ELSEVIER

Physica D 81 (1995) 280–294

PHYSICA D

Controlling chaotic transport through recurrence

Erik M. Bollt, James D. Meiss

Program in Applied Mathematics, University of Colorado, Boulder, CO 80309, USA

Received 15 April 1994; revised 18 August 1994; accepted 21 August 1994
Communicated by H. Flaschka

Abstract

Transport times for a chaotic system are highly sensitive to initial conditions and parameter values. We present a technique to find rough orbits (epsilon chains) that achieve a desired transport rapidly and which can be stabilized with small parameter perturbations. The strategy is to build the epsilon chain from segments of a long orbit – the point is that long orbits have recurrences in neighborhoods where faster orbits must also pass. The recurrences are used as the switching points between segments. The resulting epsilon chain can be refined by gluing orbit segments over the switching points, providing that a local hyperbolicity condition is satisfied. As an example, we show that transport times for the standard map can be reduced by factors of 10^4. The techniques presented here can be easily generalized to higher dimensions and to systems where the dynamics is known only as a time-series.

1. Introduction

In recent years, much attention has been given to the realization that the extreme sensitivity which characterizes chaotic dynamical systems is actually advantageous in their control since small perturbations produce large effects [7,13,14,16]. On the other hand small errors can quickly get out of control, and may overwhelm any attempted corrections. Thus the major difficulty is to find a scheme to decide when and where judiciously chosen perturbations should be applied.

This paper addresses the problem of time-optimal control, or *targeting*. This is the technique to steer a dynamical system from near an initial condition a to near a target point b in the shortest possible time. There have been several recent, successful approaches to this problem. A first approach, to use a single small perturbation in a parameter [16], has been demonstrated both numerically for one- and two-dimensional maps, and in the laboratory on dynamics which were approximately modeled by a one-dimensional map constructed from experimental time-series [15]. This method can drastically reduce the transport time providing that a fast orbit between a and b actually exists at the nominal parameters. Another interesting approach, due to Starobinets and Pikovsky [17], uses well-known time-optimal techniques to target again with a single parameter perturbation. If, however, the fastest transporting orbit (within allowed control constraints) is long or difficult to find, the algorithm becomes impractical. The standard map studied in Section 4 is just such a case [9].

Kostelich et al. [7] extended targeting to higher dimensions, and studied the dissipative

SSDI 0167-2789(94)00189-8

kicked double rotor map as an example. Here numerically determined short (typically about 30 iterates) orbit segments were arranged in a tree hierarchy defined so that there is a set of paths, that lead to a set of paths, etc., that lead to a pre-determined target point b. This tree is stored as a library of known paths. The tree branches actually provide a set of epsilon chains leading to b, pseudo-orbits that miss being exact orbits by a small phase space error at the junctions. The idea is that if there are enough branches, any arbitrary initial condition a will rapidly iterate close enough to the tree, and once there can be stabilized to reach b with small parameter perturbations. Nonetheless, the above technique may still be ineffective when natural transport is too slow to create a tree with a reasonable amount of computation.

In order to best choose the orbit segments of the epsilon chain, one would like to know where to most efficiently switch between the segments. We present here an alternative approach for building and managing a library of numerically known orbit behaviors so that this information can be quickly accessed to build fast transporting epsilon chains. Our main result is that the switching points reveal themselves by studying the path of a nonoptimal orbit that, however, eventually achieves the desired target objective.

In Section 2 we argue that recurrences are common in slow orbits and that these should be tested as switching point candidates. The resulting epsilon chain can be refined, as we show in Section 3, by gluing patches across the switching points. This requires finding the stable and unstable directions along the original orbit, and provides a patch orbit segment that:

(1) Skips the recurrent loop, often representing the bulk of the orbit's length.
(2) Converges to the original orbit backwards in time.
(3) Converges to the original orbit forward in time from the point of recurrence.

The patch size can be chosen to meet the control saturation bound. Hence, we can effectively pick and choose desired segments of a slow orbit, using hyperbolicity to our advantage to leverage away the error upon gluing in an orbit patch. Gluing has been used for example in proving the shadowing theorem for Axiom A systems [2] as well as in other contexts [18].

The obvious advantage here is the possibility of constructing fast orbits between any two points in accessible phase space. By following an arbitrary orbit for a long time and recording its local stability properties, any two points near the observed dynamics can now be reached by an epsilon chain constructed from segments of the observed dynamics. A chaotic orbit will cover all of the accessible phase space and so will have most starting and target points, close to the accessible set, somewhere within its length. We propose that this technique is also applicable in the case where we only have an approximate model of the dynamics formed by a time series of data from a real world system, and from which we can make local predictions according to the work of Farmer and Sidorowich [4]. This is possible since no inverse image of the map is necessary to our method.

We use a local linear controller at each step of the predicted orbit to diminish the effects of modeling error and system noise. Local linear controllers have been demonstrated using accessible dynamic parameters for a number of chaotic systems, and have also been shown to be effective even for dynamics specified only by time-series. An effective method is to use accessible parameters to cause the image of an initial condition to have no component on the unstable manifold of the target point [9,13], hence knowledge pertaining to the map's parameter derivatives and the unstable directions are required. More traditional "pole placement" techniques yield much the same result [14].

In Section 4 we demonstrate our method for the standard map, which has notoriously slow transport, investigating transport distributions before and after control. We also investigate the hyperbolicity of our trajectories before and after control, by computing the distribution of angles between the stable and unstable manifolds [8].

2. Chaos and recurrence

In this section we will discuss the difference between an optimal trajectory and a nonoptimal trajectory of dynamics arising from a given map. We write a map in the form

$$z' = T_k(z), \quad z \in \mathbb{R}^d \quad \text{and} \quad k \in \mathbb{R}^r, \tag{2.1}$$

which could have been derived from the continuous time flow of a differential equation by Poincaré section. We demonstrate rather general conditions under which a nonoptimal path has a nearby path that reduces the time-optimal cost function.

Our problem is to find an initial condition z_a near the starting point a, and a control strategy consisting of a set of parameter values $\{k_i\}$, that will cause z_a to iterate near the target point b as quickly as possible. Fig. 1 displays the orbit of such a control solution. Thus we wish to minimize a cost function,

$$I(\{k_i\}, z_a) = n, \tag{2.2}$$

where n is the first time such that $z_b = (\Pi_{i=1}^{n} \circ T_{k_i})(z_a) \in B_\varepsilon(b)$, subject to the constraints

$$\|k_i - k_0\|_{\mathbb{R}^r} < \Delta, \quad z_a \in B_\varepsilon(a). \tag{2.3}$$

Here $B_\varepsilon(a)$ is the ball of radius ε around the point a. Hence n applications of the map T with the control parameters $\{k_i\}_{i=1}^{n}$ near the nominal

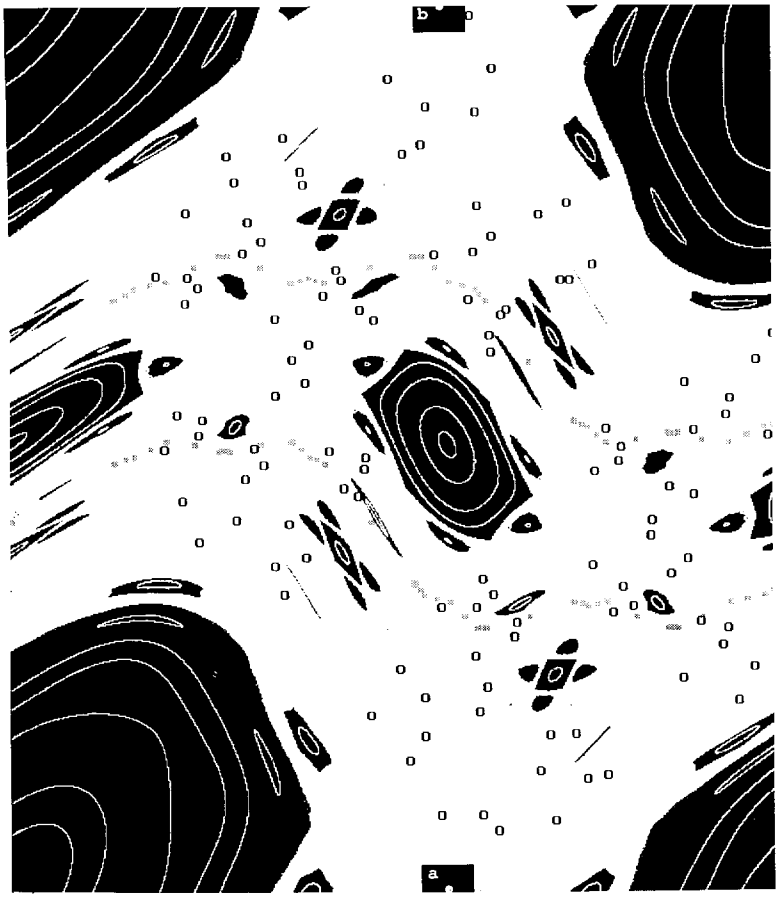

Fig. 1. A phase space portrait of the standard map for the range $0 \le x < 1$ and $0 \le y < 1$. The o's are centered on the 131-step path between a and b constructed by cutting the recurrences from a 80307-step orbit. The $1/\gamma$ and $1/\gamma^2$ golden mean cantori are represented by the gray squares in the middle region of the picture.The point a is located at (0.5,0.0) on the (0,1) hyperbolic point, and b is located at (0.5,1.0) on the (1,1) hyperbolic point.

value k_0 give a path between the two ε balls of the starting and target points respectively. This is known as the "minimum-time control problem" [6,19].

In general, the minimum occurs not at a fixed value of k, but at a program of parameter values $\{k_i\}_{i=1}^{n}$. Knowing when and where to vary k leads us to consider whether there might be regions of phase space through which transport must occur. These regions are analogous to hub airports used in deciding which trajectory an airplane takes between Los Angeles and Boston. Finding the actual route seems hopeless at first, until we realize that there are just a few possible switching points: Denver, Chicago and Houston. With this sort of model, we reduce the infinite dimensional search in all of "phase space" for places to program k to a few small switching regions. Targeting can then successfully be performed between a finite, hopefully complete set of such regions. The problem then becomes just one of permutations amongst the fastest orbits between switching regions, but one which still may be intractable if the number of switching regions is high and there is no obvious order in their importance. To find these switching regions seems at first to require detailed knowledge of the transport properties of the chaotic system; unfortunately, these are only understood for the case of two-dimensional mappings (see e.g. [11,20]).

The answer for this latter case is that transport between two regions separated by a homoclinic orbit occurs by landing in the exit lobe delineated by intersections of stable and unstable manifold segments of the homoclinic orbit. Thus a transiting orbit must always have a point in the lobe between the two regions, regardless of its complexity. The lobes are examples of switching regions. The most efficient transporting orbit lands on a subset of the exit lobe that does not intersect the lobe again before intersecting the target b. The inefficient orbit will in fact recur in the exit lobe. A complete description of the transport may even be reducible to a shift on a set of symbols such as the Smale horseshoe example [20]. Unfortunately, using such a description to quantitatively define transport between a and b requires knowing the stable and unstable manifolds that delineate the important switching lobes. In addition, the geometry for higher dimensional phase space is not yet well understood.

Even though the lobe structure is special to the case of two dimensions, there is an important aspect of it that has much wider application: recurrence.

Recurrence Lemma. Let (2.1) represent a continuous map on a compact, finite dimensional phase space $T: \Omega \to \Omega$, with metric $\rho(\cdot)$. Given an initial condition $z_0 \in \Omega$ and a $\delta > 0$, then there exists an $m \geq 0$ such that there are infinitely many times q, p where $q > p$, $q \geq m$ and such that the orbit of z_0 will recur with itself at these times to within δ. Hence, $\rho(T^q(z_0) - T^p(z_0)) < \delta$.

The proof of this lemma is quite simple, using the pigeon hole principle. Assume that the lemma is false. Recall that if we cover a compact set with δ balls $\{B_\delta(z)\}$, then we may take a finite subcover $\cup_{j=1}^{m} B_{\delta,j} \supseteq \Omega$. By assumption, the $T^i(z_0)$ must each lie in a distinct ball of the subcover. However, if $i = q - 1$ and $q \geq m$, then there are no balls left to accommodate another iterate; all the pigeon holes are filled. Hence the qth iterate must fall in an already occupied pth ball indicating a recurrence. We see that there must be infinitely many such instances if we shift $t = 0$ to $t = q$.

Now we consider what this implies in terms of minimum-time control. If z_a is in a hyperbolic set and if the recurrence distance δ between z_i and z_{i+s} is small enough, hyperbolicity implies that there exists a real orbit that converges to that of z_i backwards in time and converges to that of z_{i+s} forwards in time. Thus our original orbit could not have been time optimal since the loop $\{z_{i+1}, \ldots, z_{i+s-1}, z_{i+s}\}$ only serves to increase

$I(k)$. We can only determine it if a given recurrence is in fact close enough on a case-by-case basis by successfully constructing a patch of the shadow orbit. A technique to cut the loop and reglue it using an orbit segment is described in Section 3 below. Such a construction, when successful, monotonically decreases the value of I. Since the patched segment asymptotically converges to the original orbit, we can satisfy the constraint (2.3) as well.

The important point here is that it is very difficult to find fast orbits, or to know *a priori* when and how to apply a control sequence $\{k_i\}_{i=1}^{n}$. On the other hand, it is relatively easy to find slow orbits. However, slow orbits tend to waste time on long, sometimes very long, recurrent loops. These loops serve little more than to bide time until an appropriately aligned pass through the ball of recurrence has been achieved.

Slow orbits, while not useful in themselves, tell the story of how to find the switching points. Furthermore the orbit segments between recurrences which do not themselves recur are assumed to be locally optimal. This assumption relies on having correctly chosen the pre-assigned recurrence threshold δ so that all possible patches are glued. Choosing the threshold too large, however, wastes time checking "recurrences" that have no chance of being patched.

There are similar ideas to this in classical control literature. Dynamic programming, based on Bellman's principle of optimality, asserts that a globally optimal orbit (and it's associated control function) must also necessarily be locally optimal for segments of the orbit [6,19]. While local optimization does not in general imply global optimization, a huge improvement may nonetheless be achieved with a small amount of computer work, in the process of satisfying Bellman's necessary condition. For the two-dimensional case, if we correctly choose δ as the radius of a lobe, and each region only has one family of lobes, we believe that the restricted orbit found must be close to optimal.

In order to efficiently find the recurrences, we use the following algorithm. Given an orbit $\{z_0, \ldots, z_j, \ldots, z_N\}$ where $z_0 = z_a$ is close to a and $z_N = z_b$ is close to b, perform the following:

```
for i = 0 to N
   for j = N to i + 1, step −1
      if ρ(z_i − z_j) ≤ δ /*Is there a recurrent loop?*/
        then
        attempt to remove the loop /*Can a patch be found?*/
        if recurrent loop {z_{i+1}, ..., z_j} can be removed,
          then
          cut it and glue in patch
          let i = j
        end if
      end if
   loop
loop
```

This algorithm automatically considers the largest recurrences for removal first by working forward from z_0 and backwards from z_N. Shorter intermediate recurrences that occur inside of a bigger recurrence are automatically removed without ever being considered. This represents an improvement over a purely forward search which might require an ordering according to lengths of loops.

It may not always be necessary to have a single orbit from a to b to use this algorithm. Two regions of phase space can be explored separately by starting separate initial conditions and concatenating their resulting orbits. This can be a useful way to explore the phase space near a and b separately when a single orbit between them is particularly difficult to find. If the two orbits approach each other closely, then it might be possible to patch from one to the other. The above algorithm will automatically test all such possibilities. If, however, such a patch is not possible, then the end of the first orbit will be reached with no connection to the second orbit. In contrast, when a single orbit between the two regions can be found, the algorithm is robust, because the original, albeit slow orbit, is always available as the path.

The prerecorded orbit represents known information about transport in the visited phase

space. How we manage this information depends on our assigned task. If we are likely to be presented with a variety of initial conditions a and targets b, then the following model may be used. Ergodicity causes a long orbit to cover accessible phase space; the longer the orbit, the better the cover. Any target point close to the known orbit is feasible, and initial conditions close to the orbit allow for immediately starting stabilization. Alternatively, if no points of the known orbit are close enough to the initial condition, then uncontrolled iteration will quickly cause it to come close. This model requires directly stabilizing the initial condition to the known orbit, and restricting its length at points of recurrence, *on the fly*.

Stabilization can be performed by shooting the initial condition at the stable manifold of the known path using Newton's method to find the correct perturbations to the internal parameters. Details are discussed in [7] and represent only a slight modification to the gluing algorithm presented in the next section. Recurrences are detected and cut according to the above algorithm, where stabilization can immediately be used to skip a loop by shooting at the path after the loop. A successful cut is one in which the recurrence is close enough so that stabilization works, with a sufficiently small parameter perturbation.

Another possible control task is one where a and b are fixed in advance. We may be presented with such a model either when just a few objectives are likely, or perhaps a decision tree is to be built, and even the segments of the tree are difficult to find. In this case, time can be spent to find a more optimal solution achieving the transport. The resulting epsilon chain can be stabilized later, in real-time as above. This model tends to build faster orbits since the patches are built forwards and backwards from the switching points, as compared to the on-the-fly model described above where only the future points can be modified.

3. Cutting recurrent loops

Suppose that we find a recurrence between z_i and z_{i+s}, s steps later, i.e. $\|z_{i+s} - z_i\| < \delta$. Already, it is possible to skip $s-1$ iterates of the orbit by making the appropriate δ perturbation from z_i to z_{i+s}. Even better, it may be possible to find a patch consisting of a nearby orbit z'_i with the property that $\|z_{i-m} - z'_{i-m}\|$ and $\|z_{i+s+m} - z'_{i-m}\|$ are both small enough to satisfy the control constraint. If the orbit is hyperbolic, then we can remove the $(s-1)$-step loop using an exponentially smaller total perturbation. We find a patch $\{z'_{i-m}, \ldots, z'_{i+m}\}$ consisting of a nearby orbit segment, which is close to the pre-orbit of z_i before the recurrence, and close to the orbit of z_{i+s} after the recurrence, and which completely avoids the unwanted loop $\{z_{i+1}, \ldots, z_{i+s-1}, z_{i+s}\}$. We choose m so that the perturbation onto the patch from the natural orbit is as small as we require. The existence of such an orbit patch is guaranteed if the recurrent points are hyperbolic and the recurrence distance δ is small enough, but may be possible more generally. For the hyperbolic case, the size of δ depends on the geometry and angle of intersection between $W^{\mathrm{u}}(z_i)$ and $W^{\mathrm{s}}(z_{i+s})$. We will start by describing a point p on the patch which is between z_i and z_{i+s}. Then the rest of the patch is formed by forward and backward iteration. The point p lies on the intersection of the stable manifold of z_{i+s}, denoted $W^{\mathrm{s}}(z_{i+s})$, and the unstable manifold of z_i, similarly denoted as $W^{\mathrm{u}}(z_i)$, and therefore has the property that

$$\|T^n(p) - T^n(z_{i+s})\| \to 0 \quad \text{and}$$
$$\|T^{-n}(p) - T^{-n}(z_i)\| \to 0 \quad \text{as } n \to \infty . \tag{3.1}$$

By the proximity of z_i to z_{i+s}, p is within $h\delta$ of both z_i and z_{i+s}, where h is a constant that depends on the geometry of the intersection between $W^{\mathrm{s}}(z_{i+s})$ and $W^{\mathrm{u}}(z_i)$. To lowest order, h depends on θ, the angle of intersection between the local linear approximations to the manifolds. We expect that our technique will be less effective when θ is small, since the resulting triangle implies that p will be far from the point of recurrence, invalidating the locality assumptions. By construction, we expect that

$$\|T^m(z_{i+s}) - T^m(p)\| < h\lambda_s^m\delta \,, \tag{3.2a}$$

$$\|T^{-m}(z_i) - T^{-m}(p)\| < h\lambda_u^{-m}\delta \,, \tag{3.2b}$$

where $\lambda_s < 1$ is the local stable Lyapunov number at z_{i+s} and $\lambda_u > 1$ is the local unstable Lyapunov number at z_i. See Fig. 2.

In principle, it should be possible to choose m so that the perturbations from the original orbit onto the patch, and then back onto the original orbit, are as small as we like. However, finding points on the stable (unstable) directions numerically becomes increasingly ill conditioned if m is too large. In practice, finding the complete manifolds $W^s(z_{i+s})$ and $W^u(z_i)$ in order to find p is not practical, efficient, or even important. Instead we find p indirectly by making the approximation that $T^{-m}(p)$ lies in the tangent space of $W^u(z_{i-m})$, and likewise that $T^m(p)$ lies in the tangent space of $W^s(z_{i+s+m})$. Hence finding $T^{-m}(p)$ can be reduced to a problem of shooting.

In two dimensions we can parameterize the unstable direction with the vector $f_u(z_{i-m})$, a unit vector in the tangent space of $W^u(z_{i-m})$, by the variable t. An initial condition is then chosen,

$$z_0(t) = z_{i-m} + tf_u(z_{i-m}) \,. \tag{3.3}$$

The success of an initial condition can be measured by how closely $T^{2m}(z_0(t))$ lands on the line $z_{i+s+m} + \tau f_s(z_{i+s+m})$. We write components of the vectors $z(t) = (x(t), y(t))$ and $f_s = (f_{s,x}, f_{s,y})$. The roots of the expression

$$F(t) = f_{s,y}(x(t) - x_{i+s+m}) - f_{s,x}(y(t) - y_{i+s+m}) = 0 \tag{3.4}$$

can be found quickly using a Newton-secant method. We need only make an appropriate initial guess so that the point we find will in fact be a principal intersection point. As a rough guess, we can use Eq. (3.2) to write

$$t_0 = \lambda_u^{-m}\delta \,, \tag{3.5}$$

where we have assumed that $h \approx 1$. The extension to higher dimensions is straightforward. The number of variables needed to parameterize the initial condition must equal the dimension spanned by the unstable subspace of the tangent space at z_{i-m}. Likewise, examining the related problem of shooting at the stable manifold using the parameter k provides the controllability

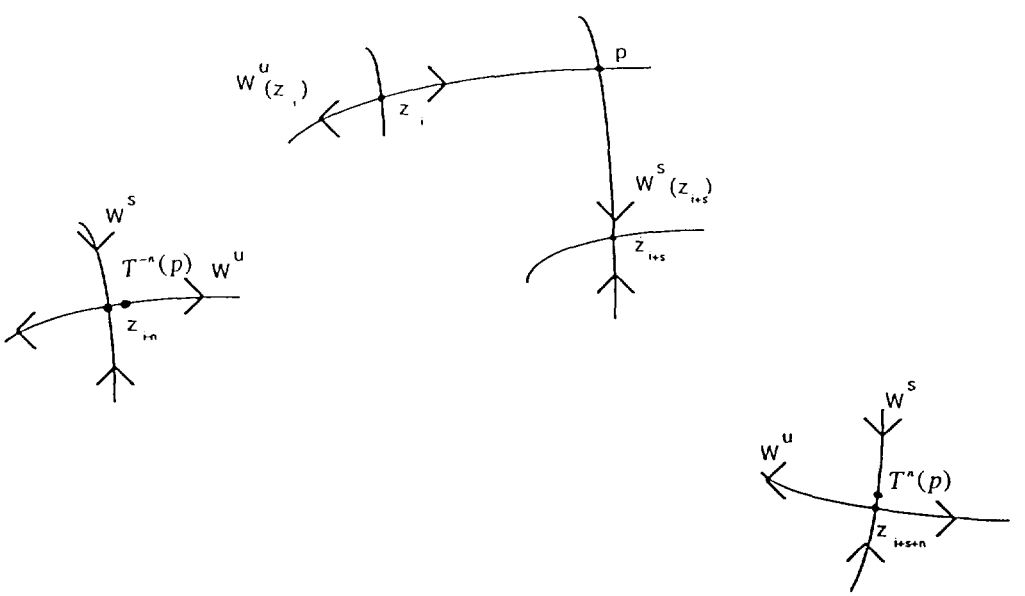

Fig. 2. Construction of a patch. When the point z_i recurs with z_{i+s}, the point of principal intersection p between the unstable manifold of z_i and the stable manifold of z_{i+s} converges to the orbit of z_{i+s} under applications of the map, and converges to the preorbit of z_i under applications of the inverse map.

condition that perturbations to k must span the unstable subspace of the tangent space at z_{i+s+m}.

Finding the stable direction f_s and the unstable direction f_u at a point z from a chaotic set first requires the complete orbit $\{\ldots, z_{-n}, \ldots, z_{-1}, z_0, z_1, \ldots, z_n, \ldots\}$. Recall that the Jacobian matrix rotates a vector in the tangent space towards the unstable direction, and the Jacobian matrix of the inverse map T^{-1} rotates a vector towards the stable direction. Therefore, in practice, we choose an arbitrary unit vector $\boldsymbol{u}$ and forward multiply, starting at z_{-n}, the Jacobian matrices along the orbit to z, normalizing the vector at each step:

$$DT^n|_{z_{-n}} \cdot \boldsymbol{u} \equiv DT|_{z_{-1}} \cdot DT|_{z_{-2}} \cdot \ldots \cdot DT|_{z_{-n}} \cdot \boldsymbol{u} \to f_u(z) \quad \text{as } n \to \infty . \tag{3.6}$$

Likewise, the stable direction is formed from the inverse Jacobian starting at $T^n(z)$. Convergence is exponential; in practice we find that $n = 20$ gives an error of 10^{-5}. We use $n = 40$ which we expect is more than adequate considering the scale of the other errors, see [8]. At the same time as the above calculation, we calculate the corresponding Lyapunov multipliers, also by the power method.

All of the required quantities for cutting and gluing are in fact accessible to a model of a dynamics formed by time-series embedding. The primary piece of information, a recurrence, requires no modeling to identify. To form the patch, however, we need to fit a piecewise model of the data in order that predictions may be formulated between known data points [4]. In such a case, a more accessible numerical technique to form the unstable and stable directions at a point z is to consider the recorded histories of nearby clusters of points in forward and backward time. Likewise, from $T^{-n}(z)$, nearby points orient along the unstable axis in forward time at z. In addition, partial derivatives of the map T in each of the parameter directions may be approximated by interpolating between three models of separate data sets which bracket the range of each parameter.

Note that further refinement to the epsilon chain can be achieved by running a second pass of the gluing algorithm, by treating chain errors as the δ and finding a patch over it to further reduce the error by a factor of λ^m. Hence a smaller error epsilon chain can be achieved with a modest m by redistributing the points of the error to the ends of the new patch.

4. The standard map, an example

4.1. Area preserving transport

We now demonstrate our method for the case of the much studied standard map. The standard map, also known as the kicked rotor, is an area preserving, twist map of the plane

$$z' = \begin{pmatrix} y' \\ x' \end{pmatrix} = \begin{pmatrix} y - \dfrac{k}{2\pi}\sin(2\pi x) \\ y - \dfrac{k}{2\pi}\sin(2\pi x) + x \end{pmatrix}. \tag{4.1}$$

The phase space structures and transport characteristics are typical of two degree of freedom Hamiltonian systems. There is periodicity in both x and y with period 1, so the phase space is the torus. Hence according to the recurrence lemma of Section 2 every orbit must eventually recur.

As a concrete example, we investigate the transport from a neighborhood of the (0,1) hyperbolic point of (4.1), to a neighborhood of the (1,1) resonance. The notation (p, q) denotes the frequency of an orbit, i.e. q iterations of the map results in exactly p wraps around the cylinder: $T^q(z) = z + p$. The starting point (0,1) a is located at $(x_a, y_a) = (0.5, 0.0)$, and (1, 1), our target point b, at $(x_b, y_b) = (0.5, 1.0)$.

It is not possible to find such an orbit if $k < k_c \approx 0.971\,635\,406\,31\ldots$ [11] which is the parameter value at which the last invariant curve dividing the phase space between (0,1) and (1,1) becomes a cantorus. The most robust curves between (0,1) and (1,1) are the circles with

rotation frequencies $1/\gamma$ and $1/\gamma^2$, where $\gamma = (1+\sqrt{5})/2$, is the golden mean. For $k > k_c$ these become cantori and they have the smallest lobe areas, and hence, represent the most difficult barriers to transport. An arbitrary orbit will typically pass through the lobes many times before finally reaching the target point b.

This effect also occurs when an orbit is trapped near an island, and near islands around islands and so on. This phenomenon has been successfully modeled using Markov trees [5,12]. It was found that a point initially "near" a KAM surface has a survival probability $F(t)$, asymptotic to $t^{-\alpha}$, that the orbit will still be near the surface at large time t, with small constant α [12]. Therefore in the presence of KAM surfaces, we find long correlations and, hence, roughly a power-law decay. However the important point is that, without knowing where the lobes are located, recurrence is a way of locally detecting globally inefficient orbits.

4.2. Transport time distributions

Before we demonstrate the improvements made by cutting recurrences, we first investigate the natural transport time distributions for a range of parameter values. We performed a Monte Carlo study within a box of size 0.1 centered around our starting point a at $(-0.5,0.0)$, from which we randomly chose an ensemble of 10^4 initial conditions, and binned them according to how long they took to arrive in a similar box around the target b at $(-0.5,1.0)$. Two of the resulting histograms are displayed together in Fig. 3 for the parameter value $k = 1.01$, just above k_c, and the moderately high value $k = 1.25$. We find results similar to the experiment performed by Chirikov [1] who observed that the transport time from $y \approx 0$ to $y \approx 0.5$ obeyed a power law. Fig. 4 displays the average transport time on a log-log plot versus $(k - k_c)$, demonstrating that the average crossing time is indeed well approximated by the singular power law $(k - k_c)^{-\eta}$ [1,3]. It is possible that

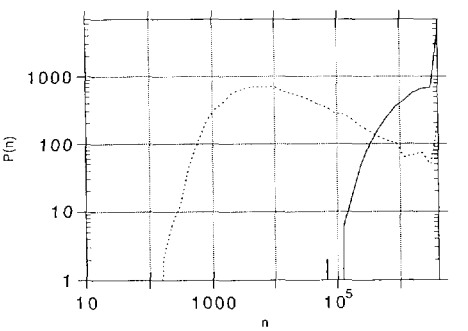

Fig. 3. Histograms of transport time for the standard map between the box of side of 0.1 centered on the starting point a at (0.5,0.0) to a similar box centered on the target point b at (0.5,1.0). 10^4 initial conditions are randomly chosen from the a box and binned according to time to transport into the b box. The number of bins alloted is 100. The solid curve is for $k = 1.01$, and the dashed curve is for $k = 1.25$. The maximum iterations performed on an initial condition is 5×10^6 before cutoff. Points requiring more iterations are found in the last bin.

many of the initial conditions are chosen in the subset of the box that is inaccessible to z_b. These, and other initial conditions have such long transport times, that we choose a cutoff time of $5 \cdot 10^6$ iterates, which is increasingly a problem for k approaching k_c. This causes the value of η that we calculate to be somewhat less than the value of 3.012 predicted [11]. The variance of the distribution is large, as seen in Fig. 3, but impossible to measure due to the

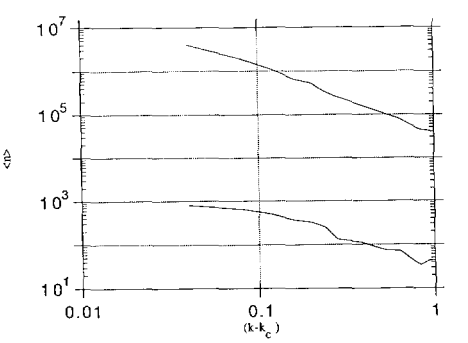

Fig. 4. Transport times from the a box to the b box as a function of $(k - k_c)$. The top curves shows the average time, calculated amongst 10^4 randomly chosen initial conditions from the a box iterated until first intersection with the b box. The bottom curve shows optimized transport times from a to b resulting from cutting recurrent loops from slower orbits.

large fraction of the box with transport times in the last bin, for any reasonable finite cut-off. Indeed, we observed that the transport rate is extremely sensitive to the initial condition, so the average transport rate is not an indicative measure of a "typical" rate. Fig. 5 displays the percentage of initial conditions from the box that never reach b. Here we find that as k approaches k_c, most of the transport times are actually larger than our cutoff. Thus our computed statistics are only lower bounds on the actual transport statistics. Nonetheless, the point is that the transport is slow as $k \to k_c$, indicating that our efforts to find faster orbits are worthwhile.

4.3. Cutting and gluing slow orbits

From the same box around the initial point a described above, we choose an initial condition which iterates *eventually* to near the target b. We restrict ourselves to the orbits of randomly chosen initial conditions that perform the transport required in less than 10^6 iterations so that we do not strain the memory capacity of our computing resources by storing uninteresting information.

As a concrete example, consider a 80307-step orbit for $k = 1.25$. Recurrences are sought following Section 2 searching for z_i from the start of the orbit, and the last recurrence z_{i+s} (largest value of s) from the end of the orbit. A certain amount of space must be reserved in order to fit the patches. We used patches consisting of $2m + 1 = 31$ steps, so $m = 15$ steps must be allowed for on either side of the recurrence in order that the error may have time to contract sufficiently such that the constraint (2.3) is satisfied. The rate of contraction is determined by the Lyapunov exponents, according to eqs. (3.2a), (3.2b). Therefore a strict lower bound on the cost function I for our technique is $(2m + 1) + 1$, the space required for one patch. If patches are forced to not overlap, then q recurrences implies that I is bounded by $q(2m + 1) + 1$. In principle, arbitrarily small constraints (2.3) can be met, but in practice, solutions become numerically ill conditioned as m gets large. We chose a modest value $m = 15$ for this example, although $m = 25$ and $m = 30$ were successfully tested.

The first recurrence that we can successfully remove from our 80307-step orbit is between z_{16} and z_{78704} which recurs to a distance $\delta = 0.08$. The cut and glue algorithm allows us to construct an orbit patch $\{z'_1, \ldots, z'_{31}\}$ such that the error to perturb on to the orbit patch is only $\|z'_1 - z_1\| = 0.002$, and the error to perturb back off of the orbit patch is $\|z'_{31} - z_{78720}\| = 0.002$. Fig. 6, showing the error between the patch orbit segment and the two ends of the recurrence on the orbit, displays how hyperbolicity is used to

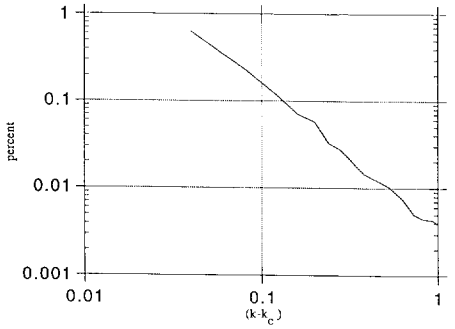

Fig. 5. The percentage of the 10^4 initial conditions from the a box which take longer than 5×10^6 iterations to reach the b box.

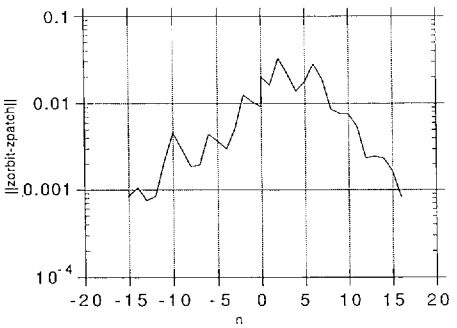

Fig. 6. The error between a patch orbit $\{z_n\}_{n=-15}^{15}$ and an original orbit at each of the points z_{i+s+n} for $n \geq 0$, and z_{i+n} for $n \leq 0$.

diminish the recurrence error. The slopes of the decaying error on either side of the recurrence represent the corresponding stable and unstable Lyapunov exponents.

With this single patch, we have already demonstrated a 1619-step epsilon chain orbit near our original orbit. By finding every recurrence within a threshold $\delta = 0.1$ and cutting those that can be patched within the error $\varepsilon = 0.005$ we eventually construct a 131-step epsilon chain orbit including 13 overlapping patches. The largest error found in this example was $\|T_{k_0}(z_{92}) - z_{93}\| = 0.003$, but there were several others of the same order. For this case, 13 important switching points in 13 important lobes are inferred, and orbit segments between the switching points are automatically found by keeping those segments that do not recur close enough to be further cut. To further demonstrate manageability of the errors, we ran a second pass of the patching algorithm over the error points of the newly formed epsilon chain. Redistribution of the errors yields a further reduction by a factor of 50 to 1000.

A phase space portrait of the final path is displayed in Fig. 1. The black regions in the plot represent inaccessible regions of phase space, corresponding to elliptic islands with rational frequencies between 0 and 1. The 131-step epsilon chain orbit, marked by black circles, manages to transport through all the resonances without getting caught in their periodicity, even though orbits trapped in a given resonance layer must move at the frequency. The efficiency of transport can be observed by the lack of getting bogged down in almost periodicities, as revealed by the lack of corresponding recurrences.

Several different runs with $k = 1.25$ for various initial conditions ranging up to a length of 10^6, and various recurrence thresholds from $\delta = 0.02$ to 0.07 yield epsilon chain orbits of length $n =$ 131 to 251. There is a trade-off when choosing the recurrence threshold value δ. Recall that the rate at which a recurrence error δ can be reduced to the tolerance ε is governed by the inequalities (3.2) which we adapt here to require that

$$\varepsilon < h\lambda_{s,u}^{\pm m}\delta \,, \tag{4.2}$$

where we choose $m = 15$. Close recurrences are more likely to be cut. The resulting t and τ, the perturbations along the vectors $f_u(z_{i-m})$ and $f_s(z_{i+s+m})$ both called ε, are small according to (4.2) for fixed m. In addition, it is often possible for (3.4) to be solved even if t and τ are large, but the linearization of the stable (unstable) manifolds is not valid, and so a contraction according to (4.2) is not expected. A successful cut is one where a given tolerance ε can be satisfied. Since zeroing (3.4) is relatively cheap, it is practical to set δ, the recurrence testing threshold, to a relatively high value where most are not successfully cut, but no opportunities are missed.

Table 1 shows that higher δ can yield faster orbits, but at a cost of many more trials. We observe in fact that $\delta \leq 0.1$ and $\delta \leq 0.2$ yield the same paths; considering $\delta \leq 0.2$ recurrences yields no extra successful patches. This is reflected in the success rate column of Table 1. All

Table 1
Various recurrence thresholds

δ	n	Loops cut	Success rate (%)
0.005	748	15	100.0
0.01	597	13	3.96
0.025	236	8	2.020
0.05	177	14	1.552
0.08	173	14	1.14
0.09	156	16	1.16
0.1	156	16	0.920
0.2	156	16	0.096
1.0	156	16	0.035

Values of δ are tested on a single 109 594-step orbit between the a box and b box for $k = 1.25$ and patch size $m = 15$ to achieve an $\varepsilon = 0.005$ tolerance. This shows how increasing computer work, to a point, yields faster paths by considering unlikely patchable recurrences. Tabulated quantities are: the threshold tested δ, the resulting epsilon chain length n, the number of loops successfully cut and the percentage ratio of successfully cut loops to those attempted when a δ recurrence was detected.

increases in δ up to 0.09 did in fact yield faster paths. By contrast, if we allow ourselves to use longer patches by increasing m, to allow more space to contract, in principle, we expect that higher values of δ are likely to be successful, but at the cost of ill conditioning the solution of (3.4). We find a good balance at $m = 15$, but the choice is arbitrary. We note that the solution of the long time targeting problem would be trivial if there were no ill-conditioning problem with growing m, since we could simply choose $\delta = 1$. There would be no need to consider intermediate switching points, the location of which is the source of the major difficulty to slow transport problems. In that case a would be recurrent with b and by choosing a very large value of m (such as $m = 65$ as indicated by our 131-step orbit above) we could shoot directly from beginning to end.

To find the optimum value it is best to choose δ equal to the "diameter" of the important lobes between a and b, in their most round iteration as they become long and thin in both forward and backward time. Define the most round iteration of a lobe as the iteration with minimum diameter, where the diameter is defined by the supremum of the distance between any two points in the set. Unfortunately, in general we do not know this value *a priori*. An efficient technique would involve several passes for increasing values of δ, first forming an epsilon chain orbit that always stays below control saturation (2.3) and then removing any possible recurrences for the next value of δ. This would represent a multi-pass algorithm since the length of the epsilon chain orbit decreases monotonically.

The lower curve of Fig. 4 displays "optimized" transport times as a function of $(k - k_c)$, for a one-pass optimization. These values can be compared to the uncontrolled transport time averages also displayed in the same figure. We see an improvement by a factor of almost 10^4 on average for the lower values of k.

4.4. Hyperbolicity in the standard map

The construction of an orbit patch as explained in Section 3 is guaranteed to work when the orbit is a hyperbolic saddle. A saddle is defined as a compact, non-attracting, invariant set with a dense orbit such that each point of the set has a stable direction and an unstable direction. A hyperbolic saddle has all the angles between stable and unstable manifolds bounded away from zero, and a non-hyperbolic saddle may have angles that approach zero. In order that the intersection point p is close to the δ ball containing the recurrent points, the constant h, depending on the angle between a stable manifold at z_{i+s} and an unstable manifold at z_i, must be small. In this section, we discuss the validity of this assumption by numerically investigating the angle distribution between stable and unstable manifolds calculated at each point along a slow orbit between the objectives a and b. This is in the spirit of the recent paper by Lai et al. [8] who perform similar calculations for the Henon map.

Figs. 7 show angle distributions for $k = 1.01$ and 1.25. The distributions found were similar in shape, average and peak for each value k, independent of the initial condition. In Figs. 7a and 7c we can see a definite spike at the peak in the angle distributions. From there the probability falls off smoothly with increasing angle. it is not clear whether the angles are bounded away from zero or not, since the curve falls off smoothly, seemingly towards zero. This situation for the standard map is qualitatively quite different from that of the Henon map, where the distributions show a "complicated structure without much regularity near $\theta = 0°$". The lowest angle for the figures displayed is $\theta = 0.0009°$ for Fig. 7a and $\theta = 0.12°$ for Fig. 7c. The results are similar when other transporting initial conditions from the a box are chosen.

The point that concerns us here is that the probability of finding angles below any reason-

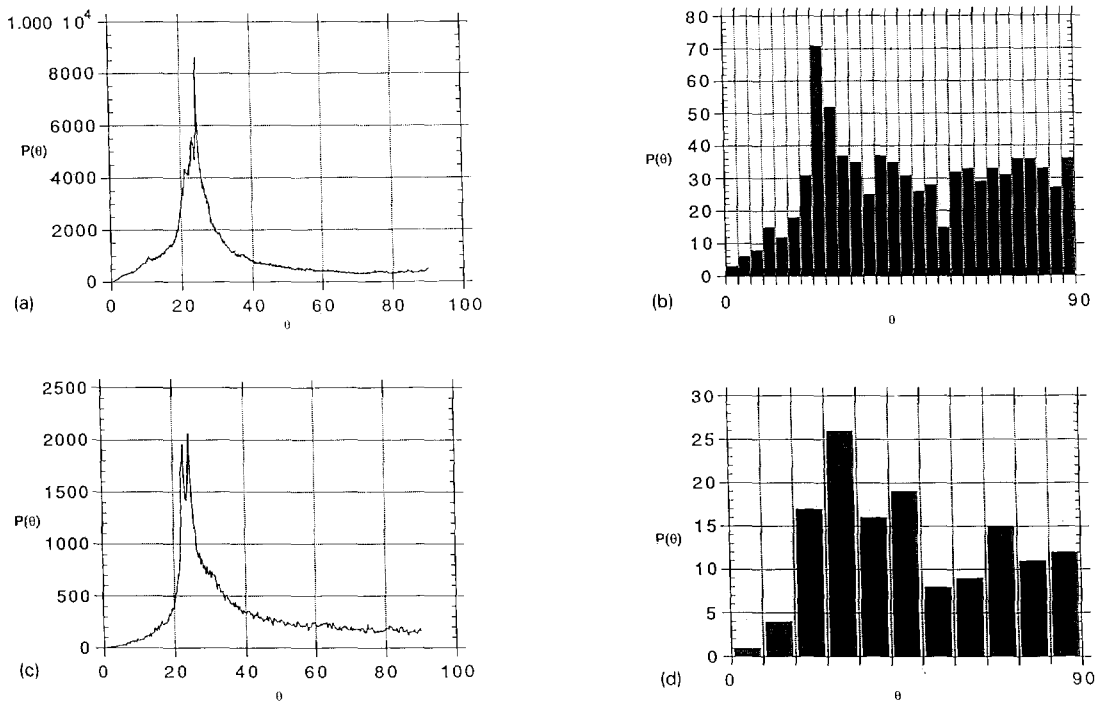

Fig. 7. Histograms of angles between stable and unstable manifolds. In (a) the distribution shows angles for each of the 975 760 points along an uncontrolled orbit between the a box and b box for $k = 1.01$. In (c) the angles along an uncontrolled orbit for $k = 1.25$ are shown. In (b) the 975 760-step orbit for $k = 1.01$ has been restricted to a 778-step epsilon chain. Similarly in (d) for $k = 1.25$, a 80307-steps orbit has been restricted to a 131-step epsilon chain.

able value is quite low. According to our algorithm, a recurrence is only cut when an orbit patch within the control tolerance can be found on a trial and error basis. So we are only concerned here with the probability of small angles: for example in Fig. 7c $P(\theta < 8°) \cong x = 0.7\%$.

It is interesting to compare the angles between $f_u(z_i)$ and $f_s(z_{i+s})$ of recurrences that cannot be successfully mended to those that can be mended. For $k = 1.25$ and $\delta = 0.01$ we recorded separately the angles of successes and failures. The failure category includes the entire range of angles. We expect a problem with small angles, but even large angles can be a problem for recurrences that are not close enough or when the manifolds curve sharply away from the linear approximations. In contrast, the smallest successfully mended angle found was 11°, and the angles tend to be much higher than that in general.

Figs. 7b and 7d show the angle distributions for the mended epsilon chain orbits. The angles are now calculated using the epsilon chain rather than following the natural orbit of the point z which may quickly diverge from the predicted pseudo-orbit. The main feature we observe in the restricted orbits is that the average angle increases invariably. For example, the average for Fig. 7a, $\langle\theta\rangle = 29.9°$ was increased to $\langle\theta\rangle = 50.9°$ along the 778-step epsilon chain. Similarly, Fig. 7d reveals a change from $\langle\theta\rangle = 40.5°$ to $\langle\theta\rangle = 46.9°$ along the 131-step path.

4.5. Stabilization

We next demonstrate that an initial condition can indeed be stabilized onto the epsilon chain

with smaller parameter perturbations as described in Section 2. As an illustration consider the same orbit as in Section 4.3 for which we found a nearby 131-step pseudo-orbit $\{z_1, \ldots, z_{131}\}$ with 13 patches. The parameter perturbation size required for stabilization depends directly on the phase space error found.

We demonstrated that initial conditions close to z_1 can be stabilized along the epsilon chain to z_{131}. Parameter perturbations were used to shoot at the stable manifold further down the path, and were calculated whenever a point of error was predicted on the epsilon chain, or the test point drifted outside a pre-determined tolerance of the known path. For this particular example, a maximum value $\Delta k = 0.016$ was required, but most perturbations were several orders smaller. Stabilization was successful for all the epsilon chains tested, for various values of k_0.

For comparison, we also tested on-the-fly stabilization directly to the long 80307-step orbit. Again, we used a randomly chosen initial condition near the known orbit, and the stabilization was turned on whenever the test orbit drifted outside a set tolerance. In addition, whenever a recurrence in the prerecorded orbit was detected along the way, stabilization was attempted by shooting at the end of the loop. We expected that the length of the test orbit would be longer, because with this method, only future events can be modified. In this example, the test orbit achieved the final destination in 464 iterations. In spite of being typically slower, on-the-fly control can be more flexible since one can rapidly retarget as needed, as the whole prerecorded orbit is available.

5. Conclusions

We have discussed the time-optimal control problem for chaotic regions, giving a method to find paths that quickly achieve transport goals and which can be stabilized with small parameter perturbations. A recurrent orbit necessarily violates Bellman's condition for optimality; we eliminated recurrences by using them as switching points between orbit segments. The resulting epsilon chain was refined by smoothing with a patch that shoots from the unstable manifold of the orbit before the recurrence onto the stable manifold of the orbit after the recurrence. The method was demonstrated on the standard map for which uncontrolled transport is extremely slow and the statistics of the transit times are anomalous. The technique, when applied to orbits of up to 10^6 iterates, typically reduces transport times by factors of up to 10^4, even for k close to k_c where previous targeting schemes are unsuccessful [9]. Finally, we demonstrated that small parameter perturbations can be used to stabilize a test orbit onto either a precalculated fast epsilon chain, or alternatively onto a long known orbit, eliminating recurrences on the fly.

The techniques in this paper can be straightforwardly extended in a number of directions. Though we used a Hamiltonian mapping as an example, the method makes no assumptions as to the nature of the dynamics – indeed since the inverse map is not needed, the dynamics can even be noninvertible. The use of recurrences as switching points requires no assumptions on the dimensionality of phase space, though recurrences will be less common in higher dimensions. As well as optimizing the transport between two separated points, the method could also be used to eliminate escape from a region, using the recurrences to construct a rough periodic orbit. This might require a smaller and less frequent control than stabilizing a fixed point in a given region. Finally, the techniques presented in this paper are also applicable to a piecewise local model of a time-series built through embedding. Our current research is aimed precisely at realizing these assertions to control a nonanalytic representation of chaotic dynamics in more than two dimensions.

Acknowledgement

This work was partially supported by NSF grant DMS-9305847, an NSF Graduate Research Traineeship, DMS-9208685, and a NATO collaborative research grant CRG 921181. We would like to thank the Mathematics Research Centre, and particularly Robert MacKay, at the University of Warwick for their hospitality for the fall of 1993. We would also like to thank Mark Muldoon for many helpful discussions.

References

[1] B.V. Chirikov, A universal instability of many dimensional oscillator systems, Phys. Rep. 52 (1979) 265.
[2] R. Bowen, On axiom A diffeomorphisms, J. Diff. Eqs. 18 (1975) 333.
[3] I. Dana and S. Fishman, Diffusion in the standard map, Physica D 17 (1985) 63.
[4] J. Farmer and J. Sidorowich, Predicting chaotic time series, Phys. Rev. Lett. 59 (1987) 845.
[5] J.D. Hanson, J.R. Cary and J.D. Meiss, Algebraic decay in self-similar Markov chains, J. Stat. Phys. 39 (1985) 327.
[6] A.V. Hsu and J.C. Meyer, Modern Control Principles and Applications (McGraw-Hill, New York, 1968).
[7] E. Kostelich, C. Grebogi, E. Ott and J.A. Yorke, Higher-dimensional targeting, Phys. Rev. E 47 (1993) 305.
[8] Y.-C. Lai, C. Grebogi, J.A. Yorke and I. Kan, How often are chaotic saddles nonhyperbolic, Nonlin. 6 (1993) 779.
[9] Y.-C. Lai, M. Dong and C. Grebogi, Controlling hamiltonian chaos, Phys. Rev. E 47 (1993) 86.
[10] R.S. MacKay, J.D. Meiss and I.C. Percival, Transport in Hamiltonian systems, Physica D 13 (1984) 55–81.
[11] J.D. Meiss, Symplectic maps, variational principles, and transport, Rev. Mod. Phys. 64 (1992) 795.
[12] J.D. Meiss and E. Ott, Markov tree model of transport in area preserving maps, Physica D 20 (1986) 387.
[13] E. Ott, C. Grebogi and J.A. Yorke, Controlling chaotic dynamical systems, in: Chaos: Soviet-American Perspective on Nonlinear Science (American Institute of Physics, New York, 1990) p. 153.
[14] F. Romeiras, C. Grebogi, E. Ott and W.P. Dayawansa, Controlling chaotic dynamical systems, Physica D 58 (1992) 165.
[15] T. Shinbrot, E. Ott, C. Grebogi and J.A. Yorke, Using chaos to direct orbits to targets in systems describable by a one dimensional map, Phys. Rev. A 45 (1992) 4165.
[16] T. Shinbrot, E. Ott, C. Grebogi and J.A. Yorke, Using chaos to direct trajectories to targets, Phys. Rev. Lett. 65 (1990) 3215.
[17] I. Starobinets and A. Pikovsky, Multistep method for controlling chaos, Phys. Lett. A 181 (1993) 149.
[18] H. Swinnerton-Dyer and C. Sparrow, The Falkner–Skan equation 1: The creation of strange invariant sets, DAMTP Cambridge report (1993).
[19] J. Tou, Modern Control Theory (McGraw Hill, New York, 1964) p. 267.
[20] S. Wiggins, Chaotic Transport in Dynamical Systems (Springer, New York, 1992).