have some important properties, as it can be expected from the fact that the Hénon mapping should not admit a finite Markov partition (see Collet and Levy[9]).

2. HYPERBOLIC SYSTEMS AND MARKOV PARTITIONS

In this section we shall present a minimal number of basic notions that we consider essential to the understanding of the applications described in the next sections. The exposition, based on Ref. 2 and sometimes on Ref. 3, will be made just having in mind the applications.

Definition 1. Let (Ω, S) be a dynamical system on some compact Riemanian manifold Ω of class C^∞ with S a diffeomorphism of class C^∞. (Ω, S) is called a *hyperbolic system* if

(a) There exist a constant $\gamma > 0$ and, for every $x \in \Omega$, two manifolds $W^s(x)$ and $W^u(x)$, which are of class C^∞ in a neighborhood of the ball $B_\gamma(x)$ (center x and radius γ) and with tangent plane at x depending on x with Hölderian regularity. Furthermore, the manifolds $W^s(x)$ and $W^u(x)$ are transversal at x and have dimensions which are complementary and positive.

(b) If $W^s_\gamma(x)$ is the connected part of $W^s(x) \cap B_\gamma(x)$ containing x and $W^u_\gamma(x)$ is the connected part of $W^u(x) \cap B_\gamma(x)$ containing x, then

$$S(W^s_\gamma(x)) \subset W^s_\gamma(Sx), \qquad S^{-1} W^u_\gamma(x) \subset W^u_\gamma(S^{-1}x)$$

(c) There exists a constant $\lambda < 1$ such that, for any $n \geqslant 0$,

$$d(S^n y, S^n z) \leqslant \lambda^n d(y, z), \qquad \forall y, z \in W^s_\gamma(x) \tag{2.1}$$

and

$$d(S^{-n} y, S^{-n} z) \leqslant \lambda^n d(y, z), \qquad \forall y, z \in W^u_\gamma(x) \tag{2.2}$$

where d is the metric on Ω.

(d) There exists a constant β, $0<\beta<\gamma$, such that, if $x, y \in \Omega$ and $d(x, y)<\beta$, the set $W_\gamma^s(x) \cap W_\gamma^u(y)$ consists of exactly one point $[x, y]$ which depends continuously on x and y.

$W^s(x)$ and $W^u(x)$ are called the *stable* and *unstable manifold* of x, respectively. Under the action of S, two points of the stable manifold approach each other with exponential rate [property (2.1)], while two points of the unstable manifold diverge with exponential rate [property (2.2)]. For this reason they are sometimes also referred to as *contracting* and *expanding* manifold.

Definition 1 contains some redundant elements and is not as general as possible. For a more general definition of hyperbolic system the reader is referred to Ruelle.[10]

Definition 2. A nonempty set $R \subset \Omega$ is an S *rectangle* for the hyperbolic system (Ω, S) if

(a) $R=\overline{\operatorname{int}(R)}$ (R is the closure of its interior)

(b) $\operatorname{diam}(R)<\beta$

(c) $x, y \in R \Rightarrow [x, y] \in R$

Let

$$\partial^s R=\{x \in R\colon x \notin \text{interior of } (W_\gamma^u(x) \cap R) \text{ in } W_\gamma^u(x)\}$$

$$\partial^u R=\{x \in R\colon x \notin \text{interior of } (W_\gamma^s(x) \cap R) \text{ in } W_\gamma^s(x)\}$$

Lemma 1. The boundary ∂R of R is $\partial^s R \cup \partial^u R$.

Definition 3. A finite covering $\mathscr{R}=(R_1, R_2,..., R_N)$ of Ω by S rectangles is a *rectangle partition* of (Ω, S) if $R_i \cap R_j=\partial R_i \cap \partial R_j$, $i \neq j$.

Definition 4. A rectangle partition $\mathscr{R}=(R_1, R_2,..., R_N)$ of (Ω, S) is a *Markov partition* if

$$\text{(a)} \quad S\left(\bigcup_{i=1}^N \partial^s R_i\right) \subset \bigcup_{i=1}^N \partial^s R_i, \qquad S^{-1}\left(\bigcup_{i=1}^N \partial^u R_i\right) \subset \bigcup_{i=1}^N \partial^u R_i \tag{2.3}$$

$$\text{(b)} \quad \text{for } i, j=1,..., N,\ R_i \cap S R_j \text{ is connected} \tag{2.4}$$

Theorem 1. Given a hyperbolic system (Ω, S) and a constant $\delta>0$, there exists a Markov partition of Ω by S rectangles of diameter less than δ.

Lemma 2. If $\mathscr{Q}=(Q_1, Q_2,..., Q_N)$ is a Markov partition for the hyperbolic system (Ω, S^M), $M \geqslant 2$, then

$$\mathscr{R}=\left\{R\colon R=\bigcap_{k=0}^{M-1} S^k Q_k,\ Q_k \in \mathscr{Q}\right\} \tag{2.5}$$

is a Markov partition for (Ω, S).

Definition 5. Given a rectangle partition $\mathscr{R}=(R_1, R_2,..., R_N)$ of (Ω, S), its *transition matrix* is the $(N \times N)$ matrix T of elements

$$t_{ij}=\begin{cases}0 & \text{if } \operatorname{int}(R_i) \cap \operatorname{int}(S^{-1}R_j)=\Phi \\ 1 & \text{if } \operatorname{int}(R_i) \cap \operatorname{int}(S^{-1}R_j) \neq \Phi\end{cases}$$

Definition 6. If $\mathscr{R}=(R_1, R_2,..., R_N)$ is a rectangle partition of (Ω, S), an $(\mathscr{R}, S)$ *history* of a point $x \in \Omega$ is a doubly infinite sequence $\boldsymbol{\sigma} \in \{1, 2,..., N\}^Z$ such that

$$x \in S^{-j}R_{\sigma_j}, \qquad \forall j \in Z$$

Definition 7. If $\mathscr{R}=(R_1, R_2,..., R_N)$ is a rectangle partition of (Ω, S) with transition matrix T, a sequence $\boldsymbol{\sigma} \in \{1, 2,..., N\}^Z$ is *T compatible* if

$$\prod_{j=-\infty}^{+\infty} t_{\sigma_j \sigma_{j+1}}=1$$

Remarks. Let $\mathscr{R}=(R_1, R_2,..., R_N)$ be a Markov partition of a hyperbolic system (Ω, S) with transition matrix T. If Σ is the set of all the T-compatible sequences, for every $\boldsymbol{\sigma} \in \Sigma$ the set

$$X(\boldsymbol{\sigma})=\bigcap_{k=-\infty}^{+\infty} S^{-k}R_{\sigma k}$$

is nonempty and consists of a single point x. Then, every $\boldsymbol{\sigma} \in \Sigma$ is the $(\mathscr{R}, S)$ history of one point $x \in \Omega$. On the other hand, for every $x \in \Omega$ there exists at least one $\boldsymbol{\sigma} \in \Sigma$ such that $X(\boldsymbol{\sigma})=\{x\}$. If $x \in \Omega / \bigcup_{k=-\infty}^{+\infty} S^{-k}\partial$, $\partial = \bigcup_{i=1}^{N} \partial R_i$, $\boldsymbol{\sigma}$ is unique, while if $x \in \bigcup_{k=-\infty}^{+\infty} S^{-k}\partial$ there are more than one $\boldsymbol{\sigma}$ (at most some n_0). Hence, if one is interested in measures which give measure zero to ∂, as it is often the case, the correspondence between Σ and Ω is almost everywhere one-to-one. Then, roughly speaking, the sequences of symbols of Σ reproduce the dynamics generated by S on Ω. For this reason the Markov partition is said to be *generating*. The transformation $X\colon \Sigma \to \Omega$ and its "inverse" transformation are called the *codes of the symbolic dynamics of S with respect to* $\mathscr{R}$.

About condition (2.4) of Definition 4, it implies that, if $t_{ij}=1$, $S^{-1}R_j$ crosses R_i only once (as well as SR_i does with respect to R_j). This guarantees the fact that $X(\boldsymbol{\sigma})$ consists of exactly one point. If this is not the case, the partition is nongenerating.

Suppose now that the dimension of Ω is two. Furthermore, we can consider partitions which consist only of connected S rectangles without losing generality. Every such rectangle R can be represented as a pair of "segments," one of the stable manifold, the other one of the unstable manifold. More precisely, if $x \in \operatorname{int}(R)$, and

$$C = W_\gamma^u(x) \cap R, \qquad D = W_\gamma^s(x) \cap R$$

it follows

$$R = [C, D] = \bigcup_{\substack{y \in C \\ z \in D}} [y, z] \tag{2.6}$$

The boundary ∂R is composed of four connected sides, two of the stable manifold, $\partial_k^s R$, and two of the unstable manifold, $\partial_k^u R$, with $k = 1, 2$. If we consider the boundaries ∂C and ∂D of the segments C and D,

$$\partial C = \{y \in C, y \notin \operatorname{int}(C)\}, \qquad \partial D = \{z \in D, z \notin \operatorname{int}(D)\}$$

both of them consist of two points, $\partial_k C$ and $\partial_k D$, $k = 1, 2$, respectively. So, we can define

$$\partial_k^s R = [\partial_k C, D], \qquad \partial_k^u R = [C, \partial_k D]$$

In the following we shall often refer to the $\partial_k^s R$ ($\partial_k^u R$), $k = 1, 2$, as the stable (unstable) sides of R.

Finally, relations (2.3) are equivalent to the following:

Property 1. For any $i \in \{1, 2, \ldots, N\}$ and for any $k \in \{1, 2\}$ there exist $i', i'' \in \{1, 2, \ldots, N\}$ and $k', k'' \in \{1, 2\}$ such that

$$S\, \partial_k^s R_i \subset \partial_{k'}^s R_{i'}, \qquad S^{-1}\, \partial_k^u R_i \subset \partial_{k''}^u R_{i''} \tag{2.7}$$

3. TWO METHODS TO CONSTRUCT MARKOV PARTITIONS OF TWO-DIMENSIONAL SYSTEMS

3.1. First Method

It applies only in the case when the transformation S has an easily accessible (from the numerical viewpoint) hyperbolic fixed point whose stable and unstable manifolds separately cover Ω densely. Let z^* be such a point and W^s (W^u) its stable (unstable) manifold. If $x \in W^s$ ($y \in W^u$), let W^s_x (W^u_y) represent the segment of W^s (W^u) included between z^* and x (y).

If x_1 and x_2 are points of W^s symmetrically placed with respect to z^*, then either

$$S(W^s_{x_1}) \subset W^s_{x_1}, \qquad S(W^s_{x_2}) \subset W^s_{x_2} \tag{3.1}$$

or

$$S(W^s_{x_1}) \subset W^s_{x_2}, \qquad S(W^s_{x_2}) \subset W^s_{x_1} \tag{3.2}$$

holds.

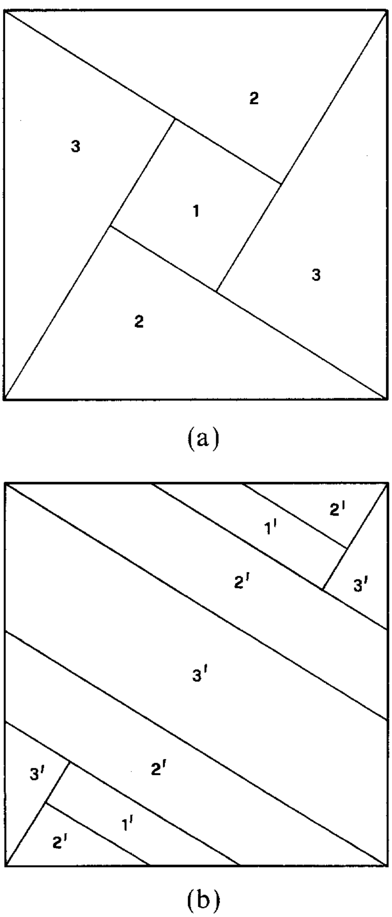

Fig. 1. (a) The partition $\mathscr{R} = \{R_1, R_2, R_3\}$ of (Ω, S) proposed by Arnold and Avez in Ref. 11; (b) image of $\mathscr{R}$ under S^{-1}. The comparison of part (a) with part (b) provides the transition matrix T: $t_{11} = t_{12} = t_{23} = t_{31} = 0$, $t_{13} = t_{21} = t_{22} = t_{32} = t_{33} = 1$.

Analogously, if y_1 and y_2 belong to W^u and are symmetrically placed with respect to z^*, then either

$$S^{-1}(W^u_{y_1}) \subset W^u_{y_1}, \qquad S^{-1}(W^u_{y_2}) \subset W^u_{y_2} \tag{3.3}$$

or

$$S^{-1}(W^u_{y_1}) \subset W^u_{y_2}, \qquad S^{-1}(W^u_{y_2}) \subset W^u_{y_1} \tag{3.4}$$

holds.

The idea for the construction of a Markov partition is implicit in these considerations. Tracing from z^* the two branches of W^s and the two branches of W^u and making each branch of W^s end onto W^u and, conversely, each branch of W^u end onto W^s, if each branch is roughly as long as the symmetric one, a rectangle partition is obtained which satisfies property (2.3). If the partition is sufficiently fine, also property (2.4) holds and it is a Markov partition. For instance, the 3-rectangle partition given by Arnold and Avez[11] for the automorphism $\binom{1\ 1}{1\ 2}$ of the 2-torus can be obtained in this way (Fig. 1a). It is not a Markov partition because its S rectangles are too large and cause two intersections $R_i \cap S^{-1}R_j$ to be disconnected (compare Fig. 1b with Fig. 1a). However, if one prolongs the four branches of the manifolds up to the next intersection, the partition becomes a Markov partition (Fig. 2).

Concerning the method as it has been proposed, we note that for most transformations the tracing from z^* of all the four branches of the manifolds is not strictly necessary. As a matter of fact, we can omit one branch of W^s if property (3.1) holds or, alternatively, one branch of W^u if (3.3) is verified.

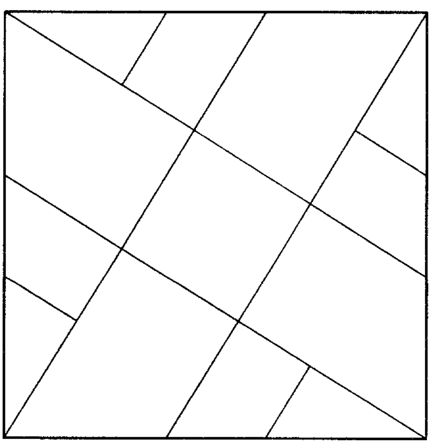

Fig. 2. Markov partition of (Ω, S) by seven rectangles.

3.2. Second Method

The proof of Theorem 1, due to Bowen,[3] is a simple constructive proof. Referring to it as exposed by Gallavotti in the two-dimensional case,[2] we can derive an algorithm for the construction of a Markov partition for any two-dimensional hyperbolic system. Such an algorithm can be summarized in four steps:

(1) Choose a suitable rectangle covering $\mathscr{A}$ of Ω.
(2) Modify $\mathscr{A}$ into $\mathscr{A}^*$ through a convergent process of successive approximations.
(3) Derive a rectangle partition $\mathscr{R}$ from the limit covering $\mathscr{A}^*$.
(4) Check whether $\mathscr{R}$ satisfies property (2.3) and (2.4).

After remarking that the algorithm may fail in producing a Markov partition only in the case of an initial covering which is not sufficiently fine, let us try to define the four steps with only a few details.

Step 1. Consider a rectangle covering $\mathscr{A}=(A_1, A_2,..., A_L)$ of Ω with the property that every point $x\in\Omega$ is interior to at least one A_j. As a consequence of this condition there exists a positive constant a such that, for every $x\in\Omega$, there is a rectangle $A_{j(x)}$ containing x and all the segments $W_a^u(y)$ and $W_a^s(z)$ of length $2a$, which are obtained by considering all the points $y\in W_\gamma^s(x)\cap A_{j(x)}$ and all the points $z\in W_\gamma^u(x)\cap A_{j(x)}$.

Step 2. Let $\mathscr{A}^0=\{A_j=A_j^0=[C_j^0, D_j^0]\}_{j=1,...,L}$. For every j, consider SC_j^0. As a consequence of the expansion, SC_j^0 is a "long" line. Find a covering of SC_j^0 which consists of $k_C(j)$ elements of $\mathscr{A}^0$, say, $A_{i_1}^0$, $A_{i_2}^0$,..., $A_{i_{k_C(j)}}^0$.Choose the $A_{i_h}^0$'s in such a way that the distance between SC_j^0 and the "parallel" unstable sides $\partial_l^u A_{i_h}^0$, $l=1, 2$, $h=1,..., k_C(j)$, is not less than a. Let the two end points of SC_j^0, $S\,\partial_1 C_j^0$ and $S\,\partial_2 C_j^0$, be covered only by $A_{i_1}^0$ and $A_{i_2}^0$, respectively. Generally, $SC_{i_l}^0$, $l=1, 2$, does not cross $A_{i_l}^0$ completely, therefore $S\,\partial_l C_j^0$ has a positive distance d_l from the stable side of $A_{i_l}^0$ that does not cross SC_j^0.

Lengthen C_j^0 so that also $A_{i_1}^0$ and $A_{i_2}^0$ are completely crossed, which means that we substitute C_j^0 by

$$C_j^1=\bigcup_{h=1}^{k_C(j)} S^{-1}([C_{i_h}^0, SC_j^0\cap A_{i_h}^0])$$

Analogously, after finding a covering $A_{i_1},..., A_{i_{k_D(j)}}$ for $S^{-1}D_j^0$,. $j=1,..., L$, replace D_j^0 by

$$D_j^1=\bigcup_{h=1}^{k_D(j)} S([S^{-1}D_j^0\cap A_{i_h}^0, D_{i_h}^0])$$

So, we can substitute the initial covering $\mathscr{A}^0$ by the covering $\mathscr{A}^1$, defined to be

$$\mathscr{A}^1 \equiv \{A_j^1 = [C_j^1, D_j^1]\}_{j=1,\dots,L}$$

The process can be iterated to produce a sequence of coverings $\mathscr{A}^n$ which may converge to a covering $\mathscr{A}^*$. The convergence is guaranteed if "$\mathrm{diam}(A_j)/\lambda$ is sufficiently smaller than a" (see Ref. 2). However, this condition is usually not necessary. In fact, we shall see later that in all the cases we consider even an initial covering by very few rectangles gives rise, independently of the value of a, to a convergent process. In this case the convergence rate is geometric by a factor λ.

Two remarks have to be made. One regards the choice of the extreme elements $A_{i_k}^n$ that determine the prolongation of each side for all the rectangles of the coverings $\mathscr{A}^n$, $n = 0, 1,\dots$. A standard choice corresponds to minimizing the distances d_l. A different choice, however, can be made at the first iterations in order to produce a Markov partition which is "better," e.g., by fewer or more balanced rectangles. The second remark concerns the substitution of C_j^n by C_j^{n+1} and D_j^n by D_j^{n+1}. There are two possible ways of operating. One consists in making the substitutions after the computation of all the C_j^{n+1}'s and the D_j^{n+1}'s, the other one consists in replacing each C_j^n (D_j^n) just after computing C_j^{n+1} (D_j^{n+1}). Usually the correctness of the final result is not invalidated by the way we operate, even if different limit coverings can be obtained.

Step 3. Consider the finest partition of Ω which is obtained by intersecting the elements of $\mathscr{A}^*$ in all the possible ways. In doing so we obtain a partition $\mathscr{R}^*$, which satisfies relations (2.7), and then relations (2.3), but $\mathscr{R}^*$ is not necessarily made only by S rectangles. It may happen, in fact, that in some vertex both stable and unstable manifolds are interrupted, causing property (c) of an S rectangle to fail. In such a case we need a further operation in order to obtain a rectangle partition $\mathscr{R}$ from $\mathscr{R}^*$. For each "anomalous" vertex we must extend at least one side by continuing it along the stable (unstable) manifold, until it meets the nearest unstable (stable) side of some other rectangle. By considering all the possible intersections of the elements of the covering $\mathscr{R}^*$ after such a prolongation of their sides, we obtain a rectangle partition $\mathscr{R}$.

Step 4. The check of condition (2.3), which is requested only in the case that some side has been extended, and that of condition (2.4) do not present any problem if a videographic device is available. Concerning the former check, it must be also remarked that it would be made unnecessary by the condition "$\mathrm{diam}(A_j)/\lambda$ sufficiently smaller than a." This condition, in fact, guarantees the preservation of property (2.3) during the operation of prolonging the sides, just as it was a sufficient condition for the convergence of the iterative procedure.

We remark again that if either check of Step 4 fails, the whole process has to be repeated with a finer initial covering.

Lemma 1 suggests an alternative use of the two methods that we have just described. Namely, we can first construct a Markov partition for the system (Ω, S^M), $M \geqslant 2$, then obtain one for (Ω, S) through the refinement (2.5). The motivations for considering S^M rather than S are different for the two methods. Concerning the first method, the most reasonable motivation could be the fact that no unstable fixed point exists, while there exists an unstable cycle of period M (with M, hopefully, not too large). The second method, instead, may fail on (Ω, S) owing to the fact that, even with a very fine initial covering, $\mathrm{diam}(A_j)/\lambda$ is not sufficiently smaller than a, which makes it necessary to work with S^M. A further reason for using S^M could be the gain in the convergence rate of the iterative procedure: the geometric ratio becomes in fact λ^M.

4. COMPUTER EXPERIMENTS

This section will be devoted to illustrating some experiments made by using the methods previously discussed. Our object is to demonstrate that both methods work and can be applied successfully in different situations. Considering that the second method is more general and more difficult to be implemented on a computer, we shall spend some space to describe the computational details that concern it. We stress however the fact that no attention will be paid to those details that appear to be only technicalities of programming solvable in several different ways.

To begin with, we consider a simple hyperbolic system (Ω, S), whose importance is emphasized by Arnold and Avez in their book.[11] Ω is the two-dimensional torus $T^2 = [0, 2\pi] \times [0, 2\pi]$ and S is the linear automorphism defined by the matrix, also labeled S, $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right)$. A point $x = (x_1, x_2) \in \Omega$ is transformed by S into the point $x' = (x_1 + x_2, x_1 + 2x_2) \in \Omega$ by identification of the coordinates that differ by 2π. The stable manifold $W^s(x)$ and the unstable one $W^u(x)$ through the point x are the two straight lines, densely covering Ω, with the direction of the eigenvectors of the matrix S. $W^s(x)$ is individuated by the vector $\mathbf{v}_1$ associated with the eigenvalue $\lambda_1 = \lambda < 1$, and $W^u(x)$ is individuated by the vector $\mathbf{v}_2$ associated with $\lambda_2 = 1/\lambda > 1$, where

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ \lambda - 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 1 \\ \dfrac{1}{\lambda} - 1 \end{pmatrix}, \qquad \lambda = (3 - \sqrt{5})/2$$

The transformation S, which is area preserving, has an unstable fixed point at the origin.

The construction of a Markov partition for (Ω, S) is considerably facilitated by the fact that the manifolds are straight lines and, moreover, by the fact that an S rectangle is a rectangle in the usual meaning. Adler and Weiss[12] gave first a very simple example of Markov partition for (Ω, S) and any other ergodic "linear automorphism" of the 2-torus. Here we have already shown a Markov partition for (Ω, S) (Fig. 2), which was obtained by using the first method. Obviously, a partition as fine as desired can be easily constructed by tracing conveniently long branches of manifolds through the origin.

The application of the second method to (Ω, S) represents a useful way to become familiar with the method. A preliminary problem to be solved regards the representation of an S rectangle inside the computer. Let (x_1, x_2) be the coordinates of a generic point $P \in \Omega$ and $\{R_1,..., R_N\}$ the set of S rectangles we are concerned with. We shall represent R_i through a pair of two opposite vertices: for instance, the bottom vertex V_i^b and the topmost one V_i^t, V_i^b being the vertex from which both sides develop with increasing x_2. The representation $R_i = (V_i^b, V_i^t)$ is very convenient and allows to pass easily to a form like (2.6).

The choice of the initial covering $\mathscr{A}^0$ of the torus can be made quite arbitrarily. We shall explain an easy technique to obtain initial coverings by S rectangles of diameter as small as one likes. Consider the segment AB of the unstable manifold which divides the square $[0, 2\pi] \times [0, 2\pi]$ into two equal parts. Trace the four segments of the stable manifold that connect the middle point of each of the four sides of the square with AB. Finally, extend AB in both directions until it meets one segment of the stable manifold. This way we obtain a partition $\mathscr{B}_0$ of Ω into three S rectangles (Fig. 3). By lengthening each side of each rectangle in both directions by an arbitrarily chosen quantity a, a covering $\mathscr{B}_0^*$ (see again Fig. 3) is produced with the desired property that every point $x \in \Omega$ is internal to at least one rectangle at a distance not less than a from its boundary. If we partition each rectangle of $\mathscr{B}_0$ into two equal parts by halving the two longer sides, we obtain a partition $\mathscr{B}_1$ by six S rectangles. By further partitioning this way, we can derive a partition $\mathscr{B}_i$, and then a covering $\mathscr{B}_i^*$, by $3 \cdot 2^i$ S rectangles. Most of our experiments were made by assuming some $\mathscr{B}_i^*$ as initial covering.

As far as steps 2 and 3 of the algorithm are concerned, their implementation involves some technical problems which derive mainly from the fact that we are dealing with a torus. They can however be overcome without too much effort, and a very efficient program for computing and drawing Markov partitions of (Ω, S) can be carried out. The iterative process is

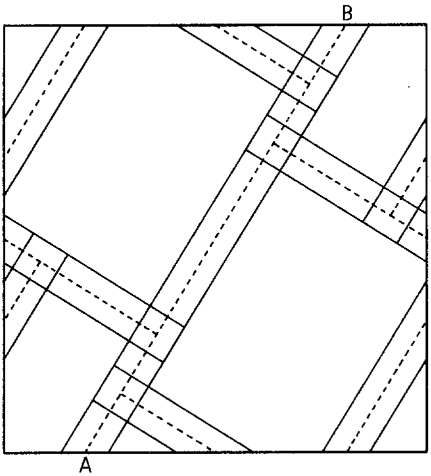

Fig. 3. The partition $\mathscr{B}_0$ (dashed line) and the rectangle covering $\mathscr{B}_0^*$ corresponding to $a = 0.3$ (solid line). $\mathscr{B}_0$ includes the segment AB.

stopped when all the prolongations of the sides become less than a prescribed tolerance α. The check of property (2.4) is performed by separately drawing the image of each rectangle of the final partition on the partition itself.

We computed several Markov partitions, in particular the ones associated with the initial coverings $\mathscr{B}_0^*$, $\mathscr{B}_1^*$, and $\mathscr{B}_2^*$ with $a = 0.01$, $a = 0.1$, and $a = 0.4$. In none of these cases the operation of prolonging the sides after the convergence was needed. On the contrary, the prolongation was necessary in a few cases of initial coverings which, having been chosen in some random way, did not exhibit any symmetry. This supports the statement that, the more randomly the initial covering is taken, the more the presence of anomalous vertices in the limit covering is likely to occur, so making the prolongation of some sides necessary.

Using a tolerance $\alpha = 10^{-10}$, the number of iterations necessary to reach the limit covering was nearly always 25, consistent with the fact that the convergence is geometric with ratio λ. The check of relations (2.3) showed that they are always satisfied with an accuracy of the same order of α. Also the transition matrix T and their eigenvalues were computed. In all cases the largest eigenvalue of T, whose logarithm provides the topological entropy of the dynamical system (Ω, S), which in this case is the same as the maximum metric entropy, was equal to $1/\lambda$, as it must be (see, for instance, Refs. 1 and 6). Figure 4 shows the three Markov partitions that correspond to $\mathscr{B}_0^*$, $\mathscr{B}_1^*$, and $\mathscr{B}_2^*$ with $a = 0.1$ and consist of 7, 15, and 35 S rectangles, respectively.

Still using the second method, we made some experiments also for S^2. In this case the convergence was obviously faster by a factor $1/\lambda$, but the

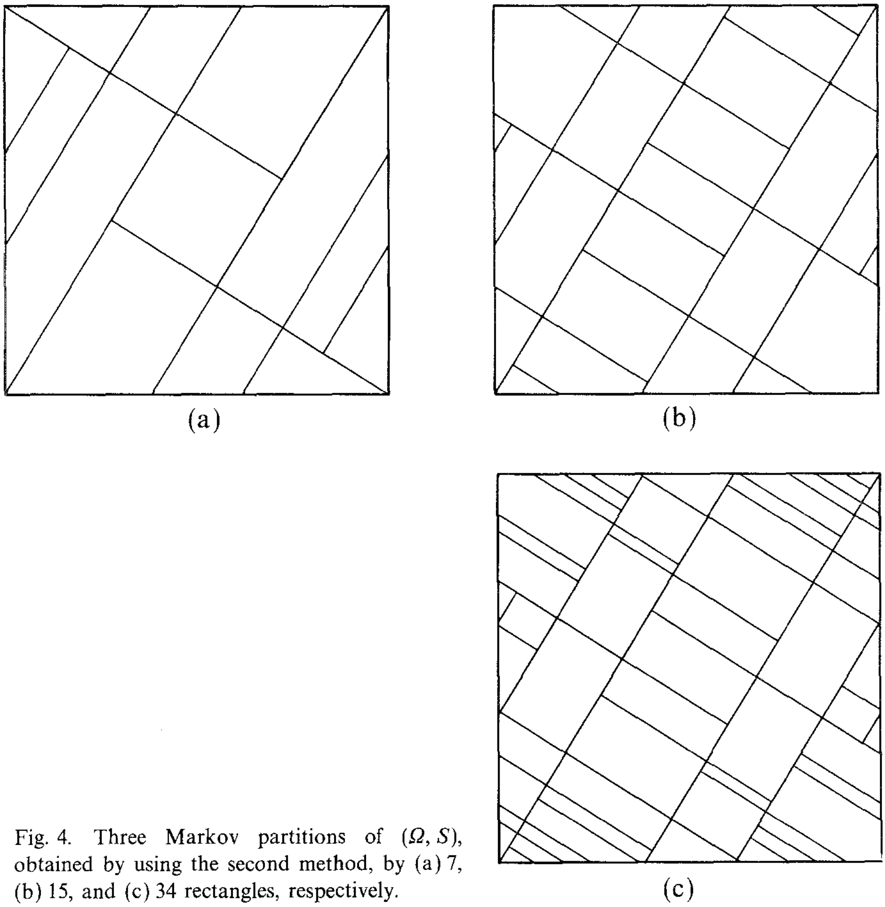

Fig. 4. Three Markov partitions of (Ω, S), obtained by using the second method, by (a) 7, (b) 15, and (c) 34 rectangles, respectively.

inconvenience of attaining a nongenerating Markov partition could occur, particularly if one assumed an initial covering of very few rectangles. Figure 5 represents a nongenerating partition that can be obtained by starting from a 3-rectangle covering. The largest eigenvalue of its transition matrix T is less than $1/\lambda^2$, which in this case means that at least one rectangle is crossed more than once by the image of some rectangle, as it is shown in the picture. A generating Markov partition of (Ω, S^2), derived from an initial covering by six rectangles, is displayed in Fig. 6. We note that now the largest eigenvalue of T is exactly $1/\lambda^2$.

The partitions obtained by working on S^2 can be used to construct Markov partitions for (Ω, S) through the refinement (2.5). A remark, which holds in general, can be made about the two partitions of (Ω, S) obtainable from the ones of Figs. 5 and 6. Without actually constructing

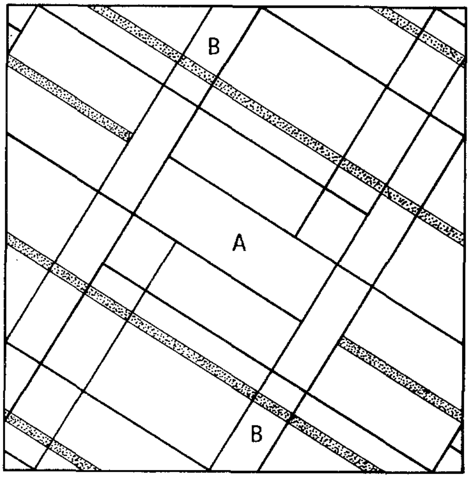

Fig. 5. Nongenerating Markov partition of (Ω, S^2) by 22 rectangles. The hatched rectangle, which is thin and long, is the inverse image of the rectangle A. It crosses the rectangle B twice.

them, we can count the number of ones in the transition matrices relative to the two partitions for S^2. This way we know that the number of the S rectangles composing the Markov partitions for S is larger than 151 in the former case and exactly 232 in the latter. Therefore, the number of elements of the resulting Markov partitions becomes much larger when one changes from S to S^2. This fact might represent a serious problem if one wants to work on S^M with M even slightly larger than two.

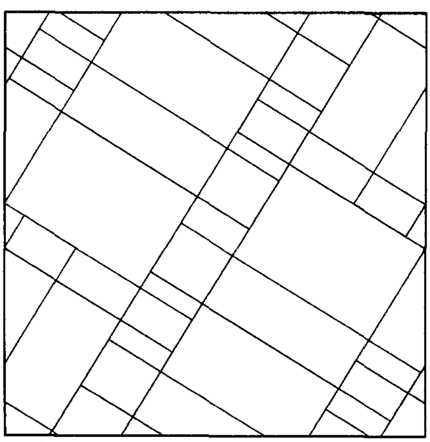

Fig. 6. Markov partition of (Ω, S^2) by 36 rectangles.

Consider now the experiments relative to a nonlinear area-preserving mapping S_ε obtained by perturbing the linear automorphism S,

$$S_\varepsilon \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 + \varepsilon \cos x_1 \\ x_1 + 2x_2 + 2\varepsilon \cos x_1 \end{pmatrix} \qquad \mod 2\pi$$

with $\varepsilon > 0$. This mapping, which is invertible for all ε and hyperbolic at least for small ε, has fixed points with both coordinates equal to ξ, where ξ is any solution of the equation $\xi + \varepsilon \cos \xi = 0$. One fixed point P_0 exists for any ε, which tends to the origin as ε tends to zero and is always unstable. Pairs of twin fixed points (P_k, P_k^*), stable the former, unstable the latter, arise for $\varepsilon = \varepsilon_k$, ε_k being the solution of the equation

$$(\varepsilon^2 - 1)^{1/2} + \arcsin(1/\varepsilon) = k\pi$$

for $k = 1, 2, \ldots$. Around P_1, which appears at $\varepsilon_1 \cong 2.972$, a region of regular motion takes place which is surrounded by a larger region of chaotic motion. P_1 becomes unstable for $\varepsilon = \varepsilon_1^* \cong 5.363$ bifurcating into a stable cycle of period two. We presume that each P_k, $k \geqslant 2$, has a behavior analogous to the one of P_1. Being interested only in constructing Markov partitions of the system (Ω, S_ε) for $\varepsilon < \varepsilon_1$, we did not investigate further the properties of the map S_ε, although we think that it could be worth.

The main problem in dealing with nonlinear transformations concerns the tracing of the stable and unstable manifolds. This problem, together with the associated fact that now the manifolds are curved lines, makes things complicated enough. Each invariant manifold can be approximated by numerically integrating a system of two first-order differential equations of the kind

$$\frac{dx_1}{dt} = u_1(x_1, x_2), \qquad \frac{dx_2}{dt} = u_2(x_1, x_2)$$

where u_1 and u_2 represent the components of a unit vector tangent to the manifold at $x = (x_1, x_2)$. The functions u_1 and u_2 can be computed taking advantage of the following consideration. Let S be the hyperbolic transformation to be studied and $\alpha_s(x)$ [$\alpha_u(x)$] the direction of W^s (W^u) at x. Consider an eigenvector $\mathbf{v}_s^{(n)}(x) = (v_{1s}^{(n)}(x), v_{2s}^{(n)}(x))$ [$\mathbf{v}_u^{(n)}(x) = (v_{1u}^{(n)}(x), v_{2u}^{(n)}(x))$] associated with the smaller eigenvalue of the Jacobian of T^n (T^{-n}) at x, $DT^n|_x$ ($DT^{-n}|_x$). Then, as n tends to infinity, the ratios

$$\alpha_s^{(n)}(x) = \frac{v_{2s}^{(n)}(x)}{v_{1s}^{(n)}(x)}, \qquad \alpha_u^{(n)}(x) = \frac{v_{2u}^{(n)}(x)}{v_{1u}^{(n)}(x)}$$

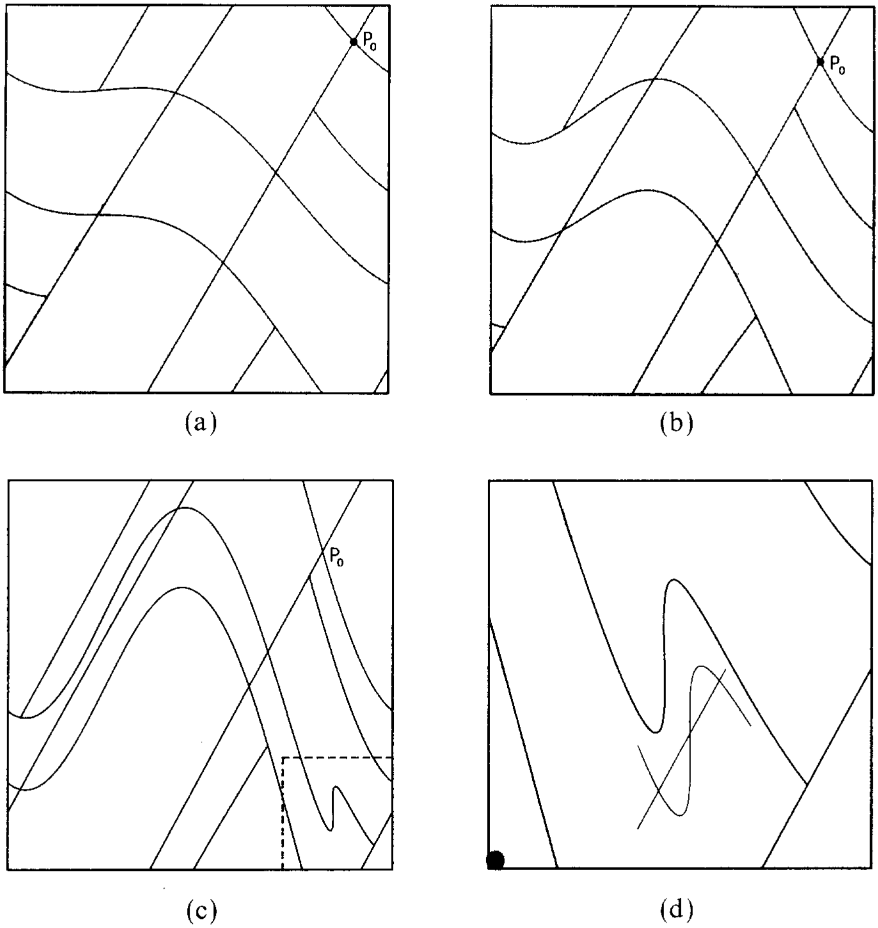

Fig. 7. (a) Seven-rectangle Markov partition of (Ω, S) obtained by using the first method for $\varepsilon = 0.7$; (b) as in (a) for $\varepsilon = 1.4$; (c) non-Markovian partition for $\varepsilon = 2.8$; (d) enlargement of the right lower corner of part (c) with the addition of two segments, one (nearly rectilinear) of the unstable manifold and one (curved) of the stable manifold, which intersect inside the same "rectangle" in three distinct points. In (a), (b), and (c) also the hyperbolic fixed point P_0 is represented.

converge to $\alpha_s(x)$ and $\alpha_u(x)$, respectively. Hence, $\alpha_s(x)$ $[\alpha_u(x)]$ can be approximated by computing the ratio $\alpha_s^{(n)}(x)$ $[\alpha_u^{(n)}(x)]$ for a n so as to guarantee the accuracy that one desires. With regard to S_ε, the number n of iterations necessary to obtain an accuracy of N exact decimal figures for both $\alpha_s(x)$ and $\alpha_u(x)$ depends on ε. n is slightly larger than N for small ε, while it is about $2N$ for ε close to ε_1. In most of our experiments we used $n = 10$. We recall that

$$DT^n|_x = DT|_{T^{n-1}(x)} \cdot \cdots \cdot DT|_{T(x)} \cdot DT|_x$$

and

$$DT^{-n}|_x = (DT|_{T^{-1}(x)} \cdot \cdots \cdot DT|_{T^{-n}(x)})^{-1}$$

The existence of the unstable fixed point P_0 allows the use of the first method to construct Markov partitions of the system (Ω, S_ε). Figure 7 exhibits the partitions corresponding to ε equal to 0.7, 1.4, and 2.8. They were obtained stopping the tracing of the manifolds at the second intersection, exactly as we did in the linear case to get Fig. 2. The sequence of pictures clearly shows how the partition changes with continuity as ε is increased, starting from the partition of Fig. 2. The largest value of ε that we considered was $\varepsilon = 3$. In this case, which corresponds to the coexistence of a region of chaos with one of regular motion, the program still produces a partition which is similar to that of Fig. 7c.

Just about the partition associated with $\varepsilon = 2.8$, an important remark has to be made. For this value of the parameter the behavior of the manifolds appear to have become quite intricated. Figure 7d, which corresponds to the right lower corner of Fig. 7c, shows one extra segment for each of the two manifolds and provides evidence for the presence of pairs x, y inside the same "rectangle" for which $[x, y]$ consists of more than one point. This fact says that the partition in question is no longer a rectangle partition, and then it is not a Markov partition although properties (2.3) and (2.4) hold.

Concerning the implementation of the second method, first of all it must be said that the nonlinearity does not allow the setting-up of a program as efficient as the one carried out for $\varepsilon = 0$. Differently from that case, in fact, many operations of the algorithm are now quite difficult to be executed directly by the computer. So, instead of a program which works in an autonomous way, we have a program which works thanks to some help with which we supply the computer, interacting with it through the graphics.

As regards the choice of the initial coverings, with a technique similar to that used in the linear case, it is possible to derive starting coverings with the requested property that every point $x \in \Omega$ is interior to at least one rectangle (Fig. 8).

We computed several Markov partitions. Two of them, relative to $\varepsilon = 1$, are shown in Fig. 9. They correspond to initial coverings by three and six S rectangles and consist of 15 and 29 elements, respectively.

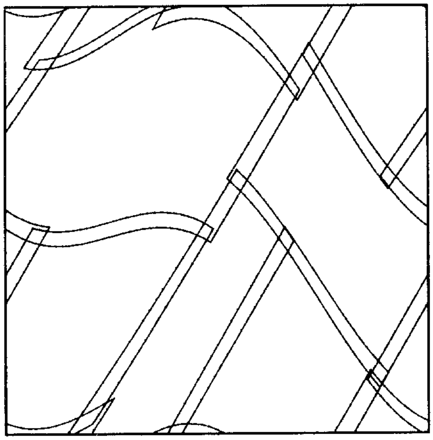

Fig. 8. A six-rectangle covering of (Ω, S_ε) for $\varepsilon = 1$.

Analogously to the linear case $\varepsilon = 0$, the check of property (2.4) was carried out by drawing on the partition itself the image of each rectangle. We note that in no case the extension of the sides at step 3 of the algorithm was necessary. This fact is presumably due to a not at all random choice of the initial coverings.

For all the Markov partitions of (Ω, S_ε) that we constructed, including the case of Fig. 7c, the associated transition matrix was computed. In all cases its largest eigenvalue turned out to be equal to $1/\lambda$, where λ is the same as in the case $\varepsilon = 0$.

A deeper analysis of the mapping S_ε and the possibility of constructing Markov partitions of (Ω, S_ε) to study its dynamics for larger values of ε appear to be subjects worth to be possibly pursued at a future time.

Finally, let us illustrate briefly some experiments that we made on a well-known dissipative mapping of the plane,

$$H(x_1, x_2) = (1 + x_2 - ax_1^2, bx_1)$$

a and b being parameters. This mapping, introduced by Hénon in 1976,[8] represents a very interesting model for chaos in dissipative dynamical systems. In fact, for a large set of parameter values and initial conditions there exists a strange attractor to which the sequence of points obtained by iteration tends. Most studies concerning the mapping have been carried out for $b = 0.3$; in particular, attention has been devoted to the strange attractor present for $a = 1.4$, known as the Hénon attractor.

This attractor is enclosed in an invariant quadrilateral on which the mapping is not hyperbolic, but it is such in an "average sense." Hence, one

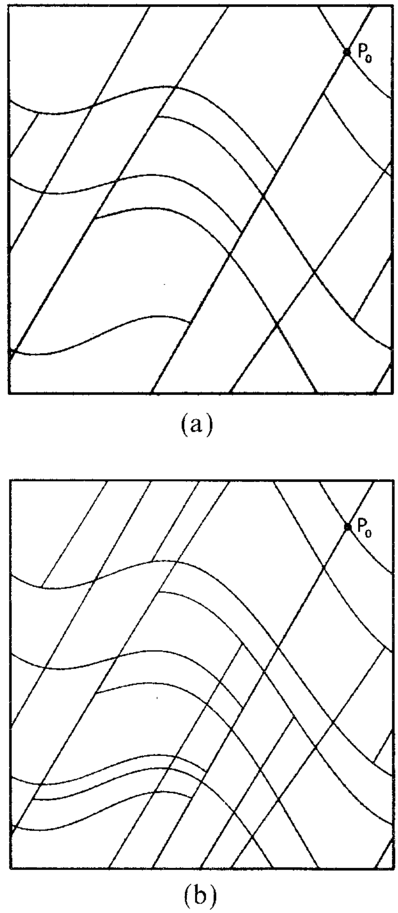

Fig. 9. Two Markov partitions of (Ω, S_ε) for $\varepsilon = 1$, by (a) 15 and (b) 29 rectangles, respectively, obtained by using the second method. The partition in (b) is derived by starting from the covering of Fig. 8.

can try to apply to the Hénon mapping the methods which work in the case of a hyperbolic system, to see whether a partition of the Hénon attractor comes out or not and, if it does, what is its meaning. We do not think that there exists a finite Markov partition; however, we do not consider useless to test the methods also in such a case of dissipative mapping.

For $a > (1-b)^2/4$ the Hénon mapping has two fixed points P and Q, given by

$$x_1 = \{b - 1 \pm [(1-b) + 4a]^{1/2}\}/2a, \qquad x_2 = bx_1$$

While Q, which is associated to the choice of the minus sign, is always unstable, P is unstable only for $a > a_1 = 3(1-b)^2/4$. We are interested in P because it is a common belief that the Hénon attractor coincides with the closure of its unstable manifold.

Our experiments are based on the parametric representation of the stable and unstable manifolds of an unstable fixed point given in Ref. 13. Referring to it for the details, we recall here the property that is at the basis of the parametrization. W^s and W^u of the point P are characterized as the images of two immersions of R into R^2 such that

$$H(W^s(t)) = W^s(\lambda_1 t), \qquad H(W^u(\tau)) = W^u(\lambda_2 \tau)$$

where λ_1 and λ_2 are the eigenvalues of the Jacobian of H at P, with $|\lambda_1| < 1$ and $|\lambda_2| > 1$ (as it is for $a > a_1$). From the two relations it is evident that the fixed point corresponds to $t = \tau = 0$.

The unstable manifold $W^u(\tau)$ is an infinite curve which is bounded. It is characterized by a longitudinal structure of pieces which are more or less parallel except for very short, in most cases sharp, arcs where the curve "turns back," and by a more complicated transversal structure of Cantor set. The most evident "turn back" arcs correspond to the first loops, starting from P, of $W^u(\tau)$, and contain all the remaining turnings in their interior. Then, roughly speaking, we can say that the whole Hénon attractor is contained inside the region delimited by a short piece of $W^u(\tau)$, say, for $|\tau| < 3$. As regards the stable manifold $W^s(t)$, it is an unbounded curve with infinitely many returns to intersect $W^u(\tau)$ for $t < 0$, while it rapidly diverges without coming back for $t > 0$. Numerical results strongly support the conjecture that for a large set of parameter values, including ($a = 1.4$, $b = 0.3$), the stable manifold $W^s(t)$ densely covers the unstable manifold $W^u(\tau)$. Figure 10, which corresponds just to ($a = 1.4$, $b = 0.3$), shows $W^u(\tau)$ for $|\tau| < 20$ and its intersections with $W^s(t)$ for $t \in (-1000, 2)$. From these considerations it follows that it is possible to cover the Hénon attractor by "rectangles" whose unstable and stable sides are pieces of $W^u(\tau)$ for low values of $|\tau|$ and pieces of $W^s(t)$ for negative t, respectively.

At this point the application of the first method is straightforward. Some simple partitions are already implicitly drawn in Fig. 8. Let us consider the simplest partition, one by three "rectangles," which is represented in Fig. 11. The three elements R_1, R_2, and R_3 that compose the partition are given by *PABC*, *PCDE*, and *EFGD*, respectively, *PA*, *BC*, *PC*, *DE*, *FG* being the stable sides and *AB*, *PC*, *CD*, *PE*, *EF*, *GD* the unstable ones. The transition matrix T is given by $t_{11} = t_{22} = t_{23} = t_{31} = 0$, $t_{12} = t_{13} = t_{21} = t_{32} = t_{33} = 1$. The numerical values of the coordinates t and τ of the points A, B, C, D, E, F, G can be easily computed with high accuracy.

This partition satisfies properties (2.3) and (2.4), but it is not a rectangle partition. In fact, analogously to the partition of Fig. 7c, property (c) of Definition 2 fails because of the presence of segments of the unstable manifold which enter and leave some "rectangle" through the same stable

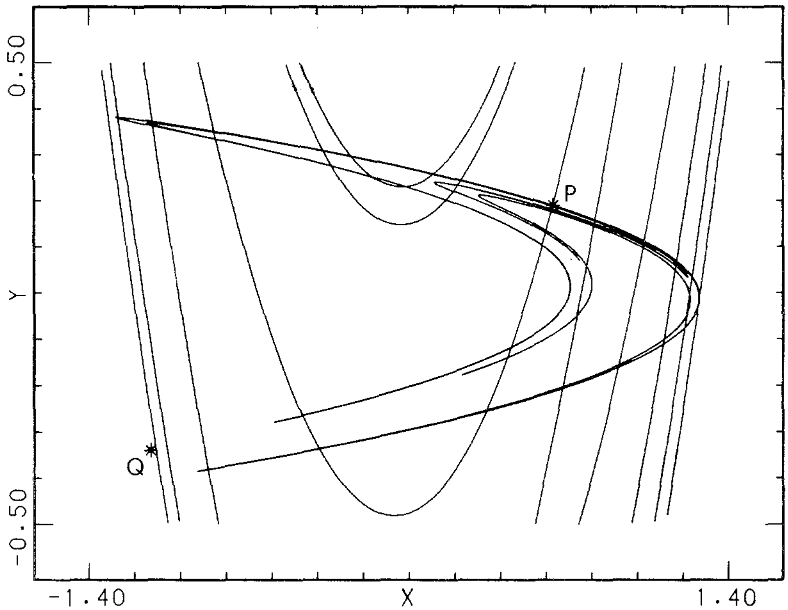

Fig. 10. Unstable manifold $W^u(\tau)$ and stable manifold $W^s(t)$ of the fixed point P of the Hénon mapping for $a = 1.4$, $b = 0.3$. $W^u(\tau)$ is traced for $\tau \in (-20, 20)$, while $W^s(t)$ is pictured for $t \in (-1000, 2)$ limitedly to the rectangle $[-1.4 \leqslant x \leqslant 1.4;\ -0.5 \leqslant y \leqslant 0.5]$. The fixed point Q is also represented.

side. This implies the possible existence of more than one point with the same H history. One can also verify directly that there exist T-compatible sequences which do not correspond to any point, for instance a sequence of the kind $\{...3213...\}$.

We considered also finer partitions obtained by tracing longer segments of the two manifolds, up to a maximum of 16 elements. All such partitions exhibited the above phenomena. Also the construction of finite partitions by many more rectangles seems to be unable to overcome the problems.

An experiment, however, seems worth trying. It would consist in the construction of a partition much finer and with the stable "sides" of some thickness. This way, which is the same as requiring some low accuracy in the approximation of the stable manifold, it might be possible to include inside the stable "sides" all the pieces of the unstable manifold that prevent the partition from being a rectangle partition. From the point of view of the applications, such an "approximate Markov partition" could be of some utility. Possibly, we will conduct this experiment next.

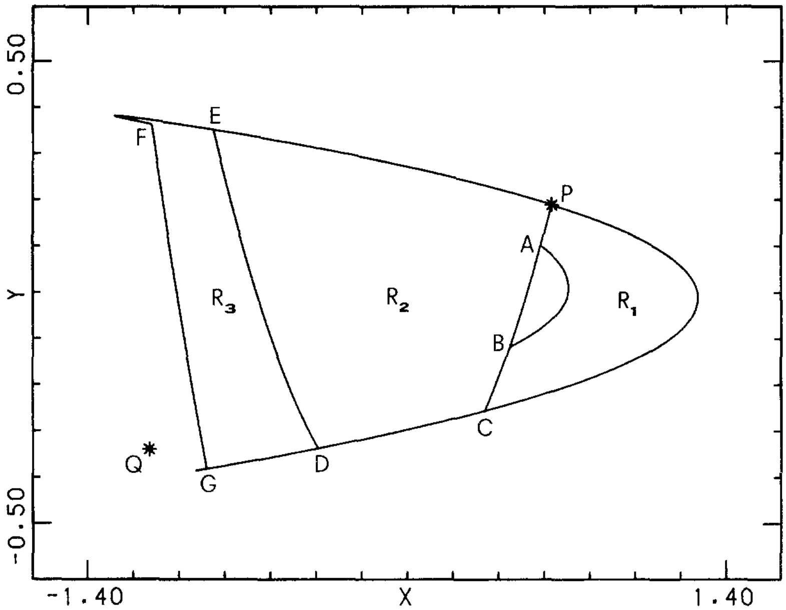

Fig. 11. Non-Markovian partition of the Hénon attractor.

To conclude the description of our computer experiments on the Hénon mapping we shall spend a few more words about the fact that also the second method was successfully applied to obtain the same partitions that were derived by using the first method. The application was not as immediate as before and required some fitting to the case in question. Having in mind how the limit covering had to be, we chose the starting covering in an appropriate way, with the unstable sides of the "rectangles" already satisfying property (2.3). Hence, the iterative procedure was needed only to correct the position of the stable sides. For this reason we were able to utilize the parametric representation of the manifolds given in Ref. 13 only in the case of $W^u(\tau)$. The stable manifold was computed by applying the same technique used for the nonlinear mapping S_ε. Another particular although not substantial modification to the method was necessary. It consisted in forcing each stable side to be transformed by H into some predetermined stable side. However, in order not to make the paper even heavier, we will not enter into further details.

ACKNOWLEDGMENTS

We are very indebted to G. Gallavotti for suggesting the paper and following it with great interest. We gratefully thank L. Russo for many useful suggestions and criticisms. We also thank M. G. Rubbiani for performing some computations and E. Caliceti for helping us with the English. The Centro di Calcolo of the University of Modena is acknowledged for providing financial support and computer facilities.

REFERENCES

1. V. M. Alekseev and M. V. Yacobson, Symbolic dynamics and hyperbolic systems, *Phys. Rep.* **75**:287–325 (1981).
2. G. Gallavotti, *Aspetti della teoria ergodica, qualitativa e statistica del moto, Quaderni dell'UMI* **21** (1981).
3. R. Bowen, Markov partitions for Axiom *A* diffeomorphisms, *Am. J. Math.* **92**:725–747 (1970).
4. Ya. G. Sinai, Markov partitions and *C*-diffeomorphisms, *Funct. Anal. Appl.* **2**:61–82 (1968).
5. Ya. G. Sinai, Construction of Markov partitions, *Funct. Anal. Appl.* **2**:245–253 (1968).
6. R. L. Adler and B. Weiss, Similarity of automorphisms of the torus, *Mem. Am. Math. Soc.* **98** (1970).
7. Ya. G. Sinai, Gibbs measures in ergodic theory, *Russ. Math. Surv.* **166**:21–69 (1972).
8. M. Hénon, A two-dimensional mapping with a strange attractor, *Commun. Math. Phys.* **50**:69–77 (1976).
9. P. Collet and Y. Levy, Ergodic properties of the Lozi mappings, *Commun. Math. Phys.*, 461–481 (1984).
10. D. Ruelle, Thermodinamic formalism, *Encyclopedia of Mathematics* (Addison-Wesley, Reading, Massachusetts, 1978), pp. 125–130.
11. V. I. Arnold and A. Avez, *Problèmes ergodiques de la mecanique classique* (Gauthier-Villars, Paris, 1967).
12. R. L. Adler and B. Weiss, Entropy, a complete metric invariant for automorphisms of the torus, *Proc. Natl. Acad. Sci. USA* **57**:1573–1576 (1967).
13. V. Franceschini and L. Russo, Stable and unstable manifolds of the Hénon mapping, *J. Stat. Phys.* **25**:757–769 (1981).