JOURNAL OF COMPUTATIONAL PHYSICS 104, 14-22 (1993)

An Efficient Method for Computing Invariant Manifolds of Planar Maps

DANA HOBSON*

Massachusetts Institute of Technology, Cambridge, Massachusetts

Received April 29, 1991; revised January 17, 1992

An efficient method for computing invariant manifolds of mappings in the plane is described and compared to a simple method in common use. The emphasis is on numerically producing smoothly resolved segments of invariant manifolds of arbitrary length while requiring a minimum number of calls to the map itself. A brief estimate of errors is given and applications are discussed. Qualitative and quantitative comparisons between the two methods are made using an example problem. The present method is applied to computing unstable manifolds for the well-known periodically forced Duffing system and the Hénon map (for which the manifold is a strange attractor). © 1993 Academic Press, Inc.

1. INTRODUCTION

In this paper we present a method for the efficient and accurate numerical computation of invariant manifolds of fixed points of planar maps. Such manifolds arise in the study of periodically perturbed Hamiltonian systems, where a continuous flow gives rise to discrete planar maps (also called diffeomorphisms) through the natural Poincaré map. The invariant manifolds (stable and unstable manifolds of fixed points) describe boundaries in the plane, analogous to discrete time samplings of streaklines in two-dimensional fluid flow. In the event that such manifolds intersect, the resulting "tangle" [6] yields several theories for the chaotic transport of regions of the plane (bounded by manifolds) to other bounded regions [3, 7]. Diffeomorphisms of the plane are also of interest for their own sake as the simplest dynamical systems which exhibit behavior similar to the Smale horseshoe. Dissipative maps can possess strange attractors, the structure of which is fundamental in understanding the behavior of more complicated systems (even infinite-dimensional systems such as the Navier Stokes equations). The celebrated Hénon map [2] has such an attracting set [1], given concisely as the closure of an invariant manifold [9]. An accurate picture of this manifold will be displayed below, showing the fine structure of the attractor.

It is generally impossible to determine an analytic expression for the manifolds in most cases of interest, so accurate numerical computations are needed. Existing numerical methods for computing invariant manifolds have several shortcomings. They are generally inefficient, requiring far too many calls to the map in question, and fail to achieve smooth resolution of long segments of manifolds. We give specific details of a method which accurately resolves arbitrarily long segments of manifolds while requiring a minimum number of calls to the map.

The paper is organized as follows. In Section 2 we give the general background for the problem we consider, including the necessary mathematical definitions and assumptions which specify the class of maps to which our method applies. We also discuss previous numerical work. We describe our method in detail in Section 3, including its implementation and a brief analysis to reveal its properties. In Section 4 we study numerical results on several test problems, measuring the results both in absolute terms and against another method. We also note various applications of the method.

2. BACKGROUND

We consider explicit planar maps of the form

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}),$$

where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ and $\mathbf{f}: \mathbb{R}^2 \to \mathbb{R}^2$. The prime denotes the image under the map $\mathbf{f}$, not differentiation. An orbit of such a map is a set of discrete points in $\mathbb{R}^2$, not a curve as is the case for ordinary differential equations. For simplicity, we assume that $\mathbf{f}$ is a C^r diffeomorphism ($r \geq 1$) so that $\mathbf{f}$ and its inverse $\mathbf{f}^{-1}$ exist and are smooth. The requirement of a smooth inverse can be relaxed in some of the computations as we shall discuss later. We also assume that $\mathbf{f}$ has an orientation-preserving saddle point $\mathbf{x}^*$ satisfying

$$\mathbf{x}^* = \mathbf{f}(\mathbf{x}^*).$$

* Present address: Lincoln Laboratory, KB318, 244 Wood St., Lexington, MA 02173-09108. dana@ll.mit.edu.

0021-9991/93 $5.00

Copyright © 1993 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Specifically, we assume that linearized matrix $D\mathbf{f}(\mathbf{x}^*)$ has two real eigenvalues λ_1 and λ_2 satisfying $0 < \lambda_1 < 1 < \lambda_2$.

These assumptions assure the existence and smoothness of stable and unstable manifolds of $\mathbf{x}^*$ which are one-dimensional curves invariant with respect to $\mathbf{f}$ [5]. The stable manifold, $W^s(\mathbf{x}^*)$, is defined as the set of points which approach $\mathbf{x}^*$ in forward iterates of the map,

$$W^s(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{f}^k(\mathbf{x}) \to \mathbf{x}^* \text{ as } k \to \infty\},$$

where $\mathbf{f}^k$ is the kth iterate of the map $\mathbf{f}$. Under our assumptions, $W^s(\mathbf{x}^*)$ is a C^r-smooth one-dimensional curve in $\mathbb{R}^2$ which is tangent at $\mathbf{x}^*$ to the local stable direction (i.e., tangent to the eigenvector corresponding to the stable eigenvalue λ_1). The global extension of this curve away from $\mathbf{x}^*$ is guaranteed by its local existence and backward mapping using the inverse map $\mathbf{f}^{-1}$. Similarly, the unstable manifold $W^u(\mathbf{x}^*)$ is defined as the set of points which approach $\mathbf{x}^*$ in backward iterates of the map,

$$W^u(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{f}^{-k}(\mathbf{x}) \to \mathbf{x}^* \text{ as } k \to \infty\},$$

where $\mathbf{f}^{-k}$ is the kth iterate of the inverse map $\mathbf{f}^{-1}$. This set forms a C^r-smooth one-dimensional manifold which is tangent at $\mathbf{x}^*$ to the local unstable direction (i.e., tangent to the eigenvector corresponding to the unstable eigenvalue λ_2). The global extension of W^u is guaranteed by its local existence and forward mapping using the map $\mathbf{f}$ itself. We note that these manifolds are not themselves single orbits of $\mathbf{f}$. They are composed of the union of all the points on all the orbits which approach $\mathbf{x}^*$ in either forward or backward iterates of $\mathbf{f}$ as appropriate. Each of these manifolds is broken into two semi-infinite branches, one on either side of the fixed point $\mathbf{x}^*$, as in Fig. 1.

The numerical methods described in this paper apply to computing the global structure of individual branches of the stable and unstable manifolds. The methods given here are directed toward computing the branches of the unstable manifold using the map $\mathbf{f}$, without need of the inverse $\mathbf{f}^{-1}$. Under the assumption that $\mathbf{f}^{-1}$ exists, these methods can be applied to compute the branches of the stable manifold as well by using the inverse map instead. In the absence of an inverse map, the branches of the stable manifold cannot be computed by the methods described here. We can relax the assumptions on the eigenvalues in practice. Provided that $\lambda_2 > 1$, we can compute branches of the unstable manifold as long as $0 < |\lambda_1| < 1$. In this case the orientation along the unstable manifold is preserved while that of the stable manifold may not be. In fact, branches of the unstable manifold can be computed whenever it is a continuous one-dimensional curve which can be approximated locally, even if the linear system near the fixed point is degenerate.

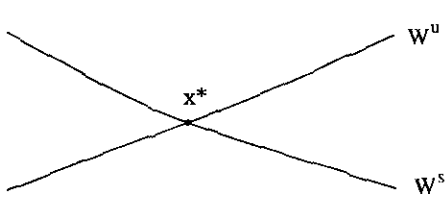

FIG. 1. Local view of the invariant manifolds of a fixed point.

A simple idea underlies all computations of branches of the unstable manifold W^u. Let U be one branch of W^u. We can break U into a sequence of *primary segments*. A primary segment is defined as a connected subsegment of U which contains a point $\mathbf{p}$ lying on U, its image $\mathbf{p}' = \mathbf{f}(\mathbf{p})$, and all the points in U which lie between $\mathbf{p}$ and $\mathbf{p}'$. Any segment which is a forward or backward image of a primary segment is also a primary segment. Given one primary segment of U, say U_0, all subsequent and previous primary segments of U can be found by taking a number of forward or backward iterates of U_0 under $\mathbf{f}$. Formally, U is the ordered union of all primary segments,

$$U = \bigcup_{i=-\infty}^{\infty} U_i,$$

where the primary segments are related by $U_{i+1} = \mathbf{f}(U_i)$, as in Fig. 2. In computations, a primary segment very close to the fixed point is mapped forward to find subsequent segments. The union of these segments traces out the portion of U beyond the initial primary segment U_0. The backward iterates merely occupy the short portion of U between $\mathbf{x}^*$ and the initial primary segment and are of little interest.

One widely used numerical method employs this idea by using the same number of points to approximate the initial primary segment and all its forward images. The initial primary segment U_0 is obtained by local analysis. It is usually taken as a straight segment along the unstable eigenvector very close to $\mathbf{x}^*$. Some large number of points N is distributed along U_0 in some fashion, usually a constant distance apart, as a discrete approximation to U_0. Each of these points is mapped forward under $\mathbf{f}$ to obtain N points which serve as a discrete approximation to the next primary segment, namely $U_1 = \mathbf{f}(U_0)$. This process is repeated M times to produce successive primary segments $U_0, U_1, \ldots, U_M$ which approximate a certain length of U when joined together end-to-end in order. The majority of

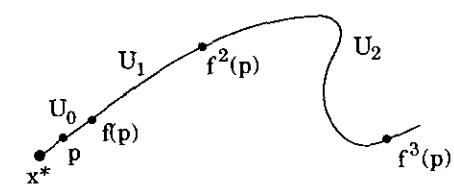

FIG. 2. Successive primary segments forming a portion of the unstable manifold.

work involved in this method is in calling the map $\mathbf{f}$, which must be done NM times to generate N points on each of the M segments after U_0.

This simple method is straightforward and very easy to implement, but it has several serious drawbacks. First, the distribution of points along each segment is not controlled. After several iterations of the initial segment, successive points on the latest segment tend to be very closely spaced in some places while large gaps between points occur elsewhere. As a result, more points than are ideally necessary are used so that the largest gap between successive points is small enough to maintain resolution of the manifold. A second drawback stems from the fact that the length of a primary segment tends to increase rapidly with the number of iterations of the initial primary segment required to produce it. In the case of tangling manifolds mentioned earlier, this length grows exponentially with the number of iterations. The number of points required to resolve the latest segment is generally much greater than the number required to resolve previous segments. Since the same number of points is used for each segment, points are wasted in the shorter segments in attempting to resolve the longest segment. Both these problems result in calling the map $\mathbf{f}$ many more times than necessary if one attempts to resolve all curves in the manifold.

An ideal method for computing U would be able to adapt the distribution and number of points on each segment. This would avoid large gaps between successive points on a segment without using too many points on the segment. It would also avoid using far too many points in the shorter segments while still resolving the longest segment. The algorithm we have developed employs such an adaptive approach in order to generate manifolds which are smoothly resolved throughout their length with a minimum number of points (i.e., a minimum number of calls to the map $\mathbf{f}$). We note that another work [10] describes a method for restricting computation to specific regions of the plane (useful when only a small region is of interest).

3. NUMERICAL METHOD

In this section we discuss a method for computing manifolds which improves upon the simple method described above. We restrict our attention to the specific problem of finding an unknown primary segment V' which is the image of a known primary segment V (i.e., $V' = \mathbf{f}(V)$). These are consecutive primary segments and are joined together along a branch of the unstable manifold so that the last point in V is the first point in V'. Subsequent primary segments can be found in the same way recursively, using the latest known segment to generate the next image segment. We assume that a set of points which lie along V is known and adequately resolve V to some order. The sense in which a set of points resolves a particular curve for our purposes will be made clear later. Our efforts concentrate on producing a set of image points which resolve the image segment V'.

We assume that the points $\{\mathbf{x}_n\}_{n=0}^{N}$ on V are known, with $\mathbf{x}_0$ lying at one end of V and $\mathbf{x}_N = \mathbf{f}(\mathbf{x}_0)$ lying at the opposite end. These points serve as a discrete approximation to V. We also assume that we have a prescription for interpolating other points from these known points in a smooth and consistent fashion. Such an interpolation scheme defines a smooth numerical curve which passes through each of the points $\mathbf{x}_n$. We work with this interpolated curve to find preimages in V (i.e., very close to V) which are to be mapped forward to image points which will specify the segment V'. In contrast, the simple method described in Section 2 used only the points $\{\mathbf{x}_n\}_{n=0}^{N}$ themselves as the preimages.

In principle, an arbitrary number of preimages can be interpolated along V, producing image points which will resolve V'. The main problem is to find a minimal set of such preimages so that the image segment V' will be resolved with a minimum number of calls to the map $\mathbf{f}$.

Our method for finding the preimages along V and their images in V' proceeds as follows. We use the first point $\mathbf{x}_0$ on V as a reference point and define s as the arclength along V measured from the point $\mathbf{x}_0$. We start with $s = 0$, corresponding to the point $\mathbf{x}_0$, and step along V in small variable increments of arclength Δs. At each step we increase s by an amount Δs. We then interpolate a point on V which lies an arclength s away from $\mathbf{x}_0$ (i.e., an arclength Δs away from the previous preimage on V). This point in V is taken as a preimage for a point in V'. The appropriate increments between preimages in V are determined in accordance with the spacing between the image points in V'. We continue this process until the latest preimage lies within a distance less than the current increments Δs of the endpoint of V, namely $\mathbf{x}_N$. We take the point $\mathbf{x}_N$ itself as the final preimage. The first forward images of these preimages under the map $\mathbf{f}$ then serve as the points which specify the primary segment V'.

The crux of the method is determining the appropriate increment Δs at each step so the resulting image points will be neither too far apart nor too close together along V'. We accomplish this essentially by trial and error at each step along V. We denote the preimages we will find in V by $\mathbf{u}_i$ and their images in V' by $\mathbf{u}_i' = \mathbf{f}(\mathbf{u}_i)$, as in Fig. 3. Given the

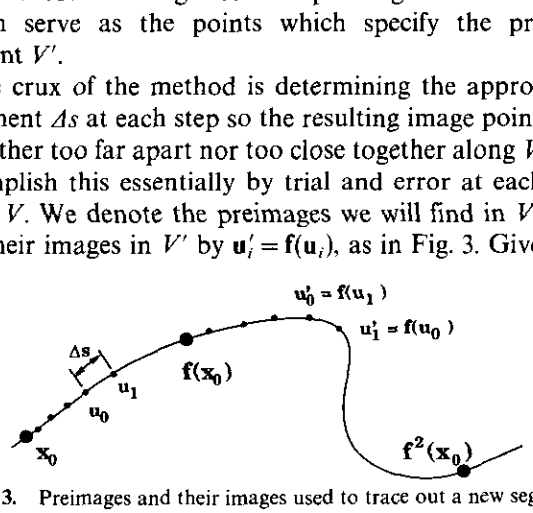

FIG. 3. Preimages and their images used to trace out a new segment.

latest preimage, call it $\mathbf{u}_0$ for convenience, and its image, $\mathbf{u}'_0 = \mathbf{f}(\mathbf{u}_0)$, we seek to determine the increment Δs so that the preimage $\mathbf{u}_1$ which lies on V a distance Δs from $\mathbf{u}_0$ will generate an image point $\mathbf{u}'_1 = \mathbf{f}(\mathbf{u}_1)$ which is an appropriate distance from the last image point $\mathbf{u}'_0$. We generate a trial preimage $\mathbf{u}_1$ by taking the increment Δs used in the previous step. If the resulting trial image $\mathbf{u}'_1$ is too far from the previous image $\mathbf{u}'_0$ we reject the points $\mathbf{u}_1$ and $\mathbf{u}'_1$ and try again using a smaller increment. If the trial image point is close enough to the previous image, we accept the trial points and repeat the process starting from these new points. If the trial image is too close to the previous image, we accept the points but also increase the increment Δs before repeating the procedure with the new points. In this way, the image points will resolve V' while using as few points as possible.

We judge whether the trial image point is an acceptable distance from the previous image point by use of a diagnostic point in addition to the points described above. Assuming the preimage $\mathbf{u}_0$ is a distance s_0 along V, the preimage $\mathbf{u}_1$ is then a distance $s_0 + \Delta s$ along V. We interpolate a third preimage $\mathbf{u}_2$ at a distance $s_0 + 2\Delta s$ along V. The image $\mathbf{u}'_2 = \mathbf{f}(\mathbf{u}_2)$ is used merely as a diagnostic point to determine how well the previous two images $\mathbf{u}'_0$ and $\mathbf{u}'_1$ resolve V' locally.

We measure the resolution by computing the angle α between two straight lines, one drawn through $\mathbf{u}'_0$ and $\mathbf{u}'_1$ and the other drawn through $\mathbf{u}'_1$ and $\mathbf{u}'_2$, as shown in Fig. 4. This angle is a measure of the spacing between the image points relative to the local radius of curvature of V'. It is straightforward to show that

$$\alpha = 2 \sin^{-1}\left(\frac{|\bar{\mathbf{u}}'_0 - \mathbf{u}'_0|}{2\,|\mathbf{u}'_1 - \mathbf{u}'_0|}\right),$$

where

$$\bar{\mathbf{u}}'_0 = \mathbf{u}'_1 + \frac{\mathbf{u}'_1 - \mathbf{u}'_2}{|\mathbf{u}'_1 - \mathbf{u}'_2|}\,|\mathbf{u}'_1 - \mathbf{u}'_0|.$$

We impose upper and lower tolerances, $\alpha_{\max}$ and $\alpha_{\min}$, in order to monitor this angle. If $\alpha > \alpha_{\max}$, the points $\mathbf{u}'_0$ and $\mathbf{u}'_1$ are too far apart, so we reduce the increment Δs and try again. If $\alpha_{\min} \leqslant \alpha \leqslant \alpha_{\max}$, we accept the trial preimage $\mathbf{u}_1$ and its image $\mathbf{u}'_1$ and proceed using these new points as the latest known points. If $\alpha < \alpha_{\min}$, the points $\mathbf{u}'_0$ and $\mathbf{u}'_1$ are too close together, so we increase the increment Δs and then proceed using the new points. In all cases, the points $\mathbf{u}_2$ and $\mathbf{u}'_2$ are discarded. In the event that $\alpha > \alpha_{\max}$ even when $\Delta s = \Delta s_{\min}$ (a prescribed minimum step size), we accept the points $\mathbf{u}_1$ and $\mathbf{u}'_1$ and proceed, recording the error.

We note here that monitoring the resolution at each step before accepting each new point is necessary to ensure the resolution of all curves in the manifold. Predicting the appropriate increment Δs at each step beforehand without monitoring the resolution afterward would result in poor resolution of the "hairpin" curves which commonly occur in these manifolds. The sudden transition from moderate values of curvature to very large values and then back to moderate values would be missed by any predictive method which assumes that the curvature does not change sharply from point to point along the manifold. The cost of resolving such sharp curves is in checking the resolution at each step and repeating steps with smaller increments where necessary.

The extra work in generating the diagnostic point at each step is also necessary. Instead of generating the diagnostic point $\mathbf{u}'_2$ to calculate the angle α formed by the points $\mathbf{u}'_0$, $\mathbf{u}'_1$, and $\mathbf{u}'_2$, we could simply use the image point just before $\mathbf{u}'_0$, namely $\mathbf{u}'_{-1}$, and calculate the angle α based on the points $\mathbf{u}'_{-1}$, $\mathbf{u}'_0$, and $\mathbf{u}'_1$. This would cut the work in half, since the map is called only once for each trial image point instead of twice. Unfortunately, this would again result in poor resolution of sharp curves. It is possible that the angle between $\mathbf{u}'_{-1}$, $\mathbf{u}'_0$, and $\mathbf{u}'_1$ will be acceptable while the angle between $\mathbf{u}'_0$, $\mathbf{u}'_1$, and any subsequent image $\mathbf{u}'_2$ will not be acceptable regardless of how close $\mathbf{u}'_2$ is to $\mathbf{u}'_1$. This occurs when the line through the points $\mathbf{u}'_0$ and $\mathbf{u}'_1$ makes an angle with the actual tangent to the curve at $\mathbf{u}'_1$ which is greater than $\alpha_{\max}$. This is detected at the next step, since no satisfactory image point after $\mathbf{u}'_1$ can be found. This problem can be avoided in all but pathological cases by using the diagnostic point as described above.

Analyzing the method's performance in an idealized setting sheds light on its properties. Figure 5 depicts three

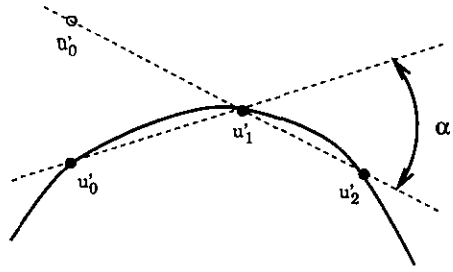

FIG. 4. The angle used in determining the resolution of the image curve by three successive image points.

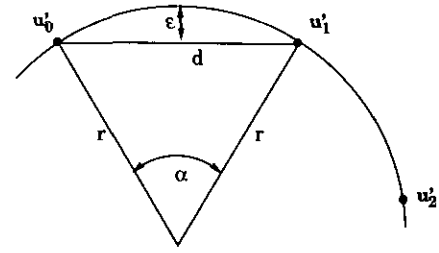

FIG. 5. Schematic in the idealized case of three successive images lying on a circle.

image points lying on a circle of radius r each a distance d from its neighbor (actual distance, not arclength). The angle α shown in Fig. 4 is equal to the interior angle in the isoceles triangle made by the points $\mathbf{u}_0'$, $\mathbf{u}_1'$, and the center of the circle. The angle is given by

$$\alpha = 2\sin^{-1}(d/2r)$$

which reduces to $\alpha \approx d/r$ for $d/r \ll 1$. Thus, for closely spaced points, α is roughly the ratio of the spacing between points to the local radius of curvature. This is a scale-independent quantity and serves as a qualitative measure of the resolution. Maintaining α between $\alpha_{\min}$ and $\alpha_{\max}$ ensures that the points will resolve tight curves (small r) in a manner similarly to nearly straight sections (large r). Points are spaced in proportion to the local radius of curvature.

Monitoring α ensures qualitative resolution of the manifold (i.e., the points appear to resolve every curve when plotted) but fails to ensure quantitative resolution. The accurate approximation of scale-dependent quantities, such as the arclength of particular segments of the manifold, can only be ensured by monitoring some scale-dependent measure. For the straight line interpolation depicted in Fig. 5, the distance between the interpolated curve (the straight line segment) and the actual manifold (the circle) is $\varepsilon \approx r\alpha^2/8 \approx d\alpha/8$ for $\alpha \ll 1$. Using a parabolic interpolation through $\mathbf{u}_0'$, $\mathbf{u}_1'$, and $\mathbf{u}_2'$ we find the maximum distance between the circle and the interpolated curve between $\mathbf{u}_0'$ and $\mathbf{u}_1'$ is $\varepsilon \approx \frac{3}{128}\, d\alpha^3$ for $\alpha \ll 1$. We find similar estimates hold for the arclength of the manifold between $\mathbf{u}_0'$ and $\mathbf{u}_1'$. The errors made in approximating these scale-dependent quantities are all proportional to $d\alpha^n$, where the constant of proportionality and the exponent n depend on the quantity being approximated and the order of the interpolation. Monitoring the product $d\alpha$ in addition to α itself would result in the ability to control the errors in all the scale-dependent quantities, since all such errors can be expressed as $d\alpha$ times some power of α. Making sure that both are small ensures that all quantitative errors will also be small.

Note that monitoring d itself in addition to α would not allow points to spread far enough apart in very straight sections. Also, monitoring just $d\alpha$ and not α would result in poor qualitative resolution of sharp corners, where the points are numerically very close to one another even if they do not geometrically resolve the sharp curve. Accordingly, in practice we monitor both $d\alpha$ and α. We decrease the increment Δs at a particular step if either measure exceeds its prescribed maximum and we increase the increment only if both fall below their prescribed minimum.

A brief error analysis provides a bound on the distance between the curve computed by the present method and the actual unstable manifold. In what follows we let U be the true unstable manifold obtained by piecing together the individual primary segments U_0, U_1, etc. which satisfy $U_{k+1} = \mathbf{f}(U_k)$. Similarly, we let X be the curve produced by piecing together the segments X_0, X_1, etc., computed using the simple method described in Section 2 (using *all* the points in the initial segment X_0 and mapping them forward successively). Also, let Y be the curve produced by the segments Y_0, Y_1, etc., computed using the present method. We will find bounds between X and U and the between Y and X. The triangle inequality will then yield the desired bound between the curve Y, computed using the present method, and the true unstable manifold U.

We first estimate the distance between the curve X, produced by the simple method, and the unstable manifold U. We assume $\mathbf{x}_0$ is a point on the approximate initial segment X_0 and $\mathbf{u}_0$ is a point on the true initial segment U_0, lying within a distance ε_a of $\mathbf{x}_0$ (i.e., $d(\mathbf{x}_0, \mathbf{u}_0) < \varepsilon_a$, where d is the Euclidean distance). The parameter ε_a is a bound on the error made by the approximation to the initial segment. A second source of error in the simple method is machine roundoff in evaluating image points numerically. The sequence $\{\mathbf{x}_k = \mathbf{f}^k(\mathbf{x}_0)\}$ formed by the images of $\mathbf{x}_0$ is a δ-pseudo-orbit. That is, $d(\mathbf{x}_{k+1}, \mathbf{f}(\mathbf{x}_k)) < \delta$ for all k, where δ is the maximum machine roundoff error in evaluating images. The shadowing lemma [8] states that for each $\varepsilon > 0$ there is a $\delta > 0$ such that each δ-pseudo-orbit is "shadowed" by an actual orbit $\{\mathbf{v}_k = \mathbf{f}(\mathbf{v}_0)\}$ within ε, i.e., $d(\mathbf{x}_k, \mathbf{v}_k) < \varepsilon$ for all k. This is useful if the roundoff errors ($\approx \delta$) are small enough to apply to values of ε of interest. This shadowing orbit $\{\mathbf{v}_k\}$ can be compared analytically to the orbit $\{\mathbf{u}_k = \mathbf{f}(\mathbf{u}_0)\}$. The mean value theorem implies that $d(\mathbf{u}_k, \mathbf{v}_k) \leqslant Cd(\mathbf{u}_{k-1}, \mathbf{v}_{k-1})$, where C is the maximum of the norm $\|D\mathbf{f}\|$ over an appropriate portion of the plane to which the manifold U is bounded. Repeated application yields $d(\mathbf{u}_k, \mathbf{v}_k) \leqslant C^k d(\mathbf{u}_0, \mathbf{v}_0) < C^k(\varepsilon + \varepsilon_a)$. Using the triangle inequality, we have finally

$$d(\mathbf{x}_k, \mathbf{u}_k) \leqslant d(\mathbf{x}_k, \mathbf{v}_k) + d(\mathbf{v}_k, \mathbf{u}_k) < \varepsilon + C^k(\varepsilon + \varepsilon_a).$$

Thus any point on the approximate initial segment X_0 generates a sequence of numerical images which stay within an exponential bound of the true curve U. So, the simple curve X stays within a bound of the unstable manifold.

We next establish a bound on the distance between the simple curve X and the present curve Y. Each of the segments Y_k comprising Y are images of curves made up of points interpolated from the previous segment. It will be necessary to consider these interpolated points explicitly. We let Z_k be the curve interpolating the segment Y_k, so preimages are taken along the curve Z_k. The points specifying the next segment Y_{k+1} are the images of points along Z_k. In principle we have $Y_{k+1} = \mathbf{f}(Z_k)$.

In what follows we neglect machine roundoff errors and initial approximation errors in comparison to interpolation errors. The error involved in interpolating preimages, absent in the simple method, is usually dominant and is the

most serious drawback of the present method. In the case of maps arising as Poincaré maps of continuous systems there will be the additional error from the solution of ordinary differential equations in order to implement the Poincaré map. We neglect these errors also. If our interpolation makes a maximum error ε_i, then for each $\mathbf{z}_k \in Z_k$ there is a point $\mathbf{y}_k \in Y_k$ such that $d(\mathbf{z}_k, \mathbf{y}_k) < \varepsilon_i$. In a manner similar to the above discussion, we can derive an inequality for the distance between the point $\mathbf{y}_k \in Y_k$ and the nearest point $\mathbf{x}_k \in X_k$,

$$\begin{aligned} d(\mathbf{y}_{k+1}, \mathbf{x}_{k+1}) = d(\mathbf{f}(\mathbf{z}_k), \mathbf{f}(\mathbf{x}_k)) &\leqslant Cd(\mathbf{z}_k, \mathbf{x}_k) \leqslant \cdots \\ &\leqslant C(d(\mathbf{z}_k, \mathbf{y}_k) + d(\mathbf{y}_k, \mathbf{x}_k)) \\ &< C(\varepsilon_i + d(\mathbf{y}_k, \mathbf{x}_k)), \end{aligned}$$

where C is defined as above. Letting $D_k = d(\mathbf{y}_k, \mathbf{x}_k)$ and noting that $D_0 = 0$ (since the present method and the simple method use the same initial segment $X_0 = Y_0$), we find that

$$\begin{aligned} D_k &< C(1 + C + C^2 + \cdots + C^{k-1})\,\varepsilon_i \\ &= C\frac{C^k - 1}{C - 1}\,\varepsilon_i. \end{aligned}$$

This holds for any point $\mathbf{y}_k \in Y$, so this is an upper bound for the distance between X and Y along their kth segments. The triangle inequality can now be used with this bound and the bound between X and U above to give a bound between Y and U. Since errors in the simple method and the present method accumulate exponentially, it is clear that any method based on mapping an initial segment forward many times will eventually fail for portions of the manifold sufficiently far from the initial segment. The above estimates are by no means sharp and we have found in practice that much higher accuracy than predicted by these bounds is achievable.

The next section gives example computations using both methods. The interpolation scheme we have used in this paper is based on third-order Taylor approximation. Other interpolation schemes, such as cubic splines or rational function interpolation, may be used instead without a significant effect on the implementation of the method. As noted above, local errors tend to scale with the maximum allowed value of the product $d\alpha$. Higher order interpolation schemes would reduce such errors by various factors of α, but we have found that extreme accuracy is best controlled by making the allowable maximum of $d\alpha$ very small instead of trying very high order interpolation.

4. RESULTS

In this section we give examples of invariant manifolds computed using the present method described in Section 3. We first compare the performance of the present method to the simple method described in Section 2 by computing an unstable manifold of the McMillan map. We go on to use the present method on unstable manifolds of the periodically forced Duffing equation and the Hénon map in order to assess its performance in absolute terms.

In our first example, we use a form of the McMillan map [4],

$$\begin{aligned} x' &= y \\ y' &= -x + 2y\left(\frac{\mu}{1 + y^2} + \varepsilon\right). \end{aligned}$$

We have chosen $\mu = 2.0$ and $\varepsilon = 0.05$ for this example. This map has a saddle point at the origin. We compute the upper-right branch U of the unstable manifold, finding the initial primary segment U_0 as follows.

We represent U near the saddle (i.e., $|x| \ll 1$) as $y = p(x) = c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n$ with n taken large enough so that the error between this value If y and the true value on U at a given x is sufficiently small. We require that the image curve also satisfy this, giving the functional equation

$$p(p(x)) = -x + 2p(x)\left(\frac{\mu}{1 + p^2(x)} + \varepsilon\right)$$

which is to be satisfied for small values of $|x|$. There are two possible values for c_1 (corresponding to the stable and unstable manifolds, respectively). Once c_1 is chosen (it is the ratio of components of the eigenvector along the unstable direction of the saddle), the remaining coefficients can be found uniquely in ascending order. A symbolic manipulator is helpful in this task. The symmetry in this map forces the even-numbered coefficients to vanish. We find $c_1 = 3.83955301681733$ and $c_3 = -4.01849014139598$.

All points on U_0 are assumed to lie on $y = c_1 x + c_3 x^3$. We take the initial point of U_0 as (0.001, 0.003839549) and the final point as (0.003839549, 0.014741924). The error between this polynomial approximation and the true manifold along this segment is less than 10^{-11}. A plot of a portion of U consisting of U_0 and its first 10 iterates joined end-to-end $(U_0, U_1, \ldots, U_{10})$ is shown in Fig. 6. The manifold starts at the origin and moves up and right before swinging clockwise and folding near the origin.

Figure 6 consists of data points generated by the present method and connected by straight line segments to form a curve. We set the upper bounds of 0.3 on the angle α and 0.001 on the product $d\alpha$. This required 1560 calls to the map while the manifold itself consists of 618 points. Figure 7 has details showing only the data points. A closeup view of the folding near the origin is shown in 7a while subsequent detail is shown in 7b–d. These plots show that the points are

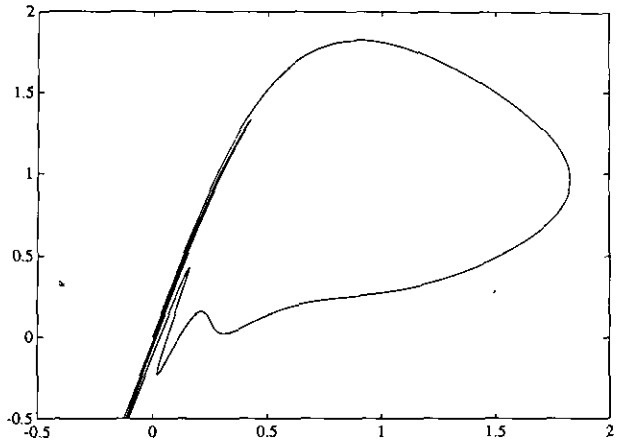

FIG. 6. A portion of the unstable manifold of the McMillan map.

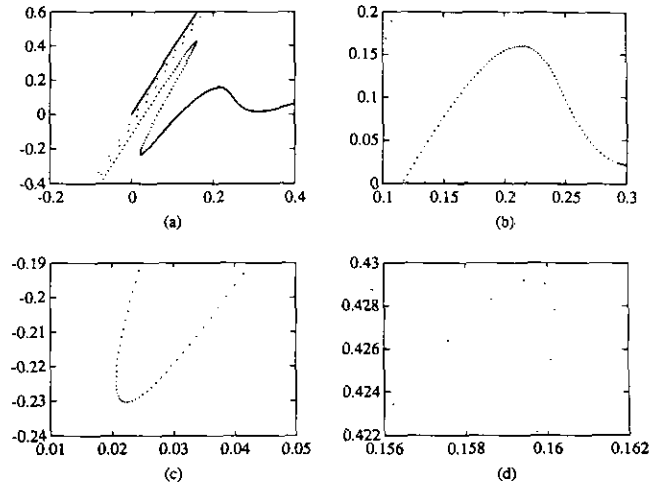

FIG. 8. Same as Fig. 7 except data is generated using simple method.

spaced roughly in proportion to the local radius of curvature, in accordance with the method's design. Even the sharp curve in d is seen to be smoothly resolved. We note that the points are only close together in the sharp curves. We also note that the resolution of all curves is similar regardless of the absolute scale of the curve, which is apparent from 7b–d. The points bunch up tightly in the sharpest part of each curve and spread farther apart as the manifold straightens.

Figure 8 shows the results of the same computation using the simple method of Section 2. In order to present a fair comparison, we have chosen the number N of points per primary segment so that the total number of calls to the map is again 1560. For this method the number of calls is equal to the number of points in the manifold. We note that in general there are many more points per unit length along the early segments than in the later segments. This is evidenced by the dense distribution of points in the initial segments of Fig. 8a which continues until the segments in which the manifold is folded sharply near the origin. In these later folded segments the points are generally spread farther apart. Within any given segment, the points are not spaced according to local curvature. The points are roughly evenly spaced within any given segment with only a slight tendency for the points to bunch up in sharp curves, as seen in 8b and c. This tendency is not sufficient to ensure resolution. The points in the later segments become spread too far apart and resolution is eventually degraded, as in 8d, where the points shown barely resolve the curve.

The tendency of the simple method to lose resolution for the latest segments is manifested most clearly in a quantitative study. Table I gives various approximations to the arclength of the primary segments U_{13}, U_{14}, and U_{15} (not plotted in Fig. 6), computed by the two methods. Columns (a) through (c) show results using the simple method. For

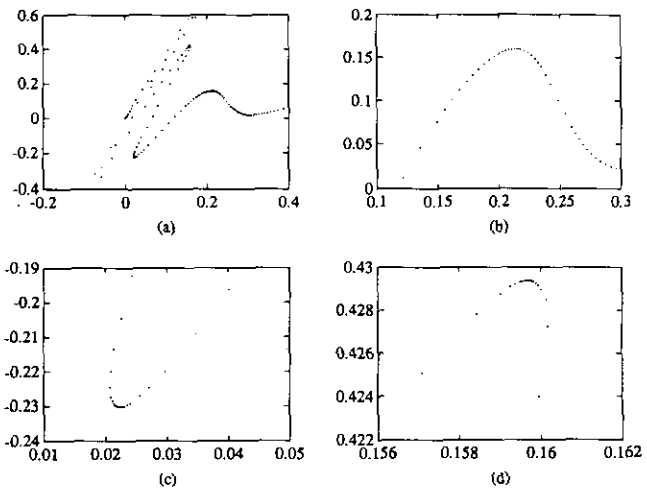

FIG. 7. Detail from Fig. 6 without curve through the data points.

TABLE I

Arclength of three consecutive primary segments

	Simple method			Present method			"Exact"
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
U_{13}	20.146	20.120	20.120	20.020	20.117	20.119	20.120
U_{14}	22.661	23.120	23.099	22.843	23.098	23.097	23.098
U_{15}	23.548	27.885	29.086	28.407	28.991	29.034	29.037
Calls	8000	32000	80000	4975	7671	11395	90813

Note. (a) Simple method, 8000 calls to the map (500 points per segment); (b) Simple method, 32,000 calls to the map (2000 points per segment); (c) Simple method, 80,000 calls to the map (5000 points per segment); (d) Present method, 4975 calls to the map, $\alpha_{max} = 0.45$, $(d\alpha)_{max} = 0.01$; (e) Present method, 7671 calls to the map, $\alpha_{max} = 0.45$, $(d\alpha)_{max} = 0.003$; (f) Present method, 11,395 calls to the map, $\alpha_{max} = 0.45$, $(d\alpha)_{max} = 0.001$; (g) Present method, 90,813 calls to the map, $\alpha_{max} = 0.15$, $(d\alpha)_{max} = 0.00001$.

this method the number of calls to the map is equal to the total number of points used, which is equal to 16 times the number of points in each segment (since there are 16 segments). Columns (d) through (g) show results of the same computation using the present method with various levels of resolution. For this method we specify upper and lower limits on α and $d\alpha$ as described in Section 3. Column (g) was computed using extreme resolution and represents the "exact" results. We can see in (a) that the simple method is fairly accurate for the arclength of U_{13} even for 8000 calls to the map, but does a poor job approximating the arclength of U_{14} and U_{15} at this level. The simple method requires 80,000 calls to resolve the arclength of U_{15} as well as the present method does, using only 7671 (less than one-tenth that required by the simple method). The simple method must waste many points over-resolving the earlier segments U_{13} and U_{14} while attempting to resolve the last segment U_{15}. The present method resolves each segment to roughly the same accuracy without wasting points in any of the earlier segments.

We now proceed to a second example, intended to illustrate the capabilities of the present method in absolute terms. We consider the periodically forced Duffing system

$$\dot{x} = y$$
$$\dot{y} = x - x^3 + \gamma \cos t$$

The Poincaré map from $t = 0$ to $t = 2\pi$ associated with these equations has a saddle point at $(x^*, y^*) = (-\gamma/2 - \gamma^3/20 + O(\gamma^5), 0)$. We have plotted a long portion of one branch of the unstable manifold of this Poincaré map for $\gamma = 0.1$ in Fig. 9a. Detail of the large cluster of curves at the upper left corner of 9a appears in 9b. Figure 9c shows the points used in 9b. Figure 9d shows the points resolving the uppermost thin curve at the far left of 9a. Figure 9e is an extreme closeup of the very thin curve coming nearly to a point around (0, 0.43). Both curves in 9d, e are smoothly resolved in their sharpest portions without being over-resolved by too many points in the nearly straight portions.

Finally, we present a highly accurate computation of a branch of the unstable manifold of the Hénon map. The closure of this branch is the much-studied strange attractor of this map [9]. This branch has been proved to be an attracting set [1]. The Hénon map [2] is given by

$$x' = 1 + y - ax^2$$
$$y' = bx.$$

For $a = 1.4$ and $b = 0.3$ we find a fixed point at $(x^*, y^*) = (-1.131354477, -0.339406343)$. We approximate the manifold near this saddle point as $y - y^* = p(x - x^*)$, where $p(x) = c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5$ with $c_1 = 0.0920295620408391$, $c_2 = 0.0120205020679811$,

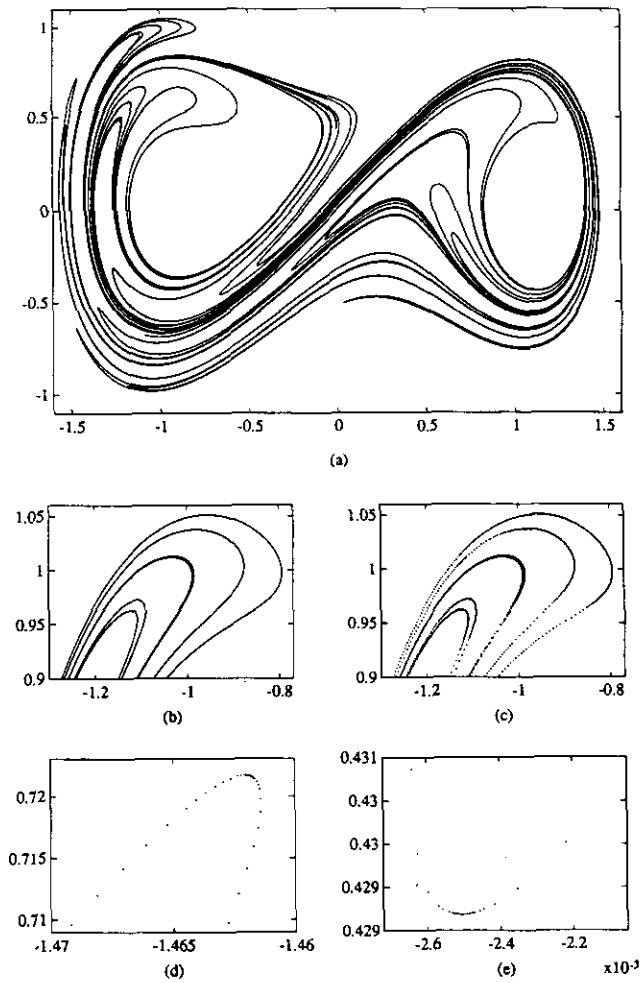

FIG. 9. The unstable manifold of the periodically forced Duffing system. Details show magnifications of the distribution of points near sharp curves.

$c_3 = 0.0031318113005733$, $c_4 = 0.0010191115245318$, and $c_5 = 0.0003712820303658$. The initial primary segment U_0 starting at $x - x^* = 0.001$ is specified by points satisfying $y - y^* = p(x - x^*)$ and differs from the true manifold by less than 10^{-16}. Figure 10 shows a portion of U consisting of primary segments U_0 through U_{18} with total arclength 408.764. For this computation we have bounded α below 0.3 and $d\alpha$ below 0.000001. Based on these bounds we estimate the maximum error introduced by interpolating preimages is less than 10^{-9} for our particular interpolation method. Figure 10 belies the fascinating complexity of this strange attractor. Figure 11 shows more detail of some sharp curves. The last part in Fig. 11 shows the detail in the very sharp curve at the bottom of Fig. 10. The minute structure apparent in these views is shown more clearly in Fig. 12. Figure 12 shows successive magnifications of a very straight portion of the manifold (similar to magnifications shown by Hénon). The small square in each view indicates the scale on which the next view is plotted. The manifold appears to be

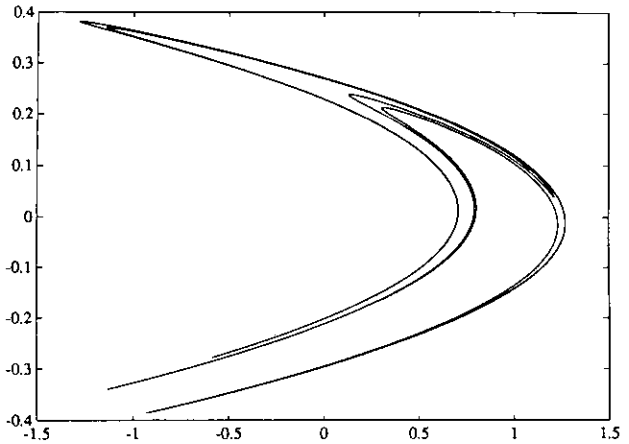

FIG. 10. The unstable manifold of the Hénon map.

identical in each of these magnifications, even though the final view is plotted on a scale 1000 times finer than the first view. This clearly shows the self-similar structure of the strange attractor. It is much like the cartesian product of a Cantor set with a straight line.

The accurate knowledge of the strange attractor of the Hénon map represented here can be used to compute some of its properties. One example is the computation of the box dimension or capacity of the attractor by finding how the number of boxes in a covering of the manifold scales with the size of the boxes. The present computation allows a more direct approach than naive numerical studies which seek to trace out the attractor by studying a single orbit of the map. Such studies result in scatter plots of points which lie on or near the attractor. These lack the sense of continuity and ordering which are inherent in the computation of the attractor as an unstably manifold. This continuity and ordering allows a constructive covering of a finite (but very long) portion of the manifold and guarantees that points are not missed along the way.

The present method can also be of use in studying

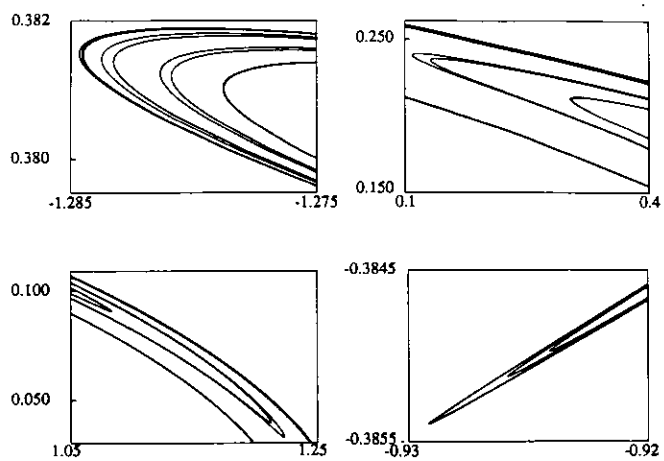

FIG. 11. Detail from Fig. 10.

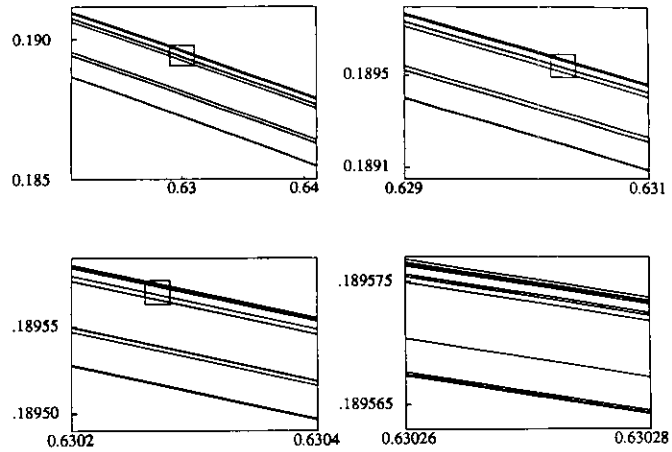

FIG. 12. Successive magnifications showing the self-similar structure of the attractor.

phenomena relating to resonance bands in two-dimensional maps. The invariant manifolds defining stochastic layers in resonance bands can be computed using the present method on the appropriate iterate of the map. The overlapping of resonance bands can thus be studied precisely by plotting the appropriate manifolds and determining if intersections between various bands occur.

Finally, an accurate knowledge of the lengths of successive segments allows estimation of the stretching rate of manifolds. This rate is important in some two-dimensional fluid problems in combustion and chemical reactions, where the interface between two reacting fluids is given by an invariant manifold. The overall length of the interface determines in part the rate at which the two fluids react.

ACKNOWLEDGMENTS

Part of this work was carried out at Caltech under a Hertz Foundation Fellowship and the Department of Energy Grant DE-FG03-89ER25073. The author acknowledges helpful comments from the referees.

REFERENCES

1. M. Benedicks and L. Carleson, *Ann. of Math.* **133**, 73 (1991).
2. M. Hénon, *Commun. Math. Phys.* **50**, 69 (1976).
3. R. S. MacKay, J. D. Meiss, and I. C. Percival, *Phys. D.* **13**, 55 (1984).
4. E. M. McMillan, in *Topics in Modern Physics* (Colorado Assoc. Univ. Press, Boulder, CO, 1971), p. 219.
5. Z. Nitecki, *Differentiable Dynamics* (MIT Press, Cambridge, MA, 1971).
6. H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Gauthier–Villars, Paris, 1899).
7. V. Rom-Kedar and S. Wiggins, *Arch. Rat. Mech. Anal.* **109**, 239 (1990).
8. D. Ruelle, *Elements of Differentiable Dynamics ad Bifurcation Theory* (Academic Press, San Diego, 1989).
9. C. Simó, *J. Stat. Phys.* **21**, 465 (1979).
10. Z. You, E. J. Kostelich, and J. A. Yorke, *Int. J. Bifurcation Chaos* **1**, 605 (1991).