INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS

SMR: 960/2

WORKSHOP ON NONLINEAR CONTROL AND CONTROL OF CHAOS

(17 - 28 June 1996)

"Control of Chaos"

presented by:

Celso Grebogi
University of Maryland
College Park
Maryland 20742
U.S.A.

AMERICAN
ASSOCIATION FOR THE
ADVANCEMENT OF
SCIENCE

SCIENCE

30 OCTOBER 1987 $2.50
VOL. 238 ■ PAGES 585–718

1

Chaos, Strange Attractors, and Fractal Basin Boundaries in Nonlinear Dynamics

CELSO GREBOGI, EDWARD OTT, JAMES A. YORKE

Recently research has shown that many simple nonlinear deterministic systems can behave in an apparently unpredictable and chaotic manner. This realization has broad implications for many fields of science. Basic developments in the field of chaotic dynamics of dissipative systems are reviewed in this article. Topics covered include strange attractors, how chaos comes about with variation of a system parameter, universality, fractal basin boundaries and their effect on predictability, and applications to physical systems.

IN THIS ARTICLE WE PRESENT A REVIEW OF THE FIELD OF chaotic dynamics of dissipative systems including recent developments. The existence of chaotic dynamics has been discussed in the mathematical literature for many decades with important contributions by Poincaré, Birkhoff, Cartwright and Littlewood, Levinson, Smale, and Kolmogorov and his students, among others. Nevertheless, it is only recently that the wide-ranging impact of chaos has been recognized. Consequently, the field is now undergoing explosive growth, and many applications have been made across a broad spectrum of scientific disciplines—ecology, economics, physics, chemistry, engineering, fluid mechanics, to name several. Specific examples of chaotic time dependence include convection of a fluid heated from below, simple models for the yearly variation of insect populations, stirred chemical reactor systems, and the determination of limits on the length of reliable weather forecasting. It is our belief that the number of these applications will continue to grow.

We start with some basic definitions of terms used in the rest of the article.

Dissipative system. In Hamiltonian (conservative) systems such as arise in Newtonian mechanics of particles (without friction), phase space volumes are preserved by the time evolution. (The phase space is the space of variables that specify the state of the system.) Consider, for example, a two-dimensional phase space (q, p), where q denotes a position variable and p a momentum variable. Hamilton's equations of motion take the set of initial conditions at time $t = t_0$ and evolve them in time to the set at time $t = t_1$. Although the shapes of the sets are different, their areas are the same. By a dissipative system we mean one that does not have this property (and cannot be made to have this property by a change of variables). Areas should typically decrease (dissipate) in time so that the area of the final set would be less than the area of the initial set. As a consequence of this, dissipative systems typically are characterized by the presence of attractors.

Attractor. If one considers a system and its phase space, then the initial conditions may be attracted to some subset of the phase space (the attractor) as time $t \to \infty$. For example, for a damped harmonic oscillator (Fig. 1a) the attractor is the point at rest (in this case the origin). For a periodically driven oscillator in its limit cycle the limit set is a closed curve in the phase space (Fig. 1b).

Strange attractor. In the above two examples, the attractors were a point (Fig. 1a), which is a set of dimension zero, and a closed curve (Fig. 1b), which is a set of dimension one. For many other attractors the attracting set can be much more irregular (some would say pathological) and, in fact, can have a dimension that is not an integer. Such sets have been called "fractal" and, when they are attractors, they are called strange attractors. [For a more precise definition see (*1*).] The existence of a strange attractor in a physically interesting model was first demonstrated by Lorenz (*2*).

Dimension. There are many definitions of the dimension d (*3*). The simplest is called the box-counting or capacity dimension and is defined as follows:

$$d = \lim_{\epsilon \to 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)} \tag{1}$$

where we imagine the attracting set in the phase space to be covered by small D-dimensional cubes of edge length ϵ, with D the dimension of the phase space. $N(\epsilon)$ is the minimum number of such cubes needed to cover the set. For example, for a point attractor (Fig. 1a), $N(\epsilon) = 1$ independent of ϵ, and Eq. 1 yields $d = 0$ (as it should). For a limit cycle attractor, as in Fig. 1b, we have that $N(\epsilon) \sim \ell/\epsilon$, where ℓ is the length of the closed curve in the figure (dotted line); hence, for this case, $d = 1$, by Eq. 1. A less trivial example is illustrated in Fig. 2, in the form of a Cantor set. This set is

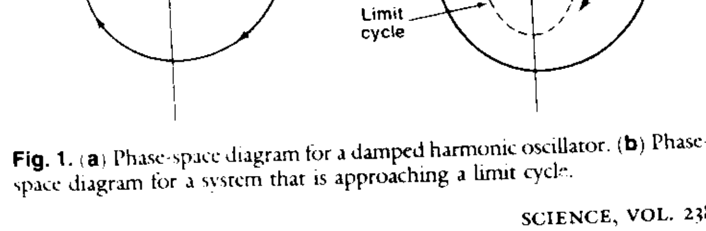


Fig. 1. (**a**) Phase-space diagram for a damped harmonic oscillator. (**b**) Phase-space diagram for a system that is approaching a limit cycle.

C. Grebogi is a research scientist at the Laboratory for Plasma and Fusion Energy Studies, E. Ott is a professor in the departments of electrical engineering and physics, and J. A. Yorke is a professor of mathematics and is the director of the Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742.

2

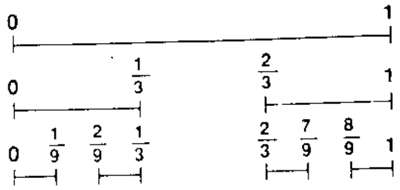


Fig. 2. Construction of a Cantor set.

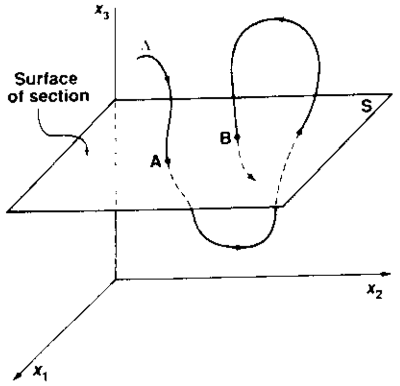


Fig. 3. Poincaré surface of section.

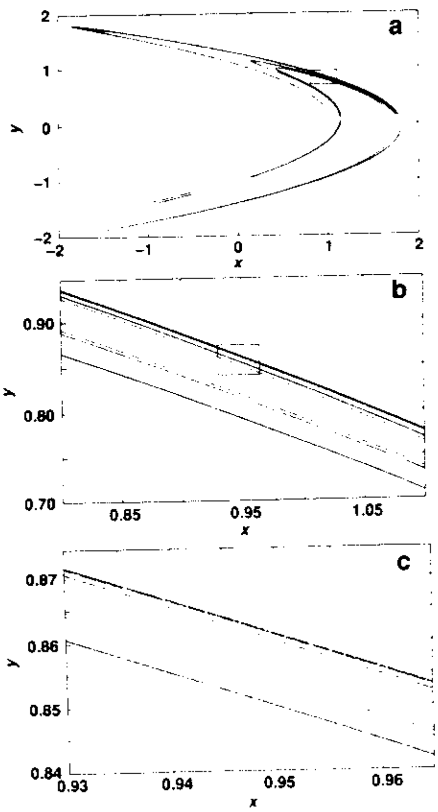


Fig. 4. The Hénon chaotic attractor. (**a**) Full set. (**b**) Enlargement of region defined by the rectangle in (a). (**c**) Enlargement of region defined by the rectangle in (b).

formed by taking the line interval from 0 to 1, dividing it in thirds, then discarding the middle third, then dividing the two remaining thirds into thirds and discarding their middle thirds, and so on ad infinitum. The Cantor set is the closed set of points that are left in the limit of this repeated process. If we take $\epsilon = 3^{-n}$ with n an integer, then we see that $N(\epsilon) = 2^n$ and Eq. 1 (in which $\epsilon \rightarrow 0$ corresponds to $n \rightarrow \infty$) yields $d = (\ln 2)/(\ln 3)$, a number between 0 and 1, hence, a fractal. The topic of the dimension of strange attractors is a large subject on which much research has been done. One of the most interesting aspects concerning dimension arises from the fact that the distribution of points on a chaotic attractor can be nonuniform in a very singular way. In particular, there can be an arbitrarily fine-scaled interwoven structure of regions where orbit trajectories are dense and sparse. Such attractors have been called multifractals and can be characterized by subsidiary quantities that essentially give the dimensions of the dense and sparse regions of the attractor. In this review we shall not attempt to survey this work. Several papers provide an introduction to recent work on the dimension of chaotic attractors (*3–5*).

Chaotic attractor. By this term we mean that if we take two typical points on the attractor that are separated from each other by a small distance $\Delta(0)$ at $t = 0$, then for increasing t they move apart exponentially fast. That is, in some average sense $\Delta(t) \sim \Delta(0)\exp(ht)$ with $h > 0$ (where h is called the Lyapunov exponent). Thus a small uncertainty in the initial state of the system rapidly leads to inability to forecast its future. [It is not surprising, therefore, that the pioneering work of Lorenz (*2*) was in the context of meteorology.] It is typically the case that strange attractors are also chaotic [although this is not always so; see (*1, 6*)].

Dynamical system. This is a system of equations that allows one, in principle, to predict the future given the past. One example is a system of first-order ordinary differential equations in time, $d\mathbf{x}(t)/dt = \mathbf{G}(\mathbf{x},t)$, where $\mathbf{x}(t)$ is a D-dimensional vector and $\mathbf{G}$ is a D-dimensional vector function of $\mathbf{x}$ and t. Another example is a map.

Map. A map is an equation of the form $\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$, where the "time" t is discrete and integer valued. Thus, given $\mathbf{x}_0$, the map gives $\mathbf{x}_1$. Given $\mathbf{x}_1$, the map gives $\mathbf{x}_2$, and so on. Maps can arise in continuous time physical systems in the form of a Poincaré surface of section. Figure 3 illustrates this. The plane x_3 = constant is the surface of section (S in the figure), and A denotes a trajectory of the system. Every time A pierces S going downward (as at points A and B in the figure), we record the coordinates (x_1,x_2). Clearly the coordinates of A uniquely determine those of B. Thus there exists a map, B = F(A), and this map (if we knew it) could be iterated to find all subsequent piercings of S.

Chaotic Attractors

As an example of a strange attractor consider the map first studied by Hénon (*7*):

$$x_{n+1} = \alpha - x_n^2 + \beta y_n \quad (2)$$

$$y_{n+1} = x_n \quad (3)$$

Figure 4a shows the result of plotting 10^4 successive points obtained by iterating Eqs. 2 and 3 with parameters $\alpha = 1.4$ and $\beta = 0.3$ (and the initial transient is deleted). The result is essentially a picture of the chaotic attractor. Figure 4, b and c, shows successive enlargements of the small square in the preceding figure. Scale invariant, Cantor set–like structure transverse to the linear structure is evident. This suggests that we may regard the attractor in Fig. 4c, for example, as being essentially a Cantor set of approximately straight parallel lines. In fact, the dimension d in Eq. 1 can be estimated numerically (*8*) to be $d \cong 1.26$ so that the attractor is strange.

As another example consider a forced damped pendulum described by the equation

$$d^2\theta/dt^2 + \nu d\theta/dt + \omega_0^2\sin\theta = f\cos(\omega t) \quad (4)$$

where θ is the angle between the pendulum arm and the rest position, ν is the coefficient of friction, ω_0 is the frequency of natural oscillation, and f is the strength of the forcing. In Eq. 4, the first term represents the inertia of the pendulum, the second term represents friction at the pivot, the third represents the gravitational force, and the right side represents an external sinusoidally varying torque of strength f and frequency ω applied to the pendulum at the pivot. In Fig. 5a, we plot the Poincaré surface of section of a strange

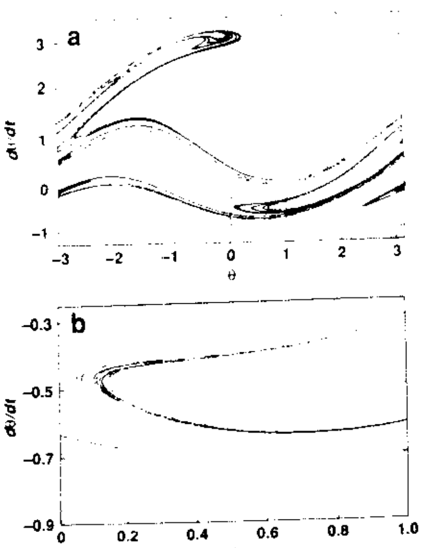


Fig. 5. (**a**) Poincaré surface of section of a pendulum strange attractor. (**b**) Enlargement of region defined by rectangle in (a).

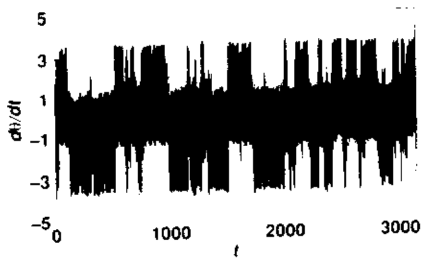


Fig. 6. Chaotic time series for pendulum shown as a plot of angular velocity versus time.

attractor for the pendulum, where we choose $\nu = 0.22$, $\omega_0 = 1.0$, $\omega = 1.0$, and $f = 2.7$ in Eq. 4. This surface of section is obtained by plotting 50,000 dots, one dot for every cycle of the forcing term, that is, one dot at every time $t = t_n = 2\pi n$ (where n is an integer). The strange attractor shown in Fig. 5a exhibits a Cantor set–like structure transverse to the linear structure. This is evident in Fig. 5b, which shows an enlargement of the square region in Fig. 5a. The dimension of this strange attractor in the surface of section is $d \cong 1.38$. Figure 6 shows the angular velocity $d\theta/dt$ as a function of t for the parameters of Fig. 5. Note the apparently erratic nature of this plot.

In general, the form of chaotic attractors varies greatly from system to system and even within the same system. This is indicated by the sequence of chaotic attractors shown in Fig. 7. All of these attractors were generated from the same map (*9*),

$$\psi_{n+1} = [\psi_n + \omega_1 + \epsilon P_1(\psi_n, \theta_n)] \bmod 1 \qquad (5)$$

$$\theta_{n+1} = [\theta_n + \omega_2 + \epsilon P_2(\psi_n, \theta_n)] \bmod 1 \qquad (6)$$

where P_1 and P_2 are periodic with period one in both their arguments. The P_1 and P_2 are the same in all of the cases shown in Fig. 7; only the parameters ω_1, ω_2, and ϵ have been varied. The results show the great variety of form and structure possible in chaotic attractors as well as their aesthetic appeal. Since ψ and θ may be regarded as angles, Eqs. 5 and 6 are a map on a two-dimensional toroidal surface. [This map is used in (*9*) to study the transition from quasiperiodicity to chaos.]

Because of the exponential divergence of nearby orbits on chaotic attractors, there is a question as to how much of the structure in these pictures of chaotic attractors (Figs. 4, 5, and 7) is an artifact due to chaos-amplified roundoff error. Although a numerical trajectory will diverge rapidly from the true trajectory with the same initial point, it has been demonstrated rigorously (*10*) in important cases [including the Hénon map (*11*)] that there exists a true trajectory with a slightly different initial point that stays near the noisy trajectory for a long time. [For example, for the Hénon map for a typical numerical trajectory computed with 14-digit precision there exists a true trajectory that stays within 10^{-7} of the numerical trajectory for 10^7 iterates (*11*).] Thus we believe that the apparently fractal structure seen in pictures such as Figs. 4, 5, and 7 is real.

The Evolution of Chaotic Attractors

In dissipative dynamics it is common to find that for some value of a system parameter only a nonchaotic attracting orbit (a limit cycle, for example) occurs, whereas at some other value of the parameter a chaotic attractor occurs. It is therefore natural to ask how the one comes about from the other as the system parameter is varied continuously. This is a fundamental question that has elicited a great deal of attention (*9, 12–19*).

To understand the nature of this question and some of the possible answers to it, we consider Fig. 8a, the so-called bifurcation diagram for the map.

$$x_{n+1} = C - x_n^2 \qquad (7)$$

where C is a constant. Figure 8a can be constructed as follows: take $C = -0.4$, set $x_0 = -0.5$, iterate the map 100 times (to eliminate transients), then plot the next 1000 values of x; increase C by a small amount, say 0.001, and repeat what was done for $C = -0.4$; increase again, and repeat; and so on, until $C = 2.1$ is reached. We see from Fig. 8a that below a certain value, $C = C_0 = -0.25$, there is no attractor in $-2 < x < 2$. In fact, in this case all orbits go to $x \to -\infty$, hence the absence of points on the plot. This is also true for C above the "crisis value" $C_c = 2.0$. Between these two values there is an attractor. As C is increased we have an attracting orbit of "period one," which, at $C = 0.75$, bifurcates to a period-two attracting orbit ($x_\alpha \to x_\beta \to x_\alpha \to x_\beta \to \cdots$), which then bifurcates (at $C = 1.25$) to a period-four orbit ($x_a \to x_b \to x_c \to x_d \to x_a \to x_b \to x_c \to x_d \to x_a \to \cdots$). In fact, there are an infinite number of such bifurcations of period 2^n to period 2^{n+1} orbits, and these accumulate as $n \to \infty$ at a finite value of C, which we denote C_∞ (from Fig. 8a, $C_\infty \cong 1.4$). [The practical importance of this phenomenology was emphasized early on by May (*12*).]

What is the situation for $C_\infty < C < C_c$? Numerically what one sees is that for many C values in this range the orbits appear to be chaotic, whereas for others there are periodic orbits. For example, Fig. 8b shows an enlargement of Fig. 8a for C in the range $1.72 < C < 1.82$. We see what appear to be chaotic orbits below $C = C_0^{(3)} = 1.75$. However, just above this value, a period-three orbit appears, supplanting the chaos. The period-three orbit then goes through a period-doubling cascade, becomes chaotic, widens into a three-piece chaotic attractor, and then finally at $C = C_c^{(3)} \cong 1.79$ widens back into a single chaotic band. We call the region $C_0^{(3)} < C < C_c^{(3)}$ a period-three window. (Such windows, but of higher period, appear throughout the region $C_\infty < C < C_c$, but are not as discernible in Fig. 8a because they are much narrower than the period-three window.)

An infinite period-doubling cascade is one way that a chaotic attractor can come about from a nonchaotic one (*13*). There are also two other possible routes to chaos exemplified in Fig. 8, a and b. These are the intermittency route (*14*) and the crisis route (*15*).

Intermittency. Consider Fig. 8b. For C just above $C_0^{(3)}$ there is a period-three orbit. For C just below $C_0^{(3)}$ there appears to be a chaotic orbit. To understand the character of this transition it is useful to examine the chaotic orbit for C just below $C_0^{(3)}$. The character of this orbit is as follows: The orbit appears to be a period-three orbit for long stretches of time after which there is a short

5

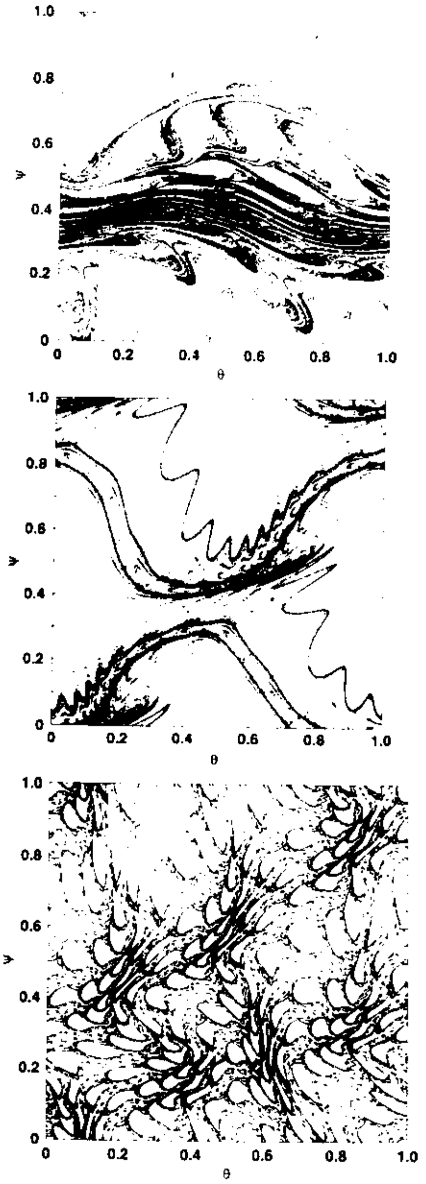


Fig. 7. Sequence of chaotic attractors for system represented by Eqs. 5 and 6. Plot shows iterated mapping on a torus for different values of ω_1, ω_2, and ϵ. (**Top**) $\omega_1 = 0.54657$, $\omega_2 = 0.36736$, and $\epsilon = 0.75$. (**Center**) $\omega_1 = 0.45922$, $\omega_2 = 0.53968$, and $\epsilon = 0.50$. (**Bottom**) $\omega_1 = 0.41500$, $\omega_2 = 0.73500$, and $\epsilon = 0.60$.

burst (the "intermittent burst") of chaotic-like behavior, followed by another long stretch of almost period-three behavior, followed by a chaotic burst, and so on. As C approaches $C_0^{(3)}$ from below, the average duration of the long stretches between the intermittent bursts becomes longer and longer (*14*), approaching infinity and proportional to $(C_0^{(3)} - C)^{-1/2}$ as $C \rightarrow C_0^{(3)}$. Thus the pure period-three orbit appears at $C = C_0^{(3)}$. Alternatively we may say that the attracting periodic attractor of period three is converted to a chaotic attractor as the parameter C decreases through the critical value $C_0^{(3)}$. It should be emphasized that, although our illustration of the transition to chaos by way of intermittency is within the context of the period-three window of the quadratic map given by Eq. 7, this phenomenon (as well as period-doubling cascades and crises) is very general; in other systems it occurs for other periods (period one, for example) in easily observable form.

Crises. From Fig. 8a we see that there is a chaotic attractor for $C < C_c = 2$, but no chaotic attractor for $C > C_c$. Thus, as C is lowered through C_c, a chaotic attractor is born. How does this occur? Note that at $C = C_c$ the chaotic orbit occupies the interval $-2 \leq x \leq 2$. If C is just slightly larger than C_c, an orbit with initial condition in the interval $-2 < x < 2$ will typically follow a chaotic-like path for a finite time, after which it finds its way out of the interval $-2 \leq x \leq 2$, and then rapidly begins to move to large negative x values (that is, it begins to approach $x = -\infty$). This is called a chaotic transient (*15*). The length of a chaotic transient will depend on the particular initial condition chosen. One can define a mean transient duration by averaging over, for example, a uniform distribution of initial conditions in the interval $-2 < x < 2$. For the quadratic map, this average duration is

$$\tau \sim 1/(C - C_c)^{\gamma} \tag{8}$$

with the exponent γ given by $\gamma = 1/2$. Thus as C approaches C_c from above, the lifetime of a chaotic transient goes to infinity and the transient is converted to a chaotic attractor for $C < C_c$. Again, this type of phenomenon occurs widely in chaotic systems. For example, the model of Lorenz (*2*) for the nonlinear evolution of the Rayleigh-Bénard instability of a fluid subjected to gravity and heated from below has a chaotic onset of the crisis type and an accompanying chaotic transient. In that case, γ in Eq. 8 is $\gamma \sim 4$ (*20*). In addition, a theory for determining the exponent γ for two-dimensional maps and systems such as the forced damped pendulum has recently been published (*21*). Thus we have seen that the period doubling, intermittency, and crisis routes to chaos are illustrated by the simple quadratic map (Eq. 7).

We emphasize that, although a map was used for illustrating these routes, all of these phenomena are present in continuous-time systems and have been observed in experiments. As an example of chaotic transitions in a continuous time system, we consider the set of three autonomous ordinary differential equations studied by Lorenz (*2*) as a model of the Rayleigh-Bénard instability,

$$dx/dt = Py - Px \tag{9}$$

$$dy/dt = -xz + rx - y \tag{10}$$

$$dz/dt = xy - bz \tag{11}$$

where P and b are adjustable parameters. Fixing $P = 10$ and $b = 8/3$ and varying the remaining parameter, r, we obtain numerical solutions that are clear examples of the intermittency and crisis types of chaotic transitions discussed above. We illustrate these in Fig. 9, a through d; the behavior of this system is as follows:

1) For r between 166.0 and 166.2 there is an intermittency transition from a periodic attractor ($r = 166.0$, Fig. 9a) to a chaotic attractor ($r = 166.2$, Fig. 9b) with intermittent turbulent bursts. Between the bursts there are long stretches of time for which the orbit oscillates in nearly the same way as for the periodic attractor (*14*) (Fig. 9a).

2) For a range of r values below $r = 24.06$ there are two periodic attractors, that represent clockwise and counterclockwise convections. For r slightly above 24.06, however, there are three attractors, one that is chaotic (shown in the phase space trajectory in Fig. 9c), whereas the other two attractors are the previously mentioned periodic attractors. The chaotic attractor comes into existence as r increases through $r = 24.06$ by conversion of a chaotic transient. Figure 9d shows an orbit in phase space executing a chaotic transient before settling down to its final resting place at one of the periodic attractors. Note the similarity of the chaotic transient trajectory in Fig. 9d with the chaotic trajectory in Fig. 9c.

The various routes to chaos have also received exhaustive experimental support. For instance, period-doubling cascades have been observed in the Rayleigh-Bénard convection (*22, 23*), in nonlinear circuits (*24*), and in lasers (*25*); intermittency has been observed in the Rayleigh-Bénard convection (*26*) and in the Belousov-Zhabotinsky reaction (*27*); and crises have been observed in nonlinear circuits (*28–30*), in the Josephson junction (*31*), and in lasers (*32*).

Finally, we note that period doubling, intermittency, and crises do not exhaust the possible list of routes to chaos. (Indeed, the

5

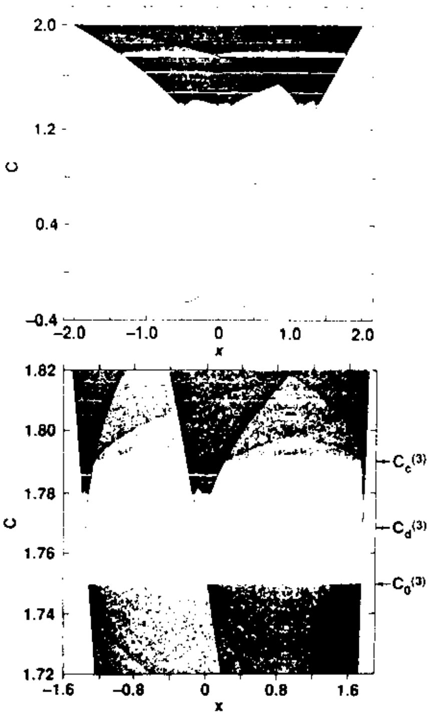


Fig. 8. (**Top**) Bifurcation diagram for the quadratic map. (**Bottom**) Period-three window for the quadratic map.

routes are not all known.) In particular, chaotic onsets involving quasiperiodicity have not been discussed here (*9, 16, 18*).

Universality

Universality refers to the fact that systems behave in certain quantitative ways that depend not on the detailed physics or model description but rather only on some general properties of the system. Universality has been examined by renormalization group (*33*) techniques developed for the study of critical phenomena in condensed matter physics. In the context of dynamics, Feigenbaum (*13*) was the first to apply these ideas, and he has extensively developed them, particularly for period doubling for dissipative systems. [See (*17*) for a collection of papers on universality in nonlinear dynamics.] For period doubling in dissipative systems, results have been obtained on the scaling behavior of power spectra for time series of the dynamical process (*34*), on the effect of noise on period doubling (*35*), and on the dependence of the Lyapunov exponent (*36*) on a system parameter. Applications of the renormalization group have also been made to intermittency (*19, 37*), and the breakdown of quasiperiodicity in dissipative (*18*) and conservative (*38*) systems.

As examples, two "universal" results can be stated within the context of the bifurcation diagrams (Fig. 8, a and b). Let C_n denote the value of C at which a period 2^n cycle period doubles to become a period 2^{n+1} cycle. Then, for the bifurcation diagram in Fig. 8a, one obtains

$$\lim_{n\to\infty} \frac{C_n - C_{n-1}}{C_{n+1} - C_n} = 4.669201\ldots \qquad (12)$$

The result given in Eq. 12 is not restricted to the quadratic map. In fact, it applies to a broad class of systems that undergo period doubling cascades (*13, 39*). In practice such cascades are very common, and the associated universal numbers are observed to be well approximated by means of fairly low order bifurcations (for example, n = 2,3,4). This scaling behavior has been observed in many experiments, including ones on fluids, nonlinear circuits, laser systems, and so forth. Although universality arguments do not explain why cascades must exist, such explanations are available from bifurcation theory (*40*).

Figure 8b shows the period-three window within the chaotic range of the quadratic map. As already mentioned, there are an infinite number of such periodic windows. [In fact, they are generally believed to be dense in the chaotic range. For example, if k is prime, there are $(2^k - 2)/(2k)$ period-k windows.] Let $C_0^{(k)}$ and $C_c^{(k)}$ denote the upper and lower values of C bounding the period-k window and let $C_d^{(k)}$ denote the value of C at which the period-k attractor bifurcates to period $2k$. Then we have that, for typical k windows (*41*).

$$\lim_{k\to\infty} \frac{C_c^{(k)} - C_0^{(k)}}{C_d^{(k)} - C_0^{(k)}} \to 9/4 \qquad (13)$$

In fact, even for the $k = 3$ window (Fig. 8b) the 9/4 value is closely approximated (it is 9/4 − 0.074 . . .). This result is universal for one-dimensional maps (and possibly more generally for any chaotic dynamical process) with windows.

Fractal Basin Boundaries

In addition to chaotic attractors, there can be sets in phase space on which orbits are chaotic but for which points near the set move away from the set. That is, they are repelled. Nevertheless, such chaotic repellers can still have important macroscopically observable effects, and we consider one such effect (*42, 43*) in this section.

Typical nonlinear dynamical systems may have more than one time-asymptotic final state (attractor), and it is important to consider the extent to which uncertainty in initial conditions leads to uncertainty in the final state. Consider the simple two-dimensional phase space diagram schematically depicted in Fig. 10. There are two attractors denoted A and B. Initial conditions on one side of the boundary, Σ, eventually asymptotically approach B; those on the other side of Σ eventually go to A. The region to the left or right of Σ is the basin of attraction for attractor A or B, respectively, and Σ is the basin boundary. If the initial conditions are uncertain by an amount ε, then for those initial conditions within ε of the boundary we cannot say a priori to which attractor the orbit eventually tends.

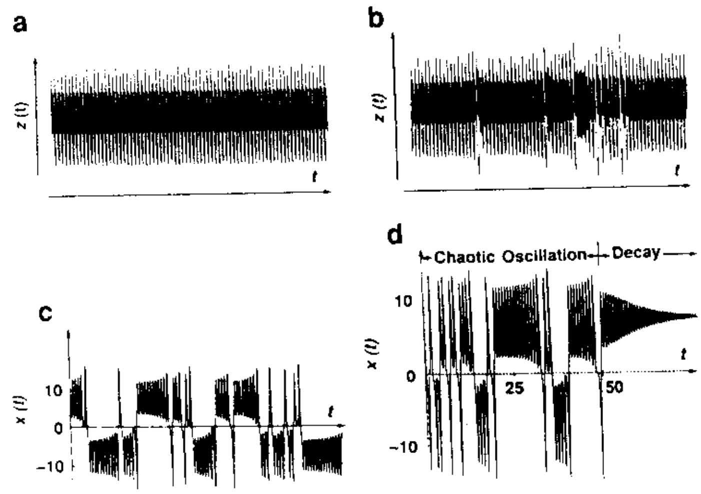
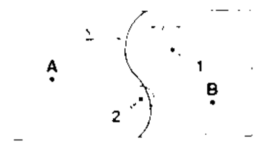


Fig. 9. Intermittency, crisis, and period doubling in continuous time systems. Intermittency in the Lorenz equations (**a**) r = 166.0; (**b**) r = 166.2. Crisis transition to a chaotic attractor in the Lorenz equations: (**c**) r = 28; (**d**) r = 22.

6

Fig. 10. A region of phase space divided by the basin boundary Σ into basins of attraction for the two attractors A and B. Points 1 and 2 are initial conditions with error ϵ.



For example, in Fig. 10, points 1 and 2 are initial conditions with an uncertainty ϵ. The orbit generated by initial condition 1 is attracted to attractor B. Initial condition 2, however, is uncertain in the sense that the orbit generated by 2 may be attracted either to A or B. In particular, consider the fraction of the uncertain phase space volume within the rectangle shown and denote this fraction f. For the case shown in Fig. 10, we clearly have $f \sim \epsilon$. The main point we wish to make in what follows is that, from the point of view of prediction, much worse scalings of f with ϵ frequently occur in nonlinear dynamics. Namely, the fraction can scale as

$$f \sim \epsilon^{\alpha} \tag{14}$$

with the "uncertainty exponent" α satisfying $\alpha < 1$ (*42, 43*). In fact, $\alpha << 1$ is fairly common. In such a case, a substantial reduction in the initial condition uncertainty, ϵ, yields only a relatively small decrease in the uncertainty of the final state as measured by f.

Although α is equal to unity for simple basin boundaries, such as that depicted in Fig. 10, boundaries with noninteger (fractal) dimension also occur. We use here the capacity definition of dimension, Eq. 1. In general, since the basin boundary divides the phase space, its dimension d must satisfy $d \geq D - 1$, where D is the dimension of the phase space. It can be proven that the following relation between the index α and the basin boundary dimension holds (*42, 43*)

$$\alpha = D - d \tag{15}$$

For a simple boundary, such as that depicted in Fig. 10, we have $d = D - 1$, and Eq. 15 then gives $\alpha = 1$, as expected. For a fractal basin boundary, $d > D - 1$, and Eq. 15 gives $\alpha < 1$.

We now illustrate the above with a concrete example. Consider the forced damped pendulum as given by Eq. 4. For parameter values $\nu = 0.2$, $\omega_0 = 1.0$, $\omega = 1.0$, and $f = 2.0$, we find numerically that the only attractors in the surface of section (θ, $d\theta/dt$) are the fixed points (-0.477, -0.609) and (-0.471, 2.037). They represent solutions with average counterclockwise and clockwise rotation at the period of the forcing. The cover shows a computer-generated picture of the basins of attraction for the two fixed point attractors. Each initial condition in a 1024 by 1024 point grid is integrated until it is close to one of the two attractors (typically 100 cycles). If an orbit goes to the attractor at $\theta = -0.477$, a blue dot is plotted at the corresponding initial condition. If the orbit goes to the other attractor, a red dot is plotted. Thus the blue and red regions are essentially pictures of the basins of attraction for the two attractors to the accuracy of the grid of the computer plotter. Fine-scale structure in the basins of attraction is evident. This is a consequence of the Cantor-set nature of the basin boundary. In fact, magnifications of the basin boundary show that, as we examine it on a smaller and smaller scale, it continues to have structure.

We now wish to explore the consequences for prediction of this infinitely fine-scaled structure. To do this, consider an initial condition (θ, $d\theta/dt$). What is the effect of a small change ϵ in the θ-coordinate? Thus we integrate the forced pendulum equation with the initial conditions (θ, $d\theta/dt$), (θ, $d\theta/dt + \epsilon$), and (θ, $d\theta/dt - \epsilon$) until they approach one of the attractors. If either or both of the perturbed initial conditions yield orbits that do not approach the same attractor as the unperturbed initial condition, we say that (θ, $d\theta/dt$) is uncertain. Now we randomly choose a large number of initial conditions and let $\hat{f}$ denote the fraction of these that we find to be uncertain. As a result of these calculations, we find that $\hat{f} \sim \epsilon^{\hat{\alpha}}$ where $\hat{\alpha} \cong 0.275 \pm 0.005$. If we assume that $\hat{f}$, determined in the way stated above, is approximately proportional to f [there is some support for this conjecture from theoretical work (*44*)], then $\alpha = \hat{\alpha}$. Thus, from Eq. 15, the dimension of the basin boundary is $d \cong 1.725 \pm 0.005$. We conclude, from Eq. 14, that in this case if we are to gain a factor of 2 in the ability to predict the asymptotic final state of the system, it is necessary to increase the accuracy in the measurement of the initial conditions by a factor substantially greater than 2 (namely by $2^{1/0.275} \cong 10$). Hence, fractal basin boundaries ($\alpha < 1$) represent an obstruction to predictability in nonlinear dynamics.

Some representative works on fractal basin boundaries, including applications, are listed in (*42–47*). Notable basic questions that have recently been answered are the following:

1) How does a nonfractal basin boundary become a fractal basin boundary as a parameter of the system is varied (*45*)? This question is similar, in spirit, to the question of how chaotic attractors come about.

2) Can fractal basin boundaries have different dimension values in different regions of the boundary, and what boundary structures lead to this situation? This question is addressed in (*46*) where it is shown that regions of different dimension can be intertwined on an arbitrarily fine scale.

3) What are the effects of a fractal basin boundary when the system is subject to noise? This has been addressed in the Josephson junction experiments of (*31*).

Conclusion

Chaotic nonlinear dynamics is a vigorous, rapidly expanding field. Many important future applications are to be expected in a variety of areas. In addition to its practical aspects, the field also has fundamental implications. According to Laplace, determination of the future depends only on the present state. Chaos adds a basic new aspect to this rule: small errors in our knowledge can grow exponentially with time, thus making the long-term prediction of the future impossible.

Although the field has advanced at a great rate in recent years, there is still a wealth of challenging fundamental questions that have yet to be adequately dealt with. For example, most concepts developed so far have been discovered in what are effectively low-dimensional systems; what undiscovered important phenomena will appear only in higher dimensions? Why are transiently chaotic motions so prevalent in higher dimensions? In what ways is it possible to use the dimension of a chaotic attractor to determine the dimension of the phase space necessary to describe the dynamics? Can renormalization group techniques be extended past the borderline of chaos into the strongly chaotic regime? These are only a few questions. There are many more, and probably the most important questions are those that have not yet been asked.

REFERENCES AND NOTES

1. C. Grebogi, E. Ott, S. Pelikan, J. A. Yorke, *Physica* **13D**, 261 (1984).
2. E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).
3. J. D. Farmer, E. Ott, J. A. Yorke, *Physica* **7D**, 153 (1983).
4. J. Kaplan and J. A. Yorke, *Lecture Notes in Mathematics No. 730* (Springer-Verlag, Berlin, 1978), p. 228; L. S. Young, *Ergodic Theory Dyn. Syst.* **1**, 381 (1981).
5. P. Grassberger and I. Procaccia, *Phys. Rev. Lett.* **50**, 346 (1983); H. G. E. Hentschel and I. Procaccia, *Physica* **8D**, 435 (1983); P. Grassberger, *Phys. Lett.* **A97**, 227 (1983); T. C. Halsey *et al.*, B. I. Shraiman, *Phys. Rev. A* **33**, 1141 (1986); C. Grebogi, E. Ott, J. A. Yorke, *ibid.* **36**, 3522 (1987).
6. A. Bondeson *et al.*, *Phys. Rev. Lett.* **55**, 2103 (1985); F. J. Romeiras, A. Bondeson, E. Ott, T. M. Antonsen, C. Grebogi, *Physica* **26D**, 277 (1987).
7. M. Hénon, *Commun. Math. Phys.* **50**, 69 (1976).
8. D. A. Russell, J. D. Hanson, E. Ott, *Phys. Rev. Lett.* **45**, 1175 (1980).
9. C. Grebogi, E. Ott, J. A. Yorke, *Physica* **15D**, 354 (1985).

4

10. D. V. Anosov, *Proc. Steklov Ins. Math.* **90** (1967); R. Bowen, *J. Differ. Equations* **18**, 333 (1975).
11. S. M. Hammel, J. A. Yorke, C. Grebogi, *J. Complexity*, 3, 136 (1987).
12. R. M. May, *Nature (London)* **261**, 459 (1976).
13. M. J. Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978).
14. Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980).
15. C. Grebogi, E. Ott, J. A. Yorke, *Physica* **7D**, 181 (1983).
16. D. Ruelle and F. Takens, *Commun. Math. Phys.* **20**, 167 (1971).
17. P. Cvitanovic, Ed. *Universality in Chaos* (Hilger, Bristol, 1984).
18. For example, S. J. Shenker, *Physica* **5D**, 405 (1982); K. Kaneko, *Progr. Theor. Phys.* **71**, 282 (1984); M. J. Feigenbaum, L. P. Kadanoff, S. L. Shenker, *Physica* **5D**, 370 (1982); D. Rand, S. Ostlund, J. Sethna, E. Siggia, *Physica* **8D**, 303 (1983); S. Kim and S. Ostlund, *Phys. Rev. Lett.* **55**, 1165 (1985); D. K. Umberger, J. D. Farmer, I. I. Satija, *Phys. Lett. A* **114**, 341 (1986); P. Bak, T. Bohr, M. H. Jensen, *Phys. Scr.* **T9**, 50 (1985); P. Bak, *Phys. Today* **39** (No. 12), 38 (1987).
19. J. E. Hirsch, M. Nauenberg, D. J. Scalapino, *Phys. Lett. A* **87**, 391 (1982).
20. J. A. Yorke and E. D. Yorke, *J. Stat. Phys.* **21**, 263 (1979); in *Topics in Applied Physics* (Springer-Verlag, New York, 1981), vol. 45, p. 77.
21. C. Grebogi, E. Ott, J. A. Yorke, *Phys. Rev. Lett.* **57**, 1284 (1986).
22. A. Libchaber and J. Maurer, *J. Phys. (Paris)* **41**, C3-51 (1980); A. Libchaber, C. Laroche, S. Fauve, *J. Phys. (Paris) Lett.* **43**, L211 (1982).
23. J. P. Gollub, S. V. Benson, J. F. Steinman, *Ann. N.Y. Acad. Sci.* **357**, 22 (1980); M. Giglio, S. Musazzi, U. Perini, *Phys. Rev. Lett.* **47**, 243 (1981).
24. P. S. Linsay, *Phys. Rev. Lett.* **47**, 1349 (1981).
25. F. T. Arecchi, R. Meucci, G. Puccioni, J. Tredicce, *ibid.* **49**, 1217 (1982).
26. M. Dubois, M. A. Rubio, P. Bergé, *ibid.* **51**, 1446 (1983).
27. J. C. Roux, P. DeKepper, H. L. Swinney, *Physica* **7D**, 57 (1983).
28. C. Jeffries and J. Perez, *Phys. Rev. A* **27**, 601 (1983); S. K. Brorson, D. Dewey, P. S. Linsay, *ibid.* **28**, 1201 (1983).
29. H. Ikezi, J. S. deGrasse, T. H. Jensen, *ibid.* **28**, 1207 (1983).
30. R. W. Rollins and E. R. Hunt, *ibid.* **29**, 3327 (1984).
31. M. Iansiti *et al.*, *Phys. Rev. Lett.* **55**, 746 (1985).
32. D. Dangoisse, P. Glorieux, D. Hannequin, *ibid.* **57**, 2657 (1986).
33. K. G. Wilson and J. Kogut, *Phys. Rep. C* **12**, 75 (1974); B. Hu, *ibid.* **91**, 233 (1982).
34. M. J. Feigenbaum, *Phys. Lett. A* **74**, 375 (1979); R. Brown, C. Grebogi, E. Ott, *Phys. Rev. A* **34**, 2248 (1986); M. Nauenberg and J. Rudnick, *Phys. Rev. B* **24**, 493 (1981); B. A. Huberman and A. B. Zisook, *Phys. Rev. Lett.* **46**, 626 (1981); J. D. Farmer, *ibid.* 47, 179 (1981).
35. J. Crutchfield, M. Nauenberg, J. Rudnick, *Phys. Rev. Lett.* **46**, 933 (1981); B. Shraiman, C. E. Wayne, P. C. Martin, *ibid.*, p. 935.
36. B. A. Huberman and J. Rudnick, *ibid.* **45**, 154 (1980).
37. B. Hu and J. Rudnick, *ibid.* **48**, 1645 (1982).
38. L. P. Kadanoff, *ibid.* 47, 1641 (1981); D. F. Escande and F. Doveil, *J. Stat. Phys.* 26, 257 (1981); R. S. MacKay, *Physica* **7D**, 283 (1983).
39. P. Collet, J. P. Eckmann, O. E. Lanford III, *Commun. Math. Phys.* 76, 211 (1980).
40. J. A. Yorke and K. A. Alligood, *ibid.* **100**, 1 (1985).
41. J. A. Yorke, C. Grebogi, E. Ott, L. Tedeschini-Lalli, *Phys. Rev. Lett.* **54**, 1095 (1985).
42. C. Grebogi, S. W. McDonald, E. Ott, J. A. Yorke, *Phys. Lett. A* **99**, 415 (1983).
43. S. W. McDonald, C. Grebogi, E. Ott, J. A. Yorke, *Physica* **17D**, 125 (1985).
44. S. Pelikan, *Trans. Am. Math. Soc.* **292**, 695 (1985).
45. C. Grebogi, E. Ott, J. A. Yorke, *Phys. Rev. Lett.* **56**, 1011 (1986); *Physica* **24D**, 243 (1987); F. C. Moon and G.-X. Li, *Phys. Rev. Lett.* **55**, 1439 (1985).
46. C. Grebogi, E. Kostelich, E. Ott, J. A. Yorke, *Phys. Lett.* **A118**, 448 (1986); *Physica* **25D**, 347 (1987); C. Grebogi, E. Ott, J. A. Yorke, H. E. Nusse, *Ann. N.Y. Acad. Sci.* **497**, 117 (1987).
47. C. Mira, *C. R. Acad. Sci.* **288A**, 591 (1979); C. Grebogi, E. Ott, J. A. Yorke, *Phys. Rev. Lett.* **50**, 935 (1983); R. G. Holt and I. B. Schwartz, *Phys. Lett.* **A105**, 327 (1984); I. B. Schwartz, *ibid.* **106**, 339 (1984); I. B. Schwartz, *J. Math. Biol.* **21**, 347 (1985); S. Takesue and K. Kaneko, *Progr. Theor. Phys.* 71, 35 (1984); O. Decroly and A. Goldbeter, *Phys. Lett.* **A105**, 259 (1984); E. G. Gwinn and R. M. Westervelt, *Phys. Rev. Lett.* **54**, 1613 (1985); *Phys. Rev. A* **33**, 4143 (1986); Y. Yamaguchi and N. Mishima, *Phys. Lett.* **A109**, 196 (1985); M. Napiorkowski, *ibid.* **113**, 111 (1985); F. T. Arecchi, R. Badii, A. Politi, *Phys. Rev. A* **32**, 402 (1985); S. W. McDonald, C. Grebogi, E. Ott, J. A. Yorke, *Phys. Lett.* **A107**, 51 (1985); J. S. Nicolis and I. Tsuda, in *Simulation, Communication, and Control*, S. G. Tzafestas, Ed. (North-Holland, Amsterdam, 1985); J. S. Nicolis, *Rep. Prog. Phys.* **49**, 1109 (1986); J. S. Nicolis, *Kybernetes* **14**, 167 (1985).
48. This work was supported by the Air Force Office of Scientific Research, the U.S. Department of Energy, the Defense Advanced Research Projects Agency, and the Office of Naval Research.

COVER Even systems as simple as a periodically forced damped pendulum can have complex behavior. This computer-generated plot shows initial pendulum velocities (measured horizontally) and positions (measured vertically). Orbits starting at points in the red region eventually settle into one type of periodic motion, while orbits starting in the blue region yield a different type of periodic motion. The boundary between these regions is fractal. The lighter the shade of red or blue, the longer it takes to settle into the corresponding motion. See page 632. [Photo courtesy of C. Grebogi, E. Ott, and J. A. Yorke, University of Maryland, College Park, MD 20742]

8

Controlling Chaos

Edward Ott,[(a),(b)] Celso Grebogi,[(a)] and James A. Yorke[(c)]
University of Maryland, College Park, Maryland 20742
(Received 22 December 1989)

It is shown that one can convert a chaotic attractor to any one of a large number of possible attracting time-periodic motions by making only *small* time-dependent perturbations of an available system parameter. The method utilizes delay coordinate embedding, and so is applicable to experimental situations in which *a priori* analytical knowledge of the system dynamics is not available. Important issues include the length of the chaotic transient preceding the periodic motion, and the effect of noise. These are illustrated with a numerical example.

PACS numbers: 05.45.+b

The presence of chaos in physical systems has been extensively demonstrated and is very common. In practice, however, it is often desired that chaos be avoided and/or that the system performance be improved or changed in some way. Given a chaotic attractor, one approach might be to make some large and possibly costly alteration in the system which completely changes its dynamics in such a way as to achieve the desired behavior. Here we assume that this avenue is not available. Thus, we address the following question: Given a chaotic attractor, how can one obtain improved performance and a desired attracting time-periodic motion by making only *small* time-dependent perturbations in an *accessible* system parameter?

The key observation is that a chaotic attractor typically has embedded within it an infinite number of unstable periodic orbits.[1] Since we wish to make only small perturbations to the system, we do not envision creating new orbits with very different properties from the existing ones. Thus, we seek to exploit the already existing unstable periodic orbits. Our approach is as follows: We first determine some of the unstable low-period periodic orbits that are embedded in the chaotic attractor. We then examine these orbits and choose one which yields improved system performance. Finally, we tailor our small time-dependent parameter perturbations so as to stabilize this already existing orbit. In this Letter we describe how this can be done, and we illustrate the method with a numerical example. The method is very general and should be capable of yielding greatly improved performance in a wide variety of situations.

It is interesting to note that if the situation is such that the suggested method is practical, then the presence of chaos can be a great advantage. The point is that any one of a number of different orbits can be stabilized, and the choice can be made to achieve the best system performance among those orbits. If, on the other hand, the attractor is not chaotic but is, say, periodic, then small parameter perturbations can only change the orbit slightly. Basically we are then stuck with whatever system performance the stable periodic orbit gives, and we have no option for substantial improvement, short of making large alterations in the system.

Furthermore, one may want a system to be used for different purposes or under different conditions at different times. Thus, depending on the use, different requirements are made of the system. If the system is chaotic, this type of multiple-use situation might be accommodated without alteration of the gross system configuration. In particular, depending on the use desired, the system behavior could be changed by switching the temporal programming of the small parameter perturbations to stabilize different orbits. In contrast, in the absence of chaos, completely separate systems might be required for each use. Thus, when designing a system intended for multiple uses, purposely building chaotic dynamics into the system may allow for the desired flexibility. Such multipurpose flexibility is essential to higher life forms, and we, therefore, speculate that chaos may be a necessary ingredient in their regulation by the brain.

To simplify the analysis we consider continuous-time dynamical systems which are *three dimensional* and depend on one system parameter which we denote p [for example, $d\mathbf{x}/dt=\mathbf{F}(\mathbf{x},p)$, where $\mathbf{x}$ is three dimensional]. We assume that the parameter p is available for external adjustment, and we wish to temporally program our adjustments of p so as to achieve improved performance. We emphasize that our restriction to a three-dimensional system is mainly for ease of presentation, and that the case of higher-dimensional (including infinite-dimensional) systems can be treated by similar methods.[2]

We imagine that the dynamical equations describing the system are not known, but that experimental time series of some scalar-dependent variable $z(t)$ can be measured. Using delay coordinates[3,4] with delay T one can form a delay-coordinate vector,

$$\mathbf{X}(t)=[z(t),z(t-T),z(t-2T),\ldots,z(t-MT)]\,.$$

We are interested in periodic orbits and their stability properties, and we shall use $\mathbf{X}$ to obtain a surface of section for this purpose. In the surface of section, a continuous-time-periodic orbit appears as a discrete-time orbit cycling through a finite set of points. We require the dynamical behavior of the surface of section map in

9

neighborhoods of these points in order to study the stability of the periodic orbits. To embed a small neighborhood of a point from $\mathbf{x}$ into $\mathbf{X}$, we typically only require as many dimensions as there are coordinates of the point. Thus, for our purposes, $M=D-1$ is generally sufficient. (This is in contrast with[3] $M+1=2D+1$, typically required for global embedding of the original phase space in the delay-coordinate space.) Hence, for the case considered ($D=3$), our surface of section is two dimensional.

We suppose that the parameter p can be varied in a small range about some nominal value p_0. Henceforth, without loss of generality, we set $p_0 \equiv 0$. Let the range in which we are allowed to vary p be $p_* > p > -p_*$.

Using an experimental surface of section for the embedding vector $\mathbf{X}$, we imagine that we obtain many experimental points in the surface of section for $p=0$. We denote these points $\xi_1, \xi_2, \xi_3, \ldots, \xi_k$, where ξ_n denotes the coordinates in the surface of section at the nth piercing of the surface of section by the orbit $\mathbf{X}(t)$. For example, a common choice of the surface of section would be $z(t-MT)$ equals a constant, and $\xi_n=[z(t_n), \ldots, z(t_n-(M-1)T)]$, where $t=t_n$ denotes the time at the nth piercing. From such experimentally determined sequences it has been demonstrated that a large number of distinct unstable periodic orbits on a chaotic attractor can be determined.[5,6] We then examine these unstable periodic orbits and select the one which gives the best performance. Again using an experimentally determined sequence, we obtain the stability properties of the chosen periodic orbit (cf. Refs. 5 and 6 for discussion of how this can be done and for descriptions of its implementation in concrete experimental cases). For the purposes of simplicity, let us assume in what follows that this orbit is a fixed point of the surface of section map (i.e., period one; the case of higher period is a straightforward extension). Let λ_s and λ_u be the experimentally determined stable and unstable eigenvalues of the surface of section map at the chosen fixed point of the map ($|\lambda_u| > 1 > |\lambda_s|$). Let $\mathbf{e}_s$ and $\mathbf{e}_u$ be the experimentally determined unit vectors in the stable and unstable directions. Let $\xi=\xi_F \equiv 0$ be the desired fixed point. We then change p slightly from $p=0$ to some other value $p=\bar{p}$. The fixed-point coordinates in the experimental surface of section will shift from 0 to some nearby point $\xi_F(\bar{p})$ and we determine this new position. For small $\bar{p}$ we approximate $\mathbf{g} \equiv \partial \xi_F(p)/\partial p\,|_{p=0} \cong \bar{p}^{-1}\xi_F(\bar{p})$, which allows an experimental determination of the vector $\mathbf{g}$.

Thus, in the surface of section, near $\xi=0$, we can use a linear approximation for the map, $\xi_{n+1}-\xi_F(p) \cong \mathbf{M}\cdot[\xi_n-\xi_F(p)]$, where $\mathbf{M}$ is a 2×2 matrix. Using $\xi_F(p) \cong p\mathbf{g}$ we have

$$\xi_{n+1} \cong p_n\mathbf{g}+[\lambda_u\mathbf{e}_u\mathbf{f}_u+\lambda_s\mathbf{e}_s\mathbf{f}_s]\cdot[\xi_n-p_n\mathbf{g}]\,. \tag{1}$$

[In the linearization (1), we have considered p_n to be small and of the same order as ξ_n.] We emphasize that $\mathbf{g}$, $\mathbf{e}_u$, $\mathbf{e}_s$, λ_u, and λ_s are all experimentally accessible by the embedding technique just discussed. In (1) $\mathbf{f}_u$ and $\mathbf{f}_s$ are contravariant basis vectors defined by $\mathbf{f}_s\cdot\mathbf{e}_s=\mathbf{f}_u\cdot\mathbf{e}_u=1$, $\mathbf{f}_s\cdot\mathbf{e}_u=\mathbf{f}_u\cdot\mathbf{e}_s=0$. Note that we have written the location of the fixed point as $p_n\mathbf{g}$ because we imagine that we adjust p to a new value p_n after each piercing of the surface of section. That is, we observe ξ_n and then adjust p to the value p_n. Thus p_n depends on ξ_n. Further, we only envision making this adjustment when the orbit falls near the desired fixed point for $p=0$.

Assume that ξ_n falls near the desired fixed point at $\xi=0$ so that (1) applies. We then attempt to pick p_n so that ξ_{n+1} falls on the stable manifold of $\xi=0$. That is, we choose p_n so that $\mathbf{f}_u\cdot\xi_{n+1}=0$. If ξ_{n+1} falls on the stable manifold of $\xi=0$, we can then set the parameter perturbations to zero, and the orbit for subsequent time will approach the fixed point at the geometrical rate λ_s. Thus, for sufficiently small ξ_n, we can dot (1) with $\mathbf{f}_u$ to obtain

$$p_n=\lambda_u(\lambda_u-1)^{-1}(\xi_n\cdot\mathbf{f}_u)/(\mathbf{g}\cdot\mathbf{f}_u)\,, \tag{2}$$

which we use when the magnitude of the right-hand side of (2) is less than p_*. When it is greater than p_*, we set $p_n=0$. We assume in (2) that the generic condition $\mathbf{g}\cdot\mathbf{f}_u \neq 0$ is satisfied. Thus, the parameter perturbations are activated (i.e., $p_n \neq 0$) only if ξ_n falls in a narrow strip $|\xi_n^u| < \xi_*$, where $\xi_n^u=\mathbf{f}_u\cdot\xi_n$, and from (2) $\xi_*=p_*|(1-\lambda_u^{-1})\mathbf{g}\cdot\mathbf{f}_u|$. Thus, for small p_*, a typical initial condition will execute a chaotic orbit, unchanged from the uncontrolled case, until ξ_n falls in the strip. Even then, because of nonlinearity not included in (1), the control may not be able to bring the orbit to the fixed point. In this case the orbit will leave the strip and continue to wander chaotically as if there was no control. Since the orbit on the uncontrolled chaotic attractor is ergodic, at some time it will eventually satisfy $|\xi_n^u| < \xi_*$ and also be sufficiently close to the desired fixed point that attraction to $\xi=0$ is achieved. [In rare cases applying Eq. (2) when the trajectory enters the strip, but is still far from 0, may result in stabilizing the wrong periodic orbit which visits the strip.]

Thus, we create a stable orbit, but, for a typical initial condition, it is preceded in time by a chaotic transient in which the orbit is similar to orbits on the uncontrolled chaotic attractor. The length τ of such a chaotic transient depends sensitively on the initial condition, and, for randomly chosen initial conditions, has an exponential probability distribution[7] $P(\tau) \sim \exp[-(\tau/\langle\tau\rangle)]$ for large τ. The average length of the chaotic transient $\langle\tau\rangle$ increases with decreasing p_* and follows a power-law relation[7] for small p_*, $\langle\tau\rangle \sim p_*^{-\gamma}$.

We will now derive a formula for the exponent γ. Dotting the linearized map for ξ_{n+1}, Eq. (1), with $\mathbf{f}_u$, we obtain $\xi_{n+1}^u \cong 0$. In obtaining this result from (1) we have substituted p_n appropriate for $|\xi_n^u| < \xi_*$. We note that the result $\xi_{n+1}^u \cong 0$ is a linearization, and typically

10

has a lowest-order nonlinear correction that is quadratic. In particular, $\xi_n^s = \mathbf{f}_s \cdot \boldsymbol{\xi}_n$ is not restricted by $|\xi_n^u| < \xi_*$, and thus may not be small when the condition $|\xi_n^u| < \xi_*$ is satisfied. Hence the correction quadratic in ξ_n^s is most significant. Including such a correction we have $\xi_{n+1}^u \cong \kappa(\xi_n^s)^2$, where κ is a constant. Thus, if $|\kappa|(\xi_n^s)^2 > \xi_*$, then $|\xi_{n+1}^u| > \xi_*$, and attraction to $\boldsymbol{\xi} = 0$ is not achieved, even though $|\xi_n^u| < \xi_*$. Attraction to $\boldsymbol{\xi} = 0$ is achieved when the orbit falls in the small parallelogram P_c given by $|\xi_n^u| < \xi_*$, $|\xi_n^s| < (\xi_*/|\kappa|)^{1/2}$. For very small ξ_*, an initial condition will bounce around on the set comprising the uncontrolled chaotic attractor for a long time before it falls in the parallelogram P_c. At any given iterate the probability of falling in P_c is $\mu(P_c)$, the measure of the uncontrolled attractor contained in P_c. Thus, $\langle\tau\rangle^{-1} = \mu(P_c)$. The scaling of $\mu(P_c)$ with ξ_* is

$$\mu(P_c) \sim (\xi_*)^{d_u}[(\xi_*/|\kappa|)^{1/2}]^{d_s} \sim \xi_*^{d_u + (1/2)d_s},$$

where d_u and d_s are the partial pointwise dimensions for the uncontrolled chaotic attractor at $\boldsymbol{\xi} = 0$ in the unstable direction and the stable direction, respectively. Thus, $\mu(P_c) = \xi_*^\gamma$, where $\gamma = d_u + d_s/2$. Since we assume the attractor to be effectively smooth in the unstable direction, $d_u = 1$. The partial pointwise dimension in the stable direction is given in terms of the eigenvalues[7] at $\boldsymbol{\xi} = 0$, $d_s = \ln|\lambda_u|/\ln|\lambda_s|^{-1}$. Thus,

$$\gamma = 1 + \tfrac{1}{2}\ln|\lambda_u|/\ln|\lambda_s|^{-1}. \tag{3}$$

To study the effect of noise we add a term $\epsilon\boldsymbol{\delta}_n$ to the right-hand side of the linearized equations for $\boldsymbol{\xi}_{n+1}$, Eq. (1), where $\boldsymbol{\delta}_n$ is a random variable and ϵ is a small parameter specifying the intensity of the noise. The quantities $\boldsymbol{\delta}_n$ are taken to have zero mean ($\langle\boldsymbol{\delta}_n\rangle = 0$), be independent ($\langle\boldsymbol{\delta}_n\boldsymbol{\delta}_m\rangle = 0$ for $m \neq n$), and have a probability density independent of n. Dotting (1) with noise included with $\mathbf{f}_u$ we obtain $\xi_{n+1}^u = \epsilon\delta_n^u$, where $\delta_n^u \equiv \mathbf{f}_u \cdot \boldsymbol{\delta}_n$. Thus, if the noise is bounded, $|\delta_n^u| < \delta_{\max}$, then the stability of $\boldsymbol{\xi} = 0$ will not be affected by the noise if the bound is small enough, $\epsilon\delta_{\max} < \xi_*$. If this condition is not satisfied, then the noise can kick an orbit which is initially in the parallelogram P_c into the region outside P_c. We are particularly interested in the case where such kickouts are caused by low-probability tails on the probability density and are thus rare. (If they are frequent, then our procedure is ineffective.) In such a case the average time to be kicked out $\langle\tau'\rangle$ will be long. Thus, an orbit will typically alternate between epochs of chaotic motion of average duration $\langle\tau\rangle$ in which it is far from $\boldsymbol{\xi} = 0$, and epochs of average length $\langle\tau'\rangle$ in which the orbit lies in the parallelogram P_c. For small enough noise the orbit spends most of its time in P_c, $\langle\tau'\rangle \gg \langle\tau\rangle$, and one might then regard the procedure as being effective.

We now consider a specific numerical example. Our purpose is to illustrate and test our analyses of the average time to achieve control and the effect of noise. To do this we shall utilize the Henon map, $x_{n+1} = A - x_n^2 + By_n$, $y_{n+1} = x_n$, where we take $B = 0.3$. We assume that the quantity A can be varied by a small amount about some value A_0. Accordingly, we write A as $A = A_0 + p$, where p is the control parameter. For the values of A_0 which we investigate, the attractor for the map is chaotic and contains an unstable period-one (fixed-point) orbit. The coordinates (x_F, y_F) of the fixed point which is in the attractor for $p = 0$ along with the associated parameters and vectors appearing in Eq. (1) may be explicitly calculated. The quantity ξ_n appearing in (1) is $\boldsymbol{\xi}_n = (x_n - x_F)\mathbf{x}_0 + (y_n - y_F)\mathbf{y}_0$. To test our prediction for the dependence of $\langle\tau\rangle$, the average time to approach $\boldsymbol{\xi} = 0$, on the maximum allowed size of the parameter perturbation p_*, we proceed as follows. We iterate the map with $p = 0$ using a large number of randomly chosen initial conditions until all these initial conditions are distributed over the attractor (500 iterates were typically used). We then turn on the parameter perturbations and determine for each orbit how many further iterates τ are necessary before the orbit falls within a circle of radius $\frac{1}{2}\xi_*$ centered at the fixed point. We then calculate the average of these times. We do this for many different values of p_* and plot the results as a function of p_*. This is shown on the log-log plot in Fig. 1 along with the theoretical straight line of slope given by the exponent (3). We see that the agreement is

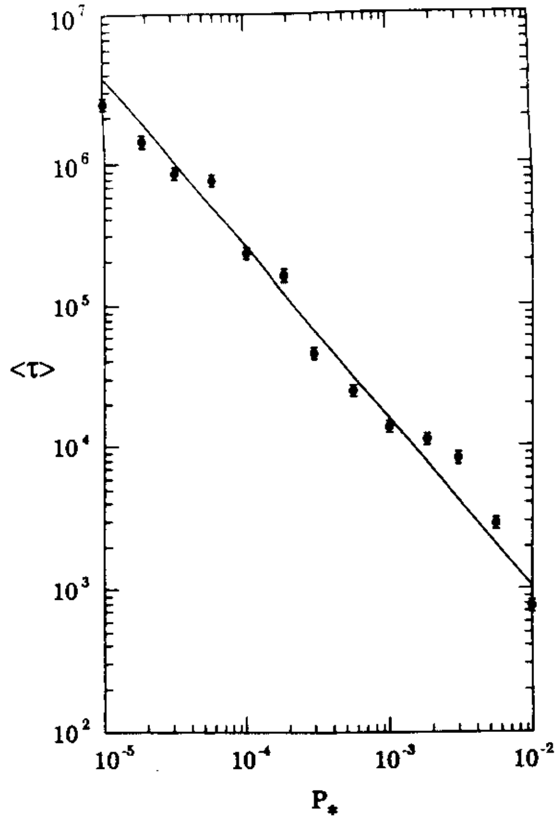


FIG. 1. $\langle\tau\rangle$ vs p_*. Points were computed using 128 randomly selected initial conditions. $A_0 = 1.4$.

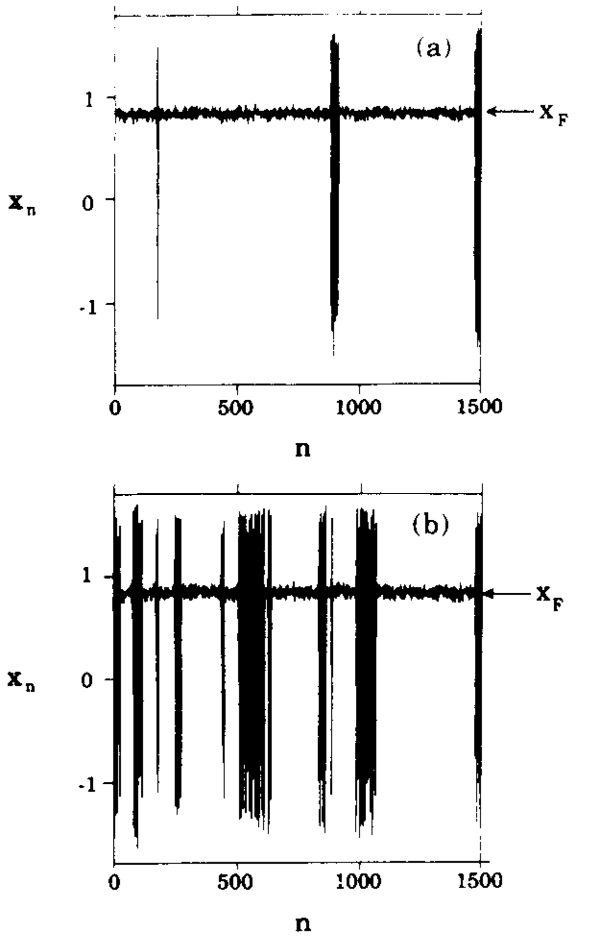


FIG. 2. x_n vs n for two cases with the same realization of the random vector δ. $p_* = 0.2$ and $A_0 = 1.29$ for both cases. (a) $\epsilon = 3.5\times10^{-2}$; (b) $\epsilon = 3.8\times10^{-2}$.

good although there are significant variations about the general power-law trend. These are to be expected due to the fractal nature of the attractor and have also been seen in numerical calculations of the pointwise dimension for points on chaotic attractors (cf. Grebogi, Ott, and Yorke[1]).

Next, we consider the issue of noise. We add terms $\epsilon\delta_{xn}$ and $\epsilon\delta_{yn}$ to the right-hand sides of the Henon map equations. The random quantities δ_{xn} and δ_{yn} are independent of each other, have mean value 0 and mean-squared value 1 ($\langle\delta_x^2\rangle = \langle\delta_y^2\rangle = 1$), and have a Gaussian probability density. Figure 2 shows orbit plots, x_n vs n for 1500 iterates of the noisy map with parameter perturbations given by (2), for two different noise levels and p_* held fixed at $p_* = 0.2$. As predicted the orbit stays near the fixed point with occasional bursts into the region far from $\xi = 0$, and these bursts are less frequent for small noise levels.

In conclusion, we have shown that there is great inherent flexibility in situations in which the dynamical motion is on a chaotic attractor. In particular, by using only small (carefully chosen) parameter perturbations it is possible to create a large variety of attracting periodic motions and to choose amongst these periodic motions the one most desirable.[8]

This research was supported by the U.S. Department of Energy (Scientific Computing Staff Office of Energy Research). The computation was done at the National Energy Research Supercomputer Center.

[(a)]Laboratory for Plasma Research.

[(b)]Departments of Electrical Engineering and of Physics.

[(c)]Institute for Physical Science and Technology and Department of Mathematics.

[1]The periodic orbits are dense in the attractor [i.e., periodic orbits pass through any neighborhood (however small) of any point on the attractor]. For discussions of the relation of ergodic properties of an attractor to its dense set of unstable periodic orbits, see, for example, C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. A **37**, 1711 (1988); **36**, 3522 (1987); D. Auerbach *et al.*, Phys. Rev. Lett. **58**, 2387 (1987); H. Hata *et al.*, Prog. Theor. Phys. **78**, 511 (1987); A. Katok, Publ. Math. IHES **51**, 137 (1980); R. Bowen, Trans. Am. Math. Soc. **154**, 377 (1971).

[2]E. Ott, C. Grebogi, and J. A. Yorke, in *Chaos: Proceedings of a Soviet-American Conference* (American Institute of Physics, New York, 1990).

[3]F. Takens, in *Dynamical Systems and Turbulence*, edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), p. 230.

[4]N. H. Packard *et al.*, Phys. Rev. Lett. **45**, 712 (1980).

[5]G. H. Gunaratne, P. S. Linsay, and M. J. Vinson, Phys. Rev. Lett. **63**, 1 (1989).

[6]D. P. Lathrop and E. J. Kostelich, "The Characterization of an Experimental Strange Attractor by Periodic Orbits" (to be published).

[7]C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **57**, 1284 (1986); P. Romeiras, C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. A **36**, 5365 (1987).

[8]The general problem of controlling chaotic systems, while clearly very important, has, so far, received almost no attention. Two exceptions (which are quite different from our approach) are the papers of Hubler (who typically requires large controlling signals) and Fowler [A. Hubler, Helv. Phys. Acta **62**, 343 (1989); T. B. Fowler, IEEE Trans. Autom. Control **34**, 201 (1989)].

12

Using Chaos to Direct Trajectories to Targets

Troy Shinbrot,[a] Edward Ott,[b),(c] Celso Grebogi,[b),(d] and James A. Yorke[d]
University of Maryland, College Park, Maryland 20742
(Received 1 June 1990)

A method is developed which uses the exponential sensitivity of a chaotic system to tiny perturbations to direct the system to a desired accessible state in a short time. This is done by applying a small, judiciously chosen, perturbation to an available system parameter. An expression for the time required to reach an accessible state by applying such a perturbation is derived and confirmed by numerical experiment. The method introduced is shown to be effective even in the presence of small-amplitude noise or small modeling errors.

PACS numbers: 05.45.+b

Chaotic systems exhibit extreme sensitivity to initial conditions. This characteristic is often regarded as an annoyance, yet it provides us with an extremely useful capability without a counterpart in nonchaotic systems. In particular, the future state of a chaotic system can be substantially altered by a tiny perturbation. If we can accurately sense the state of the system and intelligently perturb it, this presents us with the possibility of rapidly directing the system to a desired state. As we will show below for a particular example, an initial condition for a chaotic system which, if left on its own, would require more than 6000 time steps to reach a small target region can be directed with small perturbations to the desired region in only 12 time steps. In this paper, we report a method for achieving such targeting, and we show that the method can be effective even in the presence of small-amplitude noise and small modeling errors.

For simplicity, we consider a three-dimensional continuous-time dynamical system, $d\mathbf{X}/dt=\mathbf{F}(\mathbf{X})$. The extension to higher dimensions is given at the end of the paper. We assume that model equations describing the system are known, although they need not be exact. We now employ a surface of section, and denote the coordinates in the surface of section by ξ, and the Poincaré map of ξ by

$$\xi_{n+1}=\mathbf{f}(\xi_n,\alpha)\,, \tag{1}$$

where the map $\mathbf{f}$ is necessarily invertible and α is a system parameter.[1] Suppose that we wish to go from a source point $\mathbf{X}_s$ to a small region about a target point, $\mathbf{X}_t$. Following the trajectory from $\mathbf{X}_s$ forward in time, we find its first intersection with the surface of section and denote this point ξ_s. Following the trajectory through $\mathbf{X}_t$ backward in time, we similarly determine its first intersection with the surface of section and denote this point ξ_t. Thus we have reduced our problem to that of a two-dimensional map in which we desire to go from ξ_s to the vicinity of ξ_t. We assume that the system parameter α is available for adjustment at each iterate. Thus we can replace α in the map (1) by α_n. However, we also suppose that only small adjustments of α are allowed. That is, $\alpha_n=\bar{\alpha}+\delta_n$, where $\bar{\alpha}$ is a nominal value of α and the deviation from $\bar{\alpha}$, denoted δ_n, is restricted to be small.

If the process under study is ergodic, then in the absence of perturbations (i.e., for $\alpha_n=\bar{\alpha}$) the time required to travel from a source point ξ_s to a small neighborhood $\bar{\varepsilon}_t$ of linear size ε_t about a target point ξ_t in the ergodic set is typically $\tau_0\sim 1/\mu(\bar{\varepsilon}_t)$, where μ denotes the natural probability measure of the chaotic set. The measure $\mu(\bar{\varepsilon}_t)$ typically scales with the information dimension,[2] D, so the time required obeys

$$\tau_0\sim(1/\varepsilon_t)^D\,, \tag{2}$$

for small ε_t. Thus in the absence of perturbations, the amount of time required to reach a desired target increases according to a power law as the size of the target region decreases. We will see that this power law can be converted to a much weaker logarithmic increase by applying a carefully chosen small perturbation.

Now suppose that perturbations of α can be applied. After one iteration of the return map, the change of the state, $\delta\xi$, relative to the point $\mathbf{f}(\xi_s,\bar{\alpha})$, due to a small perturbation, δ_1, is given by the Taylor expansion

$$\delta\xi\cong\left.\frac{\partial\mathbf{f}(\xi_s,\alpha)}{\partial\alpha}\right|_{\bar{\alpha}}\delta_1\,. \tag{3}$$

Letting δ_1 vary through a small interval, $|\delta_1|<\delta_*$, Eq. (3) defines a small line segment through the point $\mathbf{f}(\xi_s,\bar{\alpha})$. We denote this line segment $\overline{\delta\xi}$, and we denote its length $\delta\xi$. Since our system is chaotic, the length of the image of this line segment will grow roughly geometrically with each successive iteration of the map $\mathbf{f}(\xi,\bar{\alpha})$. Let n_1 denote the number of iterates required for the small line segment to stretch to a length of order 1. This typically happens when $\delta\xi\exp(n_1\lambda_1)\sim 1$ if $\delta\xi$ is small, where λ_1 is the positive Lyapunov exponent obtained for typical initial conditions on the attractor. Defining $\tau_1=\lambda_1^{-1}\ln(1/\delta\xi)$, the length of the line segment becomes of order 1 after about τ_1 iterates (i.e., $n_1\sim\tau_1$) if $\delta\xi$ is small. Without loss of generality, we take the size of the relevant ergodic region to be of the order of 1 so that τ_1 is approximately the number of iterates required for the line segment to span the ergodic region. Like-

13

wise, if we map the region $\bar{\varepsilon}_t$ backward in time, we find that its preimage spans the ergodic region after a number of preiterates which is typically of the order of $\tau_2 = |\lambda_2|^{-1}\ln(1/\varepsilon_t)$ if ε_t is small, where λ_2 is the negative Lyapunov exponent for the map f for typical initial conditions on the attractor.

Thus we adopt the following procedure. We iterate the segment $\overline{\delta\xi}$ forward using $a_n = \bar{a}$ for n_1 iterates until its length becomes of order unity. We then iterate the region $\bar{\varepsilon}_t$ backward for n_2 iterates until it first intersects the n_1 forward iteration of the line segment $\overline{\delta\xi}$. Typically for small $\delta\xi$ and ε_t we have $n_1 \cong \tau_1$ and $n_2 \cong \tau_2$. Iterating a point in the middle of this intersection backward n_1 times, we find a point on the line segment $\overline{\delta\xi}$ which is mapped to the target region $\bar{\varepsilon}_t$ in n_1+n_2 iterates. Knowing this point we can then determine the required perturbation δ_1 to be applied on the first iterate from ξ_s by using Eq. (3). Note that, for the situation considered so far, we assume no noise and no modeling error, and consequently we can achieve targeting with $\delta_n = 0$ for $n \geq 2$. For small $\delta\xi$ and ε_t the total time to go from ξ_s to $\bar{\varepsilon}_t$ by this method scales as

$$\tau = \tau_1 + \tau_2 = \lambda_1^{-1}\ln(1/\delta\xi) + |\lambda_2|^{-1}\ln(1/\varepsilon_t). \qquad (4)$$

For example, for $\delta\xi \sim \varepsilon_t$, we have $\tau \sim \ln(1/\varepsilon_t)$ in contrast to the power law (2).

In practice, we cannot actually iterate either the line $\overline{\delta\xi}$ or the region $\bar{\varepsilon}_t$. Rather, we iterate discrete approximations to them and make successive refinements until a sufficiently accurate intersection is obtained. We do this by putting a fixed number $\hat{N}_s \gg 1$ of equally spaced points on $\overline{\delta\xi}$, iterating these points, and joining their images with straight-line segments. Similarly, we iterate $\hat{N}_t \gg 1$ points on the perimeter of $\bar{\varepsilon}_t$ backward in time and join their images with straight-line segments. Once an intersection is detected, we refine its accuracy by repeatedly halving the intersecting forward and backward line segments and determining which of the halves actually contain the intersection. We must achieve accuracy sufficient to strike the n_2 backward iterate of the target, which is a long, thin region with width of order $\varepsilon_t \exp(-\lambda_1 n_2)$. Since we have normalized the size of the attractor to be of order 1, the curvature of the line $\overline{\delta\xi}(n_1)$ and the long sides of the n_2 backward iterate of $\bar{\varepsilon}_t$ are also of order 1. Thus, taking account of the curvature, to resolve the intersection within a distance of $\varepsilon_t \exp(-\lambda_1 n_2)$, we require that the distance between points on $\overline{\delta\xi}(n_1)$ and on the n_2 backward iterate of $\bar{\varepsilon}_t$ be of order $[\varepsilon_t \exp(-\lambda_1 n_2)]^{1/2}$. The square root results because the maximum distance between the curve and its discrete straight-line approximation is quadratic in the length of the straight-line segment. Each time we halve our line segments, we increase the resolution by a factor of 2 at the expense of including three additional points [one on $\overline{\delta\xi}(n_1)$ and one each on the two segments bounding the backward iterate of $\bar{\varepsilon}_t$ near the intersection]. Thus to resolve the intersection, we require a number of points N' additional to the original $\hat{N} = \hat{N}_s + \hat{N}_t$ points, where N' obeys

$$[\varepsilon_t \exp(-\lambda_1 n_2)]^{1/2} > 2^{-N'/3}. \qquad (5)$$

Using the relation $\varepsilon \exp(|\lambda_2| n_2) \sim 1$, this can be rewritten as

$$N' \gtrsim \tfrac{3}{2} D \ln(1/\varepsilon_t), \qquad (6)$$

where $D = 1 + \lambda_1/|\lambda_2|$ is the Lyapunov dimension of the attractor. We stress that $\hat{N}$ is fixed (typically we took $\hat{N} \sim 100$) and does not depend on ε_t or $\delta\xi$. Consequently, as ε_t is reduced, the required computational effort increases logarithmically in ε_t as shown in Eq. (6).

In order to show why our method of using forward and backward iterations was employed, we now contrast it with another conceivable procedure. If one iterated the line segment $\overline{\delta\xi}$ forward until it first intersected the region $\bar{\varepsilon}_t$, it would do so on iterate $n_1 + n_2$. One could then choose a point in this intersection, iterate the point backward $n_1 + n_2$ steps to find the corresponding point on the original line segment $\overline{\delta\xi}$, and then determine δ_1 from Eq. (3). While this works in principle, the numerical requirements of this pure forward scheme are needlessly more severe than when we determine an intersection by iterating $\overline{\delta\xi}$ forward n_1 steps and $\bar{\varepsilon}_t$ backward n_2 steps. In the pure forward method, to detect an intersection between the target and the $n_1 + n_2$ iterate of the source, we require that the approximation of $\overline{\delta\xi}(n_1+n_2)$ obtained by joining the N_s points with straight-line segments intersect the region $\bar{\varepsilon}_t$. Since the curvature of $\overline{\delta\xi}(n_1+n_2)$ is typically of order 1, we thus require

$$\delta\xi(n_1+n_2)/N_s < \varepsilon_t^{1/2}. \qquad (7)$$

The source line will have length unity after n_1 iterates, and will then expand by roughly $\exp(n_2\lambda)$ during the next n_2 iterates, where λ is the topological entropy.[3] So we require

$$N_s \gtrsim \varepsilon_t^{-1/2} \exp(n_2 \lambda) \qquad (8)$$

or

$$N_s \gtrsim (1/\varepsilon_t)^{\lambda/|\lambda_2| + 1/2}. \qquad (9)$$

Thus the number of points required by the pure forward method increases exponentially with $1/\varepsilon_t$, but only increases logarithmically with $1/\varepsilon_t$ in the forward-backward method.[4]

We now illustrate the method with a specific chaotic system. In particular, we deal with the Hénon map[5] in the form $x_{n+1} = a + 0.3 y_n - x_n^2$ and $y_{n+1} = x_n$, with $\bar{a} = 1.4$. As an initial example, we choose the target region to be a small square centered on ξ_t with edge length $\varepsilon_t = 0.0038$. We find that for a representative pair of source and target points, say, $\xi_s = (0.4772, -1.188)$ and $\xi_t = (0.1371, -1.328)$, without applying a perturbation, 6062 iterations are required before the orbit from ξ_s

16

strikes within the target neighborhood, $\bar{\varepsilon}_t$. However, if we are permitted to vary a by up to 1 part in 1000 about its nominal value, we find that our targeting method directs the trajectory to the target neighborhood in only twelve iterations.

To confirm the predicted logarithmic behavior in (4), the following numerical experiment was performed. Source and target locations were chosen at random with respect to the natural measure on the Hénon attractor. Then we fix a target size ε_t, and for each pair of source and target points, our targeting algorithm, described above, was applied. The total number of iterations required to go from ξ_s to $\bar{\varepsilon}_t$ was determined for each pair, and the results were then averaged over many source-target pairs. This process was repeated for several values of ε_t. The neighborhood $\bar{\varepsilon}_t$ was chosen to be a small square of edge length ε_t centered on ξ_t. The result of this experiment is shown in Fig. 1. The solid line of slope $\lambda_1^{-1}+|\lambda_2|^{-1}$, predicted by Eq. (4), is consistent with the data. Also shown as a dashed line is the power-law dependence expected without control from Eq. (2) with $D\cong 1.26$ (the information dimension for the Hénon attractor). The logarithmic dependence of the time to reach the target on ε_t with control shows dramatic improvement over the power-law dependence without control.

The preceding discussion demonstrates that targeting can be achieved for chaotic systems using only small controls. It remains to be shown, however, that the method discussed can be effective in the presence of noise or modeling errors. Thus we suppose that the real system obeys $\xi_{n+1}=\mathbf{g}(\xi_n,a)+\Delta_n$. Here we imagine that our model map $\mathbf{f}(\xi_n,a)$ is slightly in error, and that, unknown to us, the correct form is $\mathbf{g}(\xi_n,a)$. We, furthermore, allow small-amplitude random noise to disturb the system at each iteration as indicated by the term Δ_n.

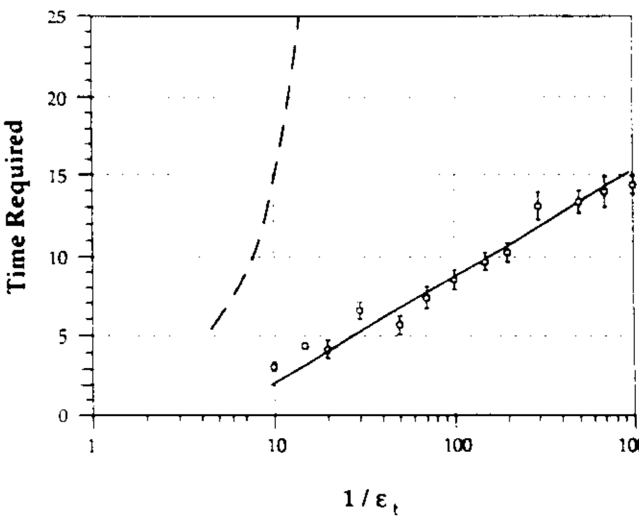


FIG. 1. Average time required to reach typical target neighborhood from typical source with control vs neighborhood size, ε_t ($\delta\xi=\varepsilon_t$). Solid line has slope predicted from Eq. (4); dashed line indicates expected behavior without control. Error bars are standard error for 25-point means.

To investigate the effect of noise alone, we take $\mathbf{f}=\mathbf{g}$. The following test was performed. Source and target locations were chosen at random on the Hénon attractor, and a trajectory between the source and the target neighborhood was found for the case without noise as previously described. As an example, the neighborhood size ε_t was chosen to be 0.01 and the time required to hit the target in the absence of noise and with only $\delta_1\neq 0$ was ten iterations. Then for each of the ten iterations, a random amount of noise was applied with Δ_n distributed uniformly in the interval $|\Delta_n|<\Delta_*$. As shown in Fig. 2 for the case $\Delta_*=0.01$, the noise displaced the tenth iteration to a point (denoted ξ_{10} in the figure) far away from ξ_t. Since the noise was applied at every iteration, we next compensated by recomputing the trajectory at every iteration and adjusting the applied perturbation correspondingly. That is, at *each* iterate we used the map $\mathbf{f}$ to determine δ_n by calculating the intersection of the forward iteration of the line determined from $\delta\xi_{n+1}=\delta_{n+1}\partial\mathbf{f}/\partial a$ with the backward iteration of the region $\bar{\varepsilon}_t$. The result of this procedure is shown in the inset of Fig. 2. The tenth iteration (denoted $\hat{\xi}_{10}$ in the figure) now lies within the target region. Thus we have shown that our method can be effective in the presence of small-amplitude noise provided that we apply a correction δ_n at each iterate.

It can similarly be shown that targeting can also be achieved even when the system is imperfectly modeled, i.e., when $\mathbf{f}$ differs slightly from the true map, $\mathbf{g}$. After each iteration, we have to compensate for the difference $\mathbf{g}-\mathbf{f}$. For example, let us consider the Hénon map for the case without noise, where $\mathbf{f}$ is the Hénon map with $\bar{a}=1.4$, and $\mathbf{g}-\mathbf{f}=0.014$; ε_t is still 0.01, and we use the same source and target as in our noise example (cf. Fig. 2). If we apply our procedure only on the first iterate

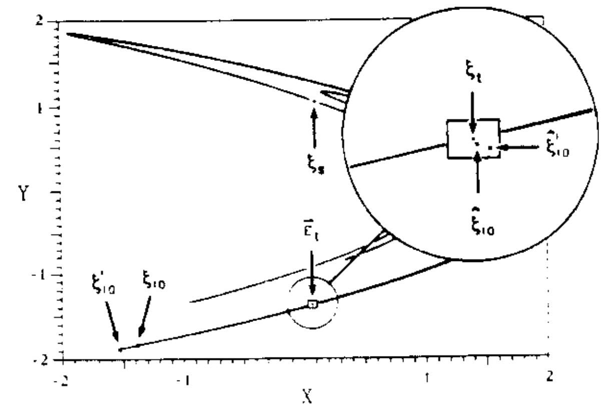


FIG. 2. Source ξ_s and target region $\bar{\varepsilon}_t$ on the Hénon attractor. Inset: When the targeting procedure is applied at every iteration, the noise or modeling error can be compensated for, and $\hat{\xi}_{10}$ and $\hat{\xi}'_{10}$ both lie within the target region.

from ξ_s (as we would if $\mathbf{f}=\mathbf{g}$), then the trajectory again ends at a point (denoted ξ'_{10} in Fig. 2) far from the target. As before, however, if our targeting algorithm is applied at every iteration, the tenth iterate (denoted $\hat{\xi}'_{10}$ in Fig. 2) arrives in the target neighborhood despite the modeling error.

To generalize our method to higher dimensions, consider that the map $\mathbf{f}$ is N dimensional and the attractor has k expanding and $N-k$ contracting directions at typical points. We note that when a k-dimensional surface and an $(N-k)$-dimensional surface intersect, generically they do so at isolated points, and small smooth perturbations will not destroy these intersections or create new ones. Thus, for a typical point and a typical small k-dimensional disk D^k centered at this point, the nth iterate of the disk $\mathbf{f}^n(D^k)$ will be a k-dimensional surface and its k-dimensional area will increase with n. Similarly, if we take an $(N-k)$-dimensional disk, D^{N-k}, centered at a typical point, then $\mathbf{f}^{-n}(D^{N-k})$ will be an $(N-k)$-dimensional surface whose area will increase with n. As these areas increase, typically they will intersect. We emphasize that targeting can typically be achieved with any dimensionality N even if we only have one available adjustable scalar parameter α. To see this we note the following. Consider a trajectory $\xi_i=\mathbf{f}^i(\xi_0,\alpha)$. If we perturb α from $\bar{\alpha}$ by an infinitesimal amount δ_i at time i, then at time $m>i$, a perturbation of ξ_m given by $\mathbf{v}_{i,m}\delta_i$ results, where $\mathbf{v}_{i,m}$ is an N-dimensional vector which is determined by the partial derivatives of the map along the trajectory. For typical ξ_0 and $\mathbf{f}$, the vectors $\mathbf{v}_{0,k},\mathbf{v}_{1,k},\ldots,\mathbf{v}_{k-1,k}$ are linearly independent and thus can be used to create the k disk D^k.

In conclusion, we have demonstrated that it is possible to rapidly reach a small, accessible target region in a chaotic system by applying small perturbations to an available parameter. The method used is robust against small-amplitude noise and small modeling errors, making it especially suited to practical applications. We emphasize that the problem addressed in this Letter is a very general one and can be expected to arise often.[6]

This research was supported by the U.S. Department of Energy (Scientific Computing Staff Office of Energy Research). The computation was done at the National Energy Research Supercomputer Center.

(a)Department of Physics.
(b)Laboratory for Plasma Research.
(c)Departments of Electrical Engineering and of Physics.
(d)Institute for Physical Science and Technology and Department of Mathematics.

[1]Alternatively, if model equations are not available, the map f might be derived from experimental data by use of the delay coordinate embedding technique [F. Takens, in *Dynamical Systems and Turbulence*, edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), p. 230; N. H. Packard *et al.*, Phys. Rev. Lett. **45**, 712 (1980); J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985)].

[2]See, for example, J. D. Farmer, E. Ott, and J. A. Yorke, Physica (Amsterdam) **7D**, 153 (1983).

[3]S. Newhouse, in *Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary (Springer-Verlag, Berlin, 1987), p. 2; Y. Yomdin, Isr. J. Math. **57**, 285 (1987). The length of a *small* line segment typically grows at the exponential rate λ_1. After the length becomes of order 1, however, it grows at the exponential rate λ.

[4]We note that some improvement can be obtained by using higher-order fitting (e.g., parabolic rather than linear) of the curves to the iterated points. The exponential and logarithmic dependences of the pure forward and the forward-backward methods remain, however.

[5]Similar results have been obtained for the area-preserving "standard map."

[6]For example, see E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990); in *Chaos*, edited by D. K. Campbell (American Institute of Physics, New York, 1990), pp. 153–172.

Communicating with Chaos

Scott Hayes[a]
U.S. Army Research Laboratory, Adelphi, Maryland 20783

Celso Grebogi[b],[c] and Edward Ott[b],[d],[e]
University of Maryland, College Park, Maryland 20742
(Received 18 December 1992)

The use of chaos to transmit information is described. Chaotic dynamical systems, such as electrical oscillators with very simple structures, naturally produce complex wave forms. We show that the symbolic dynamics of a chaotic oscillator can be made to follow a desired symbol sequence by using small perturbations, thus allowing us to encode a message in the wave form. We illustrate this using a simple numerical electrical oscillator model.

PACS numbers: 05.45.+b

Much of the fundamental understanding of chaotic dynamics involves concepts from information theory, a field developed primarily in the context of practical communication. Concepts from information theory used in chaos include metric entropy, topological entropy, Markov partitions, and symbolic dynamics [1]. On the other hand, because of their exponential sensitivity, chaotic systems are often said to evolve randomly. This terminology is partially justified if one regards the information obtained by *detailed* observation of the chaotic orbit as being less significant than the statistical properties of the orbits. The object of this Letter is to show that we can use the close connection between the theory of chaotic systems and information theory in a way that is more than purely formal. In particular, we show that the recent realization that chaos can be controlled with *small* perturbations [2] can be utilized to cause the symbolic dynamics of a chaotic system to track a prescribed symbol sequence, thus allowing us to encode any desired message in the signal from a chaotic oscillator. The natural complexity of chaos thus provides a vehicle for information transmission in the usual sense. Furthermore, we argue that this method of communication will often have technological advantages.

Specifically, assume that there is an electrical oscillator producing a large amplitude chaotic signal that one wishes to use for communication. The so-called double scroll electrical oscillator [3] yields a chaotic signal consisting of a seemingly random sequence of positive and negative peaks. If we associate a positive peak with a 1, and a negative peak with a 0, the signal yields a binary sequence. Furthermore, we can use *small* control perturbations to cause the signal to follow an orbit whose binary sequence represents the information we wish to communicate. Hence the chaotic power stage that generates the wave form for transmission can remain simple and efficient (complex chaotic behavior occurs in simple systems), while all the complex electronics controlling the output remains at the low-power microelectronic level.

The basic strategy is as follows. First, examine the free-running (i.e., uncontrolled) oscillator and extract from it a symbolic dynamics that allows one to assign symbol sequences to the orbits on the attractor. Typically, some symbol sequences are never produced by the free-running oscillator. The rules specifying allowed and disallowed sequences are called the *grammar*. Methods for determining the grammar (or an approximation to it) of specific systems have been considered in several theoretical [4] and experimental [5] works. (In the engineering literature, a similar concept exists in the context of constrained communication channels.) The next step is to choose a code whereby any message that can be emitted by the information source can be encoded using symbol sequences that satisfy suitable constraints imposed by the dynamics in the presence of the control. (The construction of codes with such constraints is a standard problem in information theory [6], and will be discussed in the context of communicating with chaos, along with the required generalizations, in a longer paper [7].) The code cannot deviate much from the grammar of the free-running oscillator because we envision using only tiny controls that cannot grossly alter the basic topological structure of the orbits on the attractor. Once the code is selected, the next problem is to specify a control method whereby the orbit can be made to follow the symbol sequence of the information to be transmitted. Finally, the transmitted signal must be detected and decoded.

We now present a simple numerical example illustrating how the preceding strategy is carried out. Figure 1(a) is a schematic diagram of the electrical circuit producing the so-called double scroll chaotic attractor [3]. The nonlinearity comes from a nonlinear negative resistance represented by the voltage v_R in Fig. 1. (Different realizations of the negative resistance are possible; we have constructed one using an operational amplifier circuit, and are designing an experiment using this oscillator to demonstrate information transmission using chaos.) The differential equations describing the double scroll system are

$$C_1\dot{v}_{C_1}=G(v_{C_2}-v_{C_1})-g(v_{C_1}),$$
$$C_2\dot{v}_{C_2}=G(v_{C_1}-v_{C_2})+i_L,$$
$$L\dot{i}_L=-v_{C_2}.$$

14

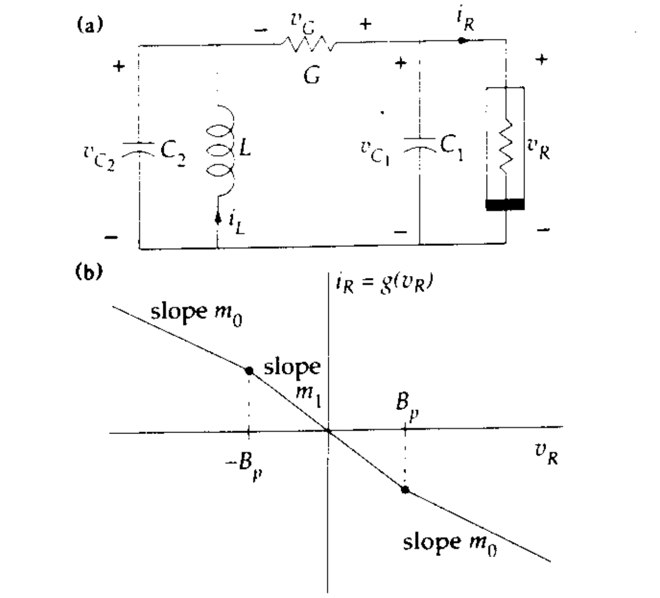


FIG. 1. Double scroll oscillator: (a) electrical schematic and (b) nonlinear negative-resistance i-v characteristic g.

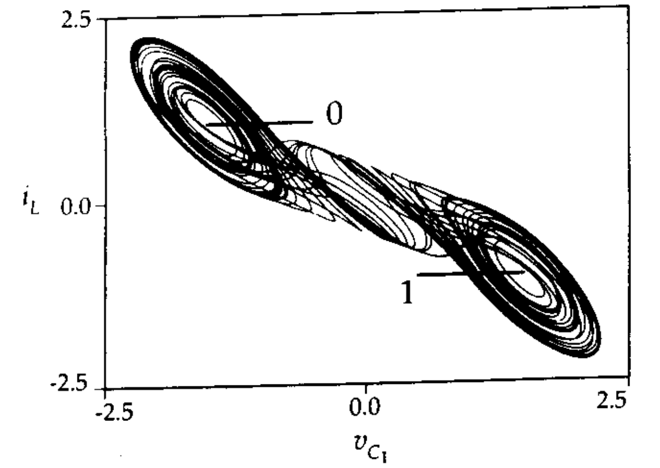


FIG. 2. Double scroll oscillator state-space trajectory projected on the i_L-v_{C_1} plane showing the two branches of the surface of section.

The negative-resistance i-v characteristic g is shown in Fig. 1(b). For our example, we use the normalized parameter values used by Matsumoto [3]: $C_1 = \frac{1}{9}$, $C_2 = 1$, $L = \frac{1}{7}$, $G = 0.7$, $m_0 = -0.5$, $m_1 = -0.8$, and $B_p = 1$. For a Poincaré surface of section (see Fig. 2), we take the surfaces $i_L = \pm GF$, $|v_{C_1}| \le F$, where $F = B_p(m_0 - m_1)/(G + m_0)$, so that these half planes intersect the attractor with edges at the unstable fixed points at the center of the attractor lobes. Figure 2 shows a trajectory of the double scroll system with the two branches of the surface of section labeled 0 and 1. (The plane surfaces are edge-on in the picture.) The intersection of the strange attractor with the surface of section is approximately a single straight line segment on each of the two branches. Let x denote the distance along this straight line segment from the fixed point at the center of the respective lobe, $x = (F - |v_{C_1}|)\cos\theta + |v_{C_2}|\sin\theta$, where θ is the angle that the line segment makes with the i_L-v_{C_1} plane. Because absolute values are used in defining x, we can use the same x coordinate for both lobes of the attractor.

To construct a description of the symbolic dynamics of the system, we run the computer simulation without control. When the free-running system state point passes through the surface of section, we record the value of the generalized coordinate x (restricted to 1000 discrete bins for the computer simulation), and then record the symbol sequence that is generated by the system after the state point crosses through the surface. Suppose the system generates the binary symbol sequence $b_1 b_2 b_3 \ldots$. We represent this by the real number $0.b_1 b_2 b_3 \ldots$, so that each symbol sequence corresponds to the real number $r = \sum_{n=1}^{\infty} b_n 2^{-n}$, and symbols that occur at earlier times are given greater weight. We refer to the number r, specifying the future symbol sequence, as the *symbolic state* of the system. This defines a function mapping the state-space coordinate x on the surface of section to the symbolic coordinate r. This function $r(x)$ (which we call the *coding function*) is shown in Fig. 3. (The function gives actual symbol sequences when referring to the 0 lobe, and the bitwise complement when referring to the 1 lobe.) Because the oscillator is only approximately described by a binary sequence, multiple values of x lead to the same future symbol sequence. (We only need to track one of them. More sophisticated techniques both for symbol assignment and symbol sequence ordering are discussed in the longer paper [7].) Because the intersection of the attractor with the surface of section is only approximately one dimensional, there is a slight uncertainty in the symbolic state for some values of x; this uncertainty is indicated by the shading in the regions between the upper and lower bounds on the value of r in Fig. 3. Observations of the time wave form produced by the oscillator suggest that the grammar is simple: Any sequence of

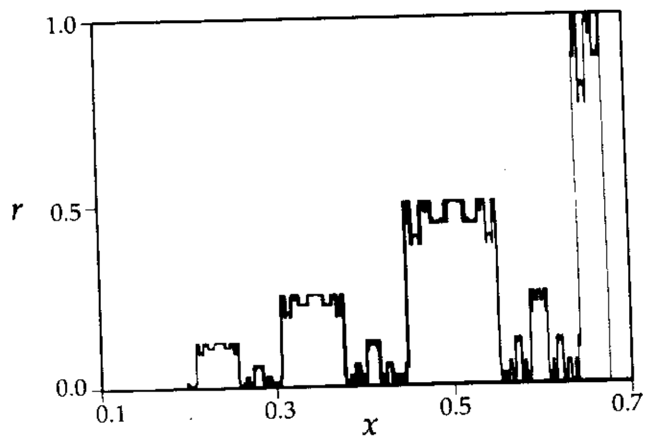


FIG. 3. Binary coding function $r(x)$ for the double scroll system.

18

binary symbols is allowed, except there can never be less than two oscillations of the same polarity. (We do not discuss the full grammar here, but instead adopt this simple grammar for simplicity of description.) This no-single-oscillation rule leads to a very simple coding: Insert an extra 1 after every block of 1's in the binary stream to be transmitted, and an extra 0 after every block of 0's. This altered data stream now satisfies the constraints of the grammar, and is uniquely decodable: Simply remove a 1 from every block of 1's upon reception, and a 0 from every block of 0's. Thus k oscillations of a given polarity represent $k-1$ information bits.

We now discuss how we control the system to follow a desired binary symbol sequence. Say the system state point passes through branch 0 of the surface of section (shown in Fig. 2) at $x=x_a$, and next crosses the surface of section (on either branch 0 or 1) at $x=x_b$. Because we have previously determined the function $r(x)$, we can use the stored values to find the symbolic state $r(x_a)$. We then convert the number $r(x_a)$ to its corresponding binary sequence truncated at some chosen length N, and store this finite-length symbol sequence in a *code register*. As the system state point travels towards its next encounter with the surface of section at $x=x_b$, we shift the sequence in the code register left, discarding the most significant bit (the leftmost bit), and insert the first desired information code bit in the now empty least significant slot (the rightmost slot) of the code register. We then convert this new symbol sequence to its corresponding symbolic state r_b'. Now, when the system state point crosses the surface of section at $x=x_b$, we use a simple search algorithm to find the nearest value of the coordinate x that corresponds to the desired symbolic state r_b'; call this x_b'. By construction, $|r(x_b)-r(x_b')| \leq 2^{-N}$. [If $r(x)$ is continuous, as in the Lorenz system, for example, this search can be replaced by a more efficient local derivative projection to find the desired value of x.] Now let $\delta x = x_b - x_b'$. Because we have chosen the branches of the surface of section at constant values of the inductor current i_L, the deviation δx in the generalized coordinate corresponds to a deviation in the voltages v_{C_1} and v_{C_2} across the two capacitors in Fig. 1. We thus apply a vector correction parallel to the surface of section (at constant i_L) along the attractor cross section to put the orbit at $x=x_b'$. This small correcting voltage perturbation is given by $\delta v_{C_1} = \pm\delta x\cos(\theta)$, $\delta v_{C_2} = \pm\delta x \sin(\theta)$, where the + signs are used for lobe 1 of the attractor, and the − signs for lobe 0. We plan to do this experimentally with current pulse generators connected in parallel with each capacitor. (Many methods of applying control perturbations are possible, but this one is particularly straightforward.) On each successive pass through the surface of section, a new code bit is shifted into the code register, and we repeat the procedure to correct the state-space coordinates, and thus the symbolic state, of the system. The coded information sequence, because it is shifted through the code register, does not begin to appear in the output wave form until N iterations of the procedure, where N is the length of the code register. If the symbol sequence is coming from a properly coded discrete ergodic information source, the process of shifting the information sequence through the code register can be viewed as locking the symbolic dynamics of the oscillator to the information source. Thus, there is a short transient phase during which the symbolic dynamics of the oscillator is being locked to the information source, and the symbolic dynamics of the oscillator is always N bits behind the information source.

Figure 4 shows an encoded wave form for the double scroll system produced by the described technique. This wave form corresponds to the voltage wave form $v_G(t)$ across the passive conductance G. If the conductance G is replaced by a transmission channel of the same impedance, the signal produced can be transmitted through the channel. We have represented each letter of the Roman alphabet by the five-bit binary number for its location in the alphabet, and added the extra bits to satisfy the no-single-oscillation constraint to encode the word "chaos." We have applied the technique to first bring the system to a periodic orbit about lobe 1 of the attractor, then to execute the writing of the word, and then to bring the system back to a periodic orbit about lobe 0. The trajectory shown in Fig. 2 is actually the encoded trajectory, but this is not apparent in the figure because the controlled trajectory approximates a possible natural trajectory. The root-mean-squared amplitude of the control signal over the writing of the word was of order 10^{-3} in the normalized units. The control probably cannot be made much smaller using this simple technique, primarily because the one-dimensional approximation in the surface of section causes the coding function to be slightly inaccurate. This control amplitude, though already very small compared to the oscillator signal voltages, does not

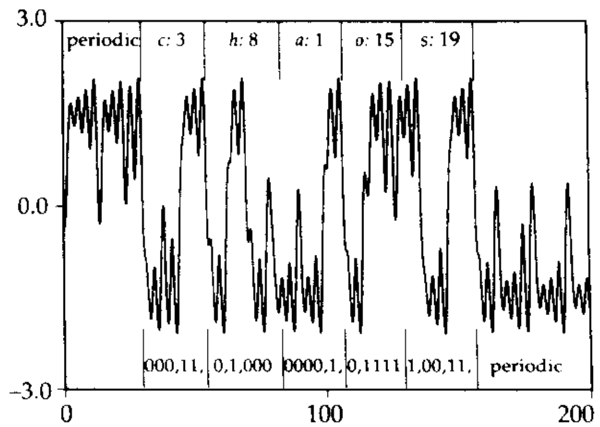


FIG. 4. Controlled $v_G(t)$ signal for the double scroll system encoded with "chaos." Each letter is shown at the top of the figure, along with its numerical position in the alphabet. Shown at the bottom are the corresponding binary code words. Extra bits (indicated by commas) are added to satisfy the constraints imposed by the grammar.

19

appear to be a fundamental limit, and we are developing control techniques to reduce it.

We conclude with some comments concerning the scope, application, and theoretical significance of our technique.

(1) Since we envision the transmitted signal to be a single scalar, its instantaneous value does not specify the full system state of the chaotic oscillator. If more state information is needed to extract the symbol sequence, time delay embedding [8] can be used. As our example using the double scroll equations shows, however, time delay embedding is not always necessary.

(2) Because our control technique uses only small perturbations [9], the dynamical motion of the system is approximately described by the equations for the uncontrolled system. Knowing the equations of motion greatly simplifies the task of removing noise [10] from a received signal. The basic bipolar nature of the signal in Fig. 4 implies that the message can still be extracted for noise amplitudes that are significant, but not too large compared to the signal. We consider the effects of additive noise on the detection of chaos signals quantitatively in the longer paper [7].

(3) Signals that are generated by chaotic dynamical systems and carry information in their symbolic dynamics have an interesting and possibly useful property: More than one encoded symbol can be extracted from a single sample of the trajectory if time delay embedding is used. This is done by using the state-space partition for a higher order iterate of the return map [7] of the system.

(4) Much of the theory needed to understand information transmission using the symbolic dynamics of chaotic systems already exists [11]. For example, because the topological entropy [12] of a dynamical system is the asymptotic growth exponent of the number of finite symbol sequences that the system can generate (given the best state-space partition), the channel capacity of a chaotic system used for information transmission is given by the topological entropy. The types of channel constraints that arise with a chaotic system will be discussed in a longer paper [7], along with other theoretical considerations.

(5) We emphasize that the particular methods for control and coding used in our double scroll example were chosen for simplicity, and that other more optimal methods are possible. Also, the double scroll oscillator itself was chosen because it is simple, and a large body of research is available about its dynamics. It is not intended as an example of a practical oscillator for communication wave form synthesis. It may be possible to use a higher-dimensional radio-frequency band-limited chaotic system for improved performance (higher information rate and better noise immunity), roughly analogous to the use of complex signaling constellations in classical communication systems. We are now developing more practical high-speed symbolic control techniques that could be used at higher bit rates than an implementation of the straightforward example given here.

(6) There has been much discussion of the role of chaos in biological systems, and we speculate that the control of chaos with tiny perturbations may be important for information transmission in nature.

This research was supported by Harry Diamond Laboratories (now the U.S. Army Research Laboratory), and by the U.S. Department of Energy (Office of Scientific Computing, Office of Energy Research).

[a] Also at Department of Physics and Astronomy, University of Maryland, College Park, MD 20742.
[b] Laboratory for Plasma Research.
[c] Institute for Physical Science and Technology and Department of Mathematics.
[d] Department of Physics and Astronomy.
[e] Department of Electrical Engineering.

[1] For a pedagogical discussion emphasizing the role of concepts from information theory in chaos, see R. Shaw, Z. Naturforsch. **36A**, 80 (1981).
[2] (a) E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990); (b) T. Shinbrot, E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **65**, 3215 (1990).
[3] Several articles about this circuit appear in Proc. IEEE **75**, No. 8 (1987): e.g., T. Matsumoto, p. 1033.
[4] P. Cvitanović, G. Gunaratne, and I. Procaccia, Phys. Rev. A **38**, 1503 (1988); P. Grassberger, H. Kantz, and U. Moenig, J. Phys. A **22**, 5217 (1989).
[5] D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989); L. Flepp, R. Holzner, E. Brun, M. Finardi, and R. Badii, Phys. Rev. Lett. **67**, 2244 (1991).
[6] R. E. Blahut, *Principles and Practice of Information Theory* (Addison-Wesley, Reading, MA, 1988); C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Chicago, 1963).
[7] S. Hayes, C. Grebogi, and E. Ott (to be published).
[8] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand, Springer Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, New York, 1981), p. 366.
[9] Our control technique can also be used to target a chaotic system [2(b)] in state space. The relationship between symbolic states and state-space coordinates is established by the coding function $r(x)$.
[10] E. J. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988); S. M. Hammel, Phys. Lett. A **148**, 421 (1990). It is easier to filter noise that is introduced in the communication channel than it is to filter noise that is present in the chaotic oscillator itself.
[11] A very different concept is that of using the synchronization of two chaotic systems [L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990)] for secure communications. In this case a small information-bearing signal is masked by a large chaotic signal. Several papers discussing this appear in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 1992). This mechanism is not, however, based on the information theoretic formalism of chaos.
[12] R. L. Adler, A. G. Konheim, and M. H. McAndrew, Trans. Am. Math. Soc. **114**, 309 (1965).

20

Experimental Control of Chaos for Communication

Scott Hayes,[1,*] Celso Grebogi,[2,†,‡] Edward Ott,[2,*,†,§] and Andrea Mark[1]
[1]*U.S. Army Research Laboratory, Adelphi, Maryland 20783*
[2]*University of Maryland, College Park, Maryland 20742*
(Received 6 May 1994)

The use of chaos to transmit information is demonstrated experimentally. The symbolic dynamics of a chaotic electrical oscillator is controlled to carry a prescribed message by use of extremely small perturbing current pulses.

PACS numbers: 05.45.+b

It has been argued on theoretical grounds that it is possible to use the close connection between chaos and information theory for communication [1]. The realization that chaos can be controlled by small perturbations [2] leads to the idea that chaotic systems can be caused to produce a signal bearing desired information. In this Letter, we report the first *experimental* verification that communicating with chaos is indeed feasible, and that one can utilize the naturally occurring chaotic orbits on the attractor to carry information. We demonstrate experimentally that we can make the symbolic dynamics of a chaotic electrical oscillator follow a prescribed symbol sequence; thus we can encode and transmit any desired message [3].

To understand the ideas behind our experiment, consider a simple electrical oscillator working in a nonlinear regime and producing a large-amplitude chaotic signal consisting of a seemingly random sequence of peaks. Now imagine that one can use small perturbing current pulses to cause the peaks to fall above or below a preassigned threshold value in a desired order. The peaks that fall above the threshold are assigned the binary symbol 1, and those that fall below are assigned a 0. By controlling this *symbol sequence*, we can encode a desired message in the signal. The message can thus be transmitted through a signal transmission path, and a receiver can extract the message by observing the sequence of peaks relative to the threshold amplitude. From the practical point of view, we envision that the power oscillator that generates the information-bearing signal can be simple and efficient, while the small controlling current pulses can be produced by a low-power microelectronic circuit.

Figure 1(a) is a schematic diagram of the electrical circuit used in our experiment [4]. The nonlinearity comes from a nonlinear negative resistance represented by the voltage v_R. The negative resistance i-v characteristic g is shown in Fig. 1(b). The frequency of oscillation is approximately 5 kHz. The capacitor C_1 is variable so that we can tune the circuit for various types of behavior. For this experiment, we tuned the circuit to produce a Rössler band. An uncontrolled experimental trajectory is shown in Fig. 2.

The first step toward controlling symbol sequences is to experimentally obtain a description of the symbolic dynamics of the uncontrolled oscillator. Because in practice some symbol sequences are not produced by the free-running oscillator, the rules specifying the allowed symbol sequences (the *grammar*) need to be determined. Methods for determining the grammar of chaotic systems are the topic of much current research in symbolic dynamics [5]. We first discuss our method for determining the symbolic dynamics and allowed symbol sequences. To determine the symbolic dynamics and to control the system, we use a computer-assisted measurement system that rapidly samples the circuit voltages and can perform all the computations necessary for control. The system can also provide current pulses to control the system. We will henceforth refer to this system as the *controller* [6].

To characterize the dynamics, we take samples of v_{C_1} on the Poincaré surface $v_{C_2} = 0$ as v_{C_2} passes through zero volts in a negative direction. The intersection of the

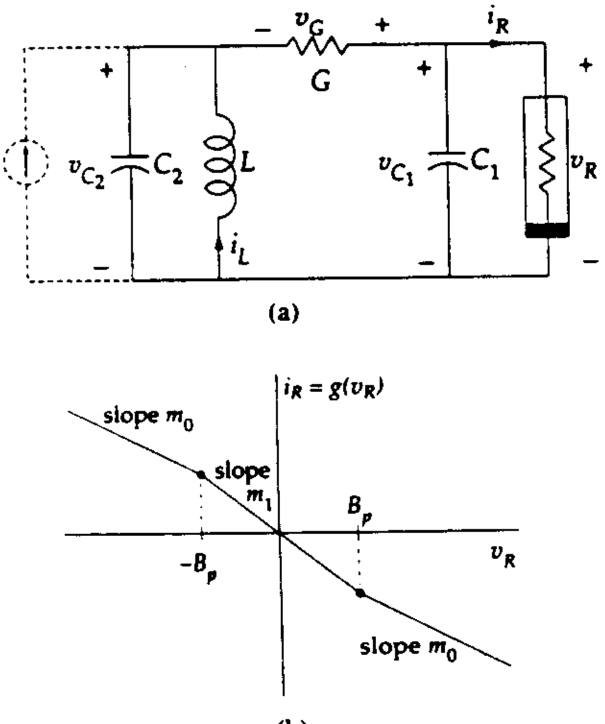


FIG. 1. Double scroll oscillator: (a) Electrical circuit with $L = 8.2$ mH, $C_1 = 0.0055$ μF, $C_2 = 0.05$ μF, and $1/G = 1.33$ kΩ. Controlled current pulses are injected by the source i_p. (b) Negative resistance i-v characteristic.

0031-9007/94/73(13)/1781(4)$06.00

21

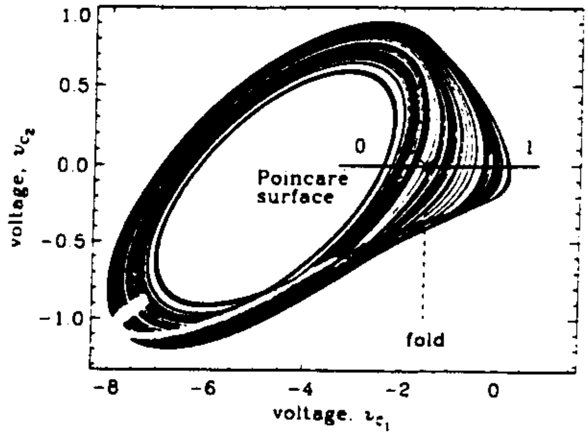


FIG. 2. Measured trajectory on Rössler-type attractor showing Poincaré surface, symbolic partition, and locus of control pulses. Units are volts.

attractor with the surface $v_{C_2} = 0$ is well approximated by a single thin arc, and thus we can describe the dynamics of v_{C_1} on the surface of section by one coordinate. Each time the voltage v_{C_2} (see Fig. 2) passes through zero in the negative direction, the controller is triggered (using the trigger circuit in an oscilloscope) to sample the value of v_{C_1}. We call the value of v_{C_1} *on the Poincaré surface the state point* $x = v_{C_1}|_{v_{C_2}=0}$. The natural way to partition the Poincaré surface into symbolic regions is to choose the partition boundary at the point on the surface of section which leads into the fold in the Rössler band under the action of the dynamics. (If the boundary is not taken at the fold, then two different state points can generate the same symbol sequence.) The symbol 0 is thus generated when the state point is to the left of the boundary shown in Fig. 2, and a 1 is generated when the state point is to the right of the boundary. The symbolic partition is shown on the Poincaré surface with respect to the state point that passes into the fold in Fig. 2. Hence, the voltages $x = v_{C_1}|_{v_{C_2}=0}$ that fall above the boundary correspond to a 1 and the voltages that fall below the boundary to a 0, and the dividing threshold is the partition boundary at the point that leads into the fold in Fig. 2. [In practice, if the transmitted signal is the scalar load voltage $v_G(t)$, one can detect the symbols by observing the *peaks* in this signal alone relative to a threshold voltage.]

Whenever the system state coordinate crosses the Poincaré surface, the state point $x = v_{C_1}|_{v_{C_2}=0}$ is stored in the controller. The controller then records the sequence of symbols generated after the crossing by determining whether the next ten crossings (including the present crossing) of the Poincaré surface are to the left or right of the partition boundary in Fig. 2. The ten-bit symbol sequence stored in the symbol register can be viewed as a binary number from 0 (all 0's) to 1023 (all 1's). We denote this number by r. We define an inverse coding function [1] $s(r)$ for r ranging from 0 to 1023, where $s(r)$ is the average value of the state point producing the symbol sequence specified by r. In practice, $s(r)$ is computed by following a long orbit and taking a running average of all x that produced the symbol sequence r. If the symbol sequence 1111111111 is produced by a trajectory passing through the state point $x = -1.5$ V, for example, we set the variable $s(1023) = -1.5$, where 1023 is the binary value of 1111111111. When another trajectory produces the same symbol sequence, perhaps at a voltage $x = -1.502$ V, we update the value to $s(1023) = -1.501$, the average of the state points for the two trajectories that produced this symbol sequence.

An experimental inverse coding function is shown in Fig. 3. This finite approximation of the inverse coding function represents a table with $2^{10} = 1024$ entries stored in controller memory of the state points s that correspond to each possible ten-bit symbol sequence r. Some of the $2^{10} = 1024$ ten-bit sequences are never produced by the free-running oscillator. This is typical of the symbolic dynamics of chaotic systems in general and is known as a constraint on the *grammar*, or, in the context of communication, it is known as a *channel constraint*. The type of constraint that occurs for our oscillator is a generalized type of *run-length constraint* [7]. (A simple run-length constraint places a maximum and minimum on the allowed length of strings of 1's and the allowed length of strings of 0's appearing in the code.)

To be able to control the system and hence produce a signal carrying a desired message by use of small perturbations, we must incorporate the effect of perturbations on our description of the dynamics. We coupled a digital-to-analog converter in our controller to the positive terminal of capacitor C_2 shown in Fig. 1(a) in series with a large buffer resistor, thus providing a simple method for injecting small current pulses. The current pulse generator is represented schematically as dashed lines in Fig. 1. These pulses had an approximate duration of 4 μs as compared to the typical cycle time of the chaotic oscillator of about 200 μs.

In our experiment we applied controlling current pulses 80 μs after each crossing of the surface of section. The

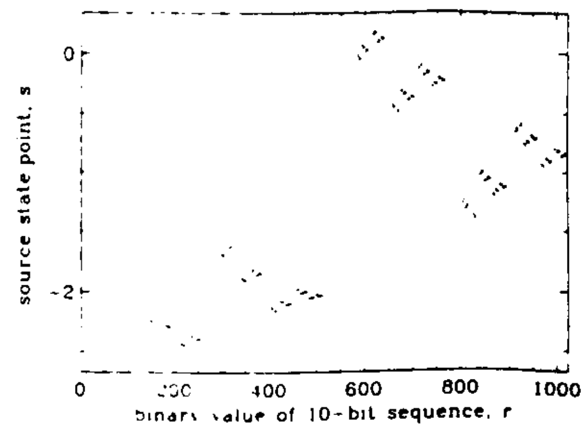


FIG. 3. Binary inverse coding function. The binary number corresponding to each binary sequence is given by its base-10 value.

22

distance around the attractor at which a pulse occurs varies depending on the average velocity of the phase space flow along a given orbit. The location on the attractor at which pulses occur is indicated in Fig. 2 by the gray swath at the bottom of the attractor. The key to our control procedure is the realization that the application of a pulse affects a very small change in the *state point* that produces a *given* symbol sequence. Because the pulse amplitudes are very small, this change Δx is approximately linear in the pulse amplitude p. That is, $p = d\Delta x$, where in general the proportionality constant d depends on x, or, equivalently, on r. We calculate $d(r)$ as follows. After a state point x is sampled, the pulse is applied using a small reference value $p = p_r$, and ten symbols (including the one corresponding to x) are shifted into the symbol register in the same way as before. Another reference pulse is not applied until this ten-bit sequence has been accumulated. The procedure is then repeated. In this way, we obtain the functional relationship between x and r in the presence of the reference pulse of amplitude $p = p_r$. We denote this relationship by $w(r)$, where w is the value of x that produces symbol sequence r in the presence of the reference perturbation. The proportionality constant $d(r)$ is then given by $d(r) = p_r/[w(r) - s(r)]$. We can thus compute the *required* pulse amplitude to affect a *desired* small change in the value of the state point by $p = d(r)[x - s(r)]$, where x is the actual state point upon crossing the Poincaré surface and r is the *desired* future symbol sequence, which will be specified by the message bits. We thus compute the required pulse amplitude using only one subtract and one multiply, a highly efficient algorithm that could be used in a *microelectronic controller* at much higher speeds than in the experiment described here.

We now discuss the procedure used to cause the oscillator to track a prescribed binary symbol stream. We now have, in controller memory, the function $s(r)$ containing the state points that will produce each allowed symbol sequence if no perturbation is applied and a function $d(r)$ that we use to compute the required control pulse amplitude. We now switch the controller from the *learn* phase to the *control* phase. When the system coordinate passes through the Poincaré surface, the controller samples v_{C_1} and obtains the first state point x_0. The controller then loads the symbol register with the symbol sequence that will naturally evolve if no control perturbation is applied; that is, the value in the symbol register is set to the value of r corresponding to the array element $s(r) = x_0$. The symbol register now contains ten bits that are determined not by our message, but by the state point that happens to occur the first time. While the trajectory travels toward its next encounter with the Poincaré surface, the controller shifts the symbol register left and inserts the first message bit into the rightmost (least significant bit) slot. The leftmost (most significant) bit is discarded; this bit corresponds to the symbol that was just produced and is no longer needed. The symbol sequence now appearing in the symbol register is the one that we *want* the oscillator to produce after the upcoming encounter with the Poincaré surface. Corresponding to this *desired* symbol sequence, r_1 is the *desired* state point $x_1 = s(r_1)$ that will produce this sequence. It is unlikely, however, that the state point x_1 will exactly correspond to $s(r_1)$, so the controller must apply a perturbation to correct the trajectory. When the trajectory passes through the Poincaré surface again, the error in the state point $\Delta x_1 = x_1 - s(r_1)$ is computed. The controller then applies the first control perturbation $p_1 = \Delta x_1 d(r_1)$. The controller then shifts the symbol register left and places the next message bit into the rightmost slot, discarding the leftmost bit corresponding to the symbol just produced. As each successive state point x_n becomes available, the controller looks up the target state point $s(r_n)$, where r_n is the *desired* symbol sequence in the symbol register, computes the error $\Delta x_n = x_n - s(r_n)$, applies a control pulse $p_n = \Delta x_n d(r_n)$, and shifts one new message bit into the symbol register. This procedure is repeated continuously as long as the system is under control. Because we first load the symbol register with the symbol sequence that will naturally evolve without control, the first ten iterates of the control procedure target the system from chaos into the information transmission.

As previously mentioned, the constraint on the grammar for our oscillator is a *generalized* run-length constraint, but we can satisfy the grammar with a *simple* run-length constraint such that no runs of more than one 0 occur. The grammar can be satisfied with this simple constraint because the dynamics of the Rössler system is well approximated by a one-dimensional single-hump map [7]. We have encoded the message [8], "Yea, verily, I say unto you: A man must have chaos yet within him to birth a dancing star. I say unto you: You have yet chaos in you." We used the standard seven-bit ASCII computer standard for the characters in the message and then mapped the message bits to code bits according to

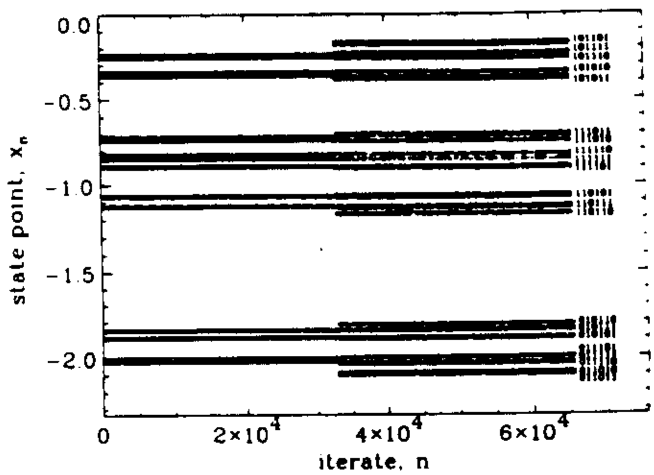


FIG. 4. State-point sequence with encoded message showing the Cantor-set structure produced by the constrained code during control with two different encoded messages.

23

the rule $0 \rightarrow 01$, $1 \rightarrow 11$, and thus no more than one 0 will ever occur in a row. (One could also use the variable-length code $0 \rightarrow 01$, $1 \rightarrow 1$ to improve the rate of information transmission.) To demonstrate the effect of the control symbol sequence on the signal, we have also produced another "message"—a random sequence of 0's and 1's constrained to no more than one 0 in a row. A state point sequence corresponding to the repeated [9] transmission of the Nietzsche quotation followed (starting at cycle $n \approx 3.3 \times 10^4$) by transmission of the random bit stream is shown in Fig. 4. Because of the overly restrictive simple run-length constraint, the state points x_n fall within bands on the Poincaré surface during transmission of both messages. (The rms control current during the whole sequence was 0.2 μA; compare this to circuit currents of a few milliamps.) Both binary streams cause the sequence of state points to approximately lie on a Cantor set [10]. This occurs because an overly constrained *symbol* sequence also constrains the *state-point* sequence. Because the binary sequence representing the quotation is *more* restrictive than the random bit sequence (because the code $0 \rightarrow 01$, $1 \rightarrow 11$ used to satisfy the run-length constraint is itself more restrictive than necessary), the signal is confined to *narrower* bands during the quotation than during the random bit sequence.

An important aspect of this type of signal is that more than one information bit can be extracted from a single state-point sample—the Cantor-set structure of the signal consists of bands that represent *sequences* of symbols not just single encoded binary symbols. One can visually resolve bands corresponding to up to six-bit-long binary sequences in our experimental signal; we have labeled these sequences in Fig. 4. Only one *new* code bit becomes available for each cycle of the system, but more than one code bit can be extracted from one state point sample. A signal such as this could thus be detected and decoded by only sampling once every three cycles of the signal, for example, which would give the observer three code bits per sampled state point.

In conclusion, the experimental results reported support the feasibility of communicating with chaos by controlling the symbolic dynamics of a chaotic oscillator and demonstrate some important features of this method of communication.

This research was supported by the U.S. Army Research Laboratory and the U.S. Department of Energy (Office of Scientific Computing, Office of Energy Research).

*Department of Physics and Astronomy, University of Maryland, College Park MD.

†Laboratory for Plasma Research.

‡Institute for Physical Science and Technology and Department of Mathematics.

§Department of Electrical Engineering and Institute for Systems Research.

[1] S. Hayes, C. Grebogi, and E. Ott, Phys. Rev. Lett. **70**, 3031 (1993); S. Hayes and C. Grebogi, Mater. Res. Soc. Symp. Proc. **203B**, 153 (1993).

[2] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).

[3] A different use of chaos in communication is the use of chaos synchronism for the purpose of secure (i.e., difficult to decipher by an interceptor) transmission of information, see L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990); K. M. Cuomo and A. V. Oppenheim, Phys. Rev. Lett. **71**, 65 (1993).

[4] T. Matsumoto, Proc. IEEE **75**, 1033 (1987). For further information about chaos in electronic circuits, see, for example, P. S. Lindsay, Phys. Rev. Lett. **47**, 1349 (1981); C. Jeffries and J. Perez, Phys. Rev. A **27**, 601 (1983); S. K. Brorson, D. Dewey, P. S. Linsay, *ibid.* **28**, 1201 (1983); H. Izeki, J. S. deGrasse, T. H. Jensen, *ibid.* **28**, 1207 (1983); R. W. Rollins and E. R. Hunt, *ibid.* **29**, 3327 (1984).

[5] P. Cvitanović, G. Gunaratne, and I. Procaccia, Phys. Rev. A **38**, 1503 (1988); P. Grassberger, H. Kantz, and U. Moenig, J. Phys. A **22**, 5217 (1989); D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989); L. Flepp, R. Holzner, E. Brun, M. Finardi, and R. Badii, Phys. Rev. Lett. **67**, 2244 (1991); G. B. Mindlin, H. G. Solari, M. A. Natiello, R. Gilmore, and X.-J. Hou, J. Nonlinear Sci. **1**, 147 (1991).

[6] The controller is an HP3567A signal analyzer containing a high-speed signal processing computer and is used to digitize the incoming signal, compute the control algorithm, and provide a control pulse within one cycle of the oscillator.

[7] An introductory discussion of codes for constrained communication channels can be found in R. E. Blahut, *Principles and Practice of Information Theory* (Addison-Wesley, Reading, MA, 1988). The general constraint on symbol sequences in systems described by single-hump maps will be discussed in a longer paper (to be published).

[8] Friedrich Nietzsche, *Thus Spake Zarathustra*. We have this liberal yet appealing translation of the German "Ich sage euch: man muß noch Chaos in sich haben, um einen tanzenden Stern gebären zu können. Ich sage euch: ihr habt noch Chaos in euch." from the frontispiece of the book by Manfred Schroeder, *Fractals, Chaos, Power Laws—Minutes from an Infinite Paradise* (W. H. Freeman and Company, New York, 1991). The more literal translation by Thomas Common (Random House Modern Library, New York) is "I tell you: one must still have chaos in one, to give birth to a dancing star. I tell you: ye have still chaos in you."

[9] For the first 3.3×10^4 cycles, the oscillator repeats the transmission of the Nietzsche quotation. There are 133 characters in the quotation, thus 7×133 ASCII bits, and $7 \times 133 \times 2 = 1862$ code bits. The encoded quotation thus appears in the signal about 17 times during the first half of the controlled state-point sequence.

[10] This signal represents the first experimental stabilization of a Cantor-set invariant density.

84

Controlling Complexity

Leon Poon and Celso Grebogi
Institute for Plasma Research, University of Maryland, College Park, Maryland 20742
(Received 11 July 1995)

Complex systems have the property that many competing behaviors are possible, and the system tends to alternate among them. In fact, the ability of a complex system to access many different states, combined with its sensitivity, offers great flexibility in manipulating the system's dynamics to select a desired behavior. By understanding dynamically how some of the complex features arise, we show that it is indeed possible to control a complex system's behavior. This is illustrated by using the noisy double rotor map as a paradigm.

PACS numbers: 05.45.+b

Scientists have come to the realization that many naturally occurring systems are neither completely ordered and predictable nor completely random and unpredictable. In fact, the behavior of many systems in nature fall between these two opposite ends. The term "complexity" has been coined to denote the study of "complex systems," that is, systems with complicated and intricate features having both elements of order and elements of randomness. They arise in fields as diverse as biology, chemistry, computer science, geology, physics, and fluid mechanics. Some systems that exhibit apparent complex behavior are Rayleigh-Bernard convection [1], an extended optical system [2], neuronal activity [3], and fluidized beds [4].

But what is a complex system? While there is no general consensus on what constitutes a good quantitative measure of complexity, the following general traits are some that most agree a complex system often exhibits [5]. (i) A complex system is composed of many parts that are interrelated in a complicated manner. Usually, these intricate mutual relations result in some form of coherent structure. (ii) A complex system possesses both "ordered" and "random" behaviors. (iii) A complex system often exhibits a hierarchy of structures; that is, nontrivial structures exist over a wide range of time and/or length scales.

These complex features are generally, but not necessarily, found in systems with many degrees of freedom. Instead of evolving towards one dominating attracting set, which is common in lower dimensional systems, the interactions among the large number of attracting and unstable sets often result in rich and varied dynamics where many competing behaviors are possible. As a result, the dynamics of complex systems tend to alternate among these different behaviors and which of them are observed at a given time are often sensitive to minor perturbations. These two key attributes, accessibility to many states and sensitivity, present us with an opportunity to influence and manipulate a complex system's dynamics.

In this paper, we show how complexity can be realized and how its dynamics can be manipulated using small perturbations. Specifically, we use the double rotor system as a paradigm of a relatively low-dimensional dynamical system (as opposed to an extended system), which exhibits many characteristics typical of complex systems when it is subjected to random external noise. We argue that the three traits mentioned above are apparent in many of its behaviors. Furthermore, we also show how to use small amplitude feedback control to influence and manipulate the behavior of the system such that its trajectories, which formerly traversed the various attracting and unstable chaotic sets, will now be confined to a neighborhood of one of the attracting states of our choice.

The double rotor map is a rich dynamical system with many complex features [6]. For a wide range of parameters, the double rotor map has a multitude of periodic attractors (the coexistence of tiny chaotic attractors is also possible for some parameters). Moreover, only a small portion of the basins of attraction for these periodic attractors have "smooth" structure; that is, only initial conditions within small balls centered at each component of the periodic attractor unambiguously asymptote to their respective attractors. With the exception of these small open neighborhoods about the periodic attractors, the majority of phase space is occupied by fractal basin boundaries whose dimension ($\approx$3.998 for the parameters we study) is very close to the dimension of the phase space.

The complicated basin structures of the double rotor system plays an important role in the system's complex dynamics. The presence of unstable invariant sets embedded in the fractal basin boundaries has two appreciable effects on the system: long chaotic transient behavior and final state sensitivity [7]. Typical trajectories with initial conditions near the boundaries (which is highly probable since most of the phase space is dominated by basin boundaries) will undergo chaotic motion for long times before settling on one of the periodic attractors. The dynamics is then characterized by a large number of periodic attractors "embedded" in a sea of transient chaos. The fine scale intermingling among the various basins makes it very difficult to predict the future state of trajectories for arbitrary initial conditions [7]. Thus, the system is extremely sensitive to perturbations.

Because fractal basin boundaries permeate most of the phase space, the addition of small amplitude noise prevents the trajectories from settling into any of the stable

25

periodic behavior. What happens instead is that a trajectory will come close to one of the periodic attractors and stay in its neighborhood for some time. For this period of time, the trajectory's behavior is governed by the periodic attractor, and it is, thereby, ordered. If one takes a large number of trajectories near this periodic attractor and then a snapshot is taken after the noisy system has evolved for some time, one would see coherent structures in phase space in the vicinity of the periodic attractor as shown in Fig. 1. (Coherent structures are colored with darker shades. What the picture represents and how it is generated will be discussed shortly.) However, this ordered behavior, for a particular trajectory, is transitory, and noise will eventually move the trajectory out of this state into the fractal boundary region. The trajectory will then spend some amount of time within the massive basin boundary region executing an apparently chaotic motion before approaching the same or another periodic attractor. The period of time in the fractal basin boundaries corresponds to the trajectory's "random" behavior. For an ensemble of trajectories, these random structures are represented by the lighter shades in Fig. 1. Hence, a typical noisy trajectory alternates between intervals of chaotic motion and intervals of nearly periodic behavior as shown in Fig. 2. A physical system that suggests the above behavior was observed by Bergé and Dubois in a Rayleigh-Bernard convection experiment [1].

The dynamics we have just described correspond to traits (i) and (ii) expected in complex systems. The fractal basin boundaries with the embedded chaotic sets provide

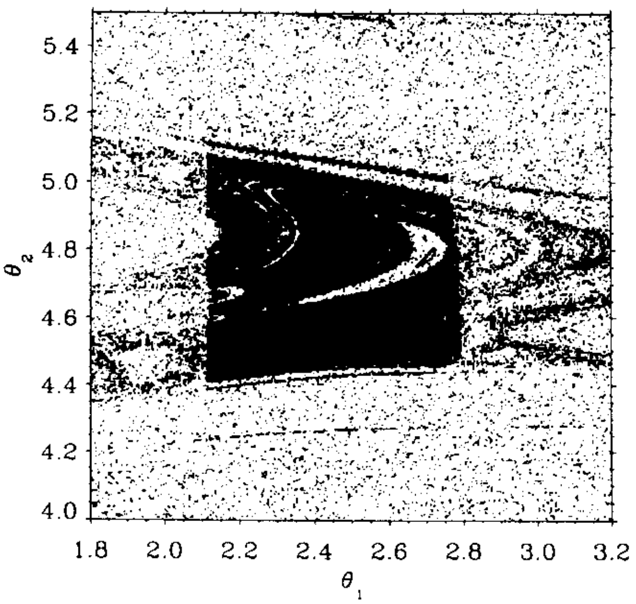


FIG. 1(color). A blowup of a region that exhibits a combination of coherent and random structures. The picture is generated after 100 iterations of the noisy map. The apparent sharp vertical boundaries in the picture is an artifact of projecting a four-dimensional figure onto a two-dimensional space, namely, $(\theta_1, \theta_2, 2.24, -3.65)$.

the link among the various attractors. The presence of both "ordered" and "random" behaviors mentioned in (ii) is evidenced by the trajectories of our system cascading from the random structures down to the coherent structures. Noise, on the other hand, displaces trajectories from the coherent structures back to the random structures. Therefore, dynamically there is such an interplay between random and coherent structures that the system is neither completely predictable nor completely random.

The coherent structures mentioned above and shown in Fig. 1 are found by examining the dynamics of a group of trajectories (10^5) near one of the periodic orbits. We follow the evolution of an ensemble of initial conditions in physical space (varying only θ_1 and θ_2) near one of the periodic attractors while the system is subjected to random noise with uniform distribution. In order to get an indication of whether a trajectory is evolving chaotically or has settled into some sort of periodic behavior after a certain number of iterates, say n, we compute the largest eigenvalue of the Jacobian matrix for each of the trajectories in the ensemble at the next iterate, i.e., $n + 1$. The largest eigenvalue is a measure of the local instability, and hence it indicates the "jump" the trajectory will make from its current location. We then assign a color to a point corresponding to the trajectory's initial condition according to how large a trajectory "jumped" at the $(n + 1)$th iterate, where the lighter the shade the bigger the jumps. Thus, a trajectory point with the largest eigenvalue less than 1 would be various shades of brown, while a trajectory point with the largest eigenvalue more than 1 would be various shades of white. In Fig. 1, we see an intricate picture exhibiting coherence and randomness. Thus, the combination of deterministic chaos and stochastic noise has resulted in the emergence of coherent structures.

We can quantify order and randomness by encoding the dynamics of the system into symbolic sequences. We choose to encode the trajectory by the sequence in which it visits the neighborhood of the attractors. Of course, the trajectory visits the different attracting sets by traversing through the chaotic sets in the boundaries. These chaotic excursions are implicit in our choice of the alphabet, each

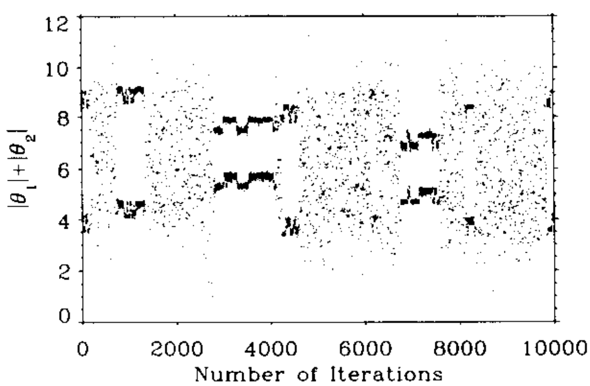


FIG. 2. Time series of a typical noisy trajectory.

26

symbol corresponding to an attractor. For the parameters we study, eight symbols are needed in the alphabet [8]. We compute now the "transition" probabilities among the various attracting sets. The complex dynamics can now be characterized by the Kolmogorov-Sinai (KS) entropy [9]

$$h = \lim_{n\to\infty} \frac{H_n}{n} = \lim_{n\to\infty} \frac{1}{n}\left(-\sum_{|S|=n} p(S)\ln p(S)\right),$$

where $S = s_1 s_2 \cdots s_n$ denotes a finite symbol sequence; $p(S)$, the relative frequency of S; and H_n, block entropy of block length n. Since $s_i \in \{1,\ldots,8\}$, sequences where the letters are completely uncorrelated would have a KS entropy value of $\ln 8 \simeq 2.08$, while a periodic sequence would yield a value of 0. Our computation shows that H_n/n converges fairly rapidly to a value of $h \approx 1.42$. KS entropy can be thought of as a measure of the coherence of a trajectory as it evolves, so an intermediate value of h means not only that the symbolic sequence is unpredictable (since $h > 0$), but there is also structure in the set of all possible symbol sequences. It should be noted that the encoding scheme we have implemented does not take into account the amount of time the trajectory spends in the ordered as well as the chaotic regions.

A consequence of the complex interplay between the coherent and random structures, and the irregular switchings among them, is the appearance of nontrivial length and/or time scalings in the noisy double rotor system, a quality [trait (iii)] that we expect in complex systems. First, there is the length of the chaotic transients that is a measure of how long a trajectory spends in the vicinity of each chaotic saddle embedded in the fractal basin boundary. It is known that the average length of a chaotic transient is related to the dimension and the Lyapunov exponents of the chaotic saddle [10]. Each chaotic saddle, in general, contributes a distinct time scale, and the overall chaotic transient $\langle\tau\rangle$ would then be a conglomeration of all these different time scales. Another relevant measure of time is the mean escape time $\langle T\rangle$ for a trajectory to leave the neighborhood of an attracting set, and it will in general be different for the different attractors as can be seen in Fig. 3.

After establishing the complex features of the double rotor system, we proceed to show how this knowledge can help us in manipulating and controlling the behavior of this complex system. Unlike low-dimensional chaotic systems that are commonly controlled using the ideas introduced in Ref. [11], complex systems are not characterized by the existence of one large chaotic attractor but by the coexistence of many attractors. While the existence of a large chaotic attractor is critical to those control schemes [11], we argue that for a complex system, the unstable chaotic sets in the boundaries provide us with the necessary sensitivity and flexibility to gear the dynamics toward a specific periodic behavior using small perturbations. We can elect to stabilize an unstable periodic orbit embedded within a chaotic saddle in the boundary [12] or stabilize one of the (metastable) attracting sets as we show next.

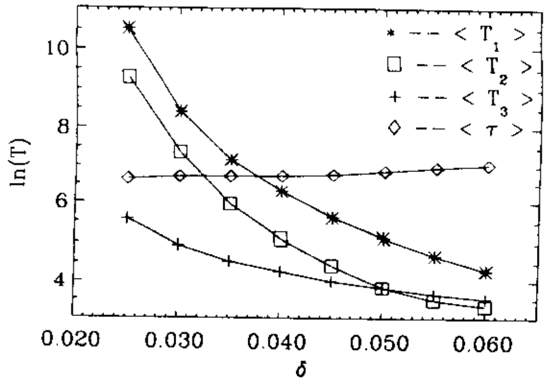


FIG. 3. The mean escape times $\langle T_i\rangle$ for some of the attractors and the average length of the chaotic transient $\langle\tau\rangle$ as a function of noise amplitude δ.

In controlling a metastable state in the complex double rotor map, we employ a simple feedback scheme. We denote the noiseless double rotor map as $\mathbf{x} \mapsto F(\mathbf{x})$ where $\mathbf{x}$ is the four-dimensional phase space coordinate, and the noisy double rotor map as $\tilde{F}(\mathbf{x}) \equiv F(\mathbf{x}) + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is the noise vector whose norm is bounded by δ. For simplicity, we assume the periodic orbit to be controlled is a fixed point (generalization to higher periodic orbit is fairly straightforward). If we labeled this fixed point as $\mathbf{x}_\star$, then in a neighborhood of it, we have the following linearization $F(\mathbf{x}_\star + \boldsymbol{\varepsilon}) = \mathbf{x}_\star + DF(\mathbf{x}_\star)\boldsymbol{\varepsilon}$, where the eigenvalues of $DF(\mathbf{x}_\star)$ are inside the unit circle (since $\mathbf{x}_\star$ is stable without noise). Suppose now that on the ith iterate, the trajectory lands in a neighborhood of this fixed point, so $\mathbf{x}_i = \mathbf{x}_\star + \boldsymbol{\varepsilon}$. Without control, $\mathbf{x}_i \mapsto \mathbf{x}_{i+1} = \tilde{F}(\mathbf{x}_i)$. However, assuming the linearization holds approximately for the noisy map near $\mathbf{x}_\star$, we can stabilize the fixed point with the addition of a controlling term, or $\hat{\mathbf{x}}_{i+1} = \mathbf{x}_{i+1} - DF(\mathbf{x}_\star)\boldsymbol{\varepsilon} = \tilde{F}(\mathbf{x}_i) - DF(\mathbf{x}_\star)(\mathbf{x}_i - \mathbf{x}_\star)$. Thus, the noisy trajectory with the above perturbation approaches the fixed point in due course. Since we want to achieve control using only small perturbations, the correction $|DF(\mathbf{x}_\star)\boldsymbol{\varepsilon}|$ is scaled when necessary so it will not exceed some predetermined upper bound of our choice.

Applying the correction to the noisy double rotor map, we control the dynamics of the system. In Fig. 4, we follow a typical noisy trajectory until it lands in a neighborhood of the desired metastable attractor we wish to control, then we turn on the control and let the system evolve for another thousand iterates in the neighborhood of the desired attractor. Control is then turned off, and we let the trajectory wander until it falls near the next desired metastable attractor. In this fashion, we stabilize, say, eight of the metastable attractors in the order we desire as demonstrated in Fig. 4. Furthermore, if the trajectory is caught in the vicinity of an attracting set that is undesirable, we can destabilize it by applying a small amount of noise. In fact, this was done in a brain experiment to control epilepsy [13] and a fluidized bed experiment to control slugging [14].

27

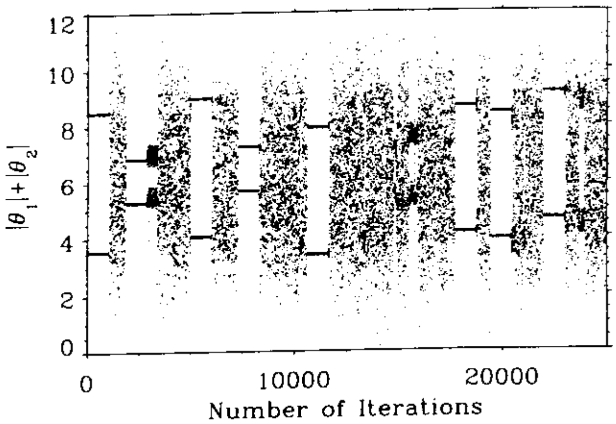


FIG. 4. Time series showing the result of applying the simple feedback control scheme to successively control eight different metastable states.

These two experiments demonstrate the general result that small perturbations, if chosen judiciously, not only can affect a desired outcome in these systems, but it also validates and extends these ideas to other systems.

In conclusion, we see that the myriad of possible behaviors in a complex system is of great utility if we are able to harness it. In fact, the ability of a complex system to access many different states, combined with its sensitivity, offers great flexibility in manipulating and controlling its dynamics. In general, many complex systems' behavior can be modified to suit our needs using only small perturbation strategies provided we are able to exploit their sensitivity.

This research was supported by a grant from DOE (Office of Scientific Computing). The numerical computations reported in this paper were supported in part by a grant from the W. M. Keck Foundation.

[1] P. Bergé and M. Dubois, Phys. Lett. **93A**, 365 (1983).

[2] F. T. Arecchi, G. Giacomelli, P. L. Ramazza, and S. Residori, Phys. Rev. Lett. **65**, 2531 (1990).

[3] P. E. Rapp *et al.*, J. Neuroscience **14**, 4731 (1994).

[4] C. S. Daw, C. E. A. Finney, M. Vasudevan, N. A. van Goor, K. Nguyen, D. D. Bruns, E. J. Kostelich, C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **75**, 2308 (1995).

[5] R. Badii and A. Politi, "Complexity: Hierarchical Structures and Scaling in Physics" (to be published); *Measures of Complexity*, edited by L. Peliti and A. Vulpiani (Springer-Verlag, Berlin, 1988); P. Grassberger, in *Fifth Mexican School on Statistical Mechanics*, edited by F. Ramos-Gómez (World Scientific, Singapore, 1991), p. 57; R. Badii, in *Chaotic Dynamics: Theory and Practice*, edited by T. Bountis (Plenum Press, New York, 1992), p. 1; J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989); J. Horgan, Sci. Am. **272**, No. 6, 104 (1995).

[6] The kicked double rotor consists of two connected massless rods (lengths l_1 and l_2) where one end of the first rod pivots about a fixed point, while the other end is attached to and pivots about the middle of the second rod. A mass m_1 is placed at the end of the first rod and two masses $m_2/2$ are placed at each end of the second rod. The coefficients of friction at the pivots are denoted ν_1 and ν_2. Kicks of constant strength ρ is applied to one end of the second rod at periodic intervals. The state of the system after each kick is described by the angular displacements and velocities of the rods $(\theta_1, \theta_2, \dot\theta_1, \dot\theta_2)$. The parameter values used in the study are $m_1 = m_2 = 1$, $\nu_1 = \nu_2 = 0.2$, $l_1 = 1/\sqrt{2}$, $l_2 = 1$, and $\rho = 6.5$. See the following references for more details: C. Grebogi, E. Kostelich, E. Ott, and J. A. Yorke, Physica (Amsterdam) **25D**, 347 (1987); F. Romeiras, C. Grebogi, E. Ott, and W. P. Dayawansa, Physica (Amsterdam) **58D**, 165 (1992); S. Dawson, C. Grebogi, T. Sauer, and J. A. Yorke, Phys. Rev. Lett. **73**, 1927 (1994).

[7] H. E. Nusse, C. Grebogi, E. Ott, and J. A. Yorke, in *Lecture Notes in Mathematics*, edited by J. C. Alexander (Springer-Verlag, New York, 1988), Vol. 1342, p. 220; C. Grebogi, S. W. McDonald, E. Ott, and J. A. Yorke, Phys. Lett. **99A**, 415 (1983).

[8] Actually there are 16 attractors for the parameters chosen, but, because of a simple symmetry in the double rotor system (Romeiras *et al.* in [6]), we can group the attractors into eight distinct pairs.

[9] A. N. Kolmogorov, Dok. Acad. Nauk SSSR **119**, 861 (1958); Ya. G. Sinai, Dok. Acad. Nauk SSSR **124**, 768 (1959).

[10] H. Kantz and P. Grassberger, Physica (Amsterdam) **17D**, 75 (1985); G.-H. Hsu, E. Ott, and C. Grebogi, Phys. Lett. A **127**, 199 (1988).

[11] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990); T. Shinbrot, C. Grebogi, E. Ott, and J. A. Yorke, Nature (London) **363**, 3 (1993), and references therein.

[12] Z. Kovács, K. G. Szabó, and T. Tél, in *Nonlinearity and Chaos in Engineering Dynamics*, edited by J. M. T. Thompson and S. R. Bishop (Wiley, Chichester, 1994).

[13] S. J. Schiff *et al.*, Nature (London) **370**, 615 (1994).

[14] C. S. Daw, C. E. A. Finney, and M. Vasudevan, in Proceedings of the Third Experimental Chaos Conference, edited by R. Harrison *et al.* (World Scientific, Singapore, to be published).

28

Using small perturbations to control chaos

Troy Shinbrot, Celso Grebogi, Edward Ott & James A. Yorke

The extreme sensitivity of chaotic systems to tiny perturbations (the 'butterfly effect') can be used both to stabilize regular dynamic behaviours and to direct chaotic trajectories rapidly to a desired state. Incorporating chaos deliberately into practical systems therefore offers the possibility of achieving greater flexibility in their performance.

CHAOTIC systems are characterized by extreme sensitivity to tiny perturbations. This characteristic, known as the 'butterfly effect', is often regarded as a troublesome property, and for many years it has been generally believed that chaotic motions are neither predictable nor controllable. The first reference we find to a differing view is due to John von Neumann, who is reported[1] to have stated around 1950 that small, carefully chosen, pre-planned atmospheric disturbances could lead after some time to desired large-scale changes in the weather. Although this specific application might be problematic, the basic idea of using chaotic sensitivity seems to have been clearly appreciated by von Neumann. Here we review recent work which demonstrates that the butterfly effect permits the use of tiny feedback perturbations to control trajectories in chaotic systems—a capability without a counterpart in nonchaotic systems. These considerations apply to cases in which the chaotic dynamics can in principle be defined by only a few variables, so systems where there are many active degrees of freedom (for example, the weather, and high-Reynolds-number flows) may not be tractable. However, we emphasize that cases of high (or infinite) dimensional systems for which the 'attractor' (and hence the dynamics) is low dimensional are common.

The research that we will review fits broadly into two categories. First we will discuss how, as proposed in ref. 2, one can select a desired behaviour from among the infinite variety of behaviours naturally present in chaotic systems, and then stabilize this behaviour by applying only tiny changes to an accessible system parameter (related work appears in refs 3-32, 77, 78). Moreover, we will show how one can switch between behaviours as circumstances change, again using only tiny perturbations. This means that chaotic systems can achieve great flexibility in their ultimate performance. Second, we will show how one can use the sensitivity of chaotic systems to direct trajectories rapidly to a desired state[33-38]. For example, a few years ago, NASA scientists used only small amounts of residual hydrazine fuel to send the spacecraft ISEE-3/ICE more than 50 million miles across the Solar System, thereby achieving the first scientific cometary encounter[39-43]. This feat was made possible by the sensitivity of the three-body problem of celestial mechanics to small perturbations, and would not have been possible in a nonchaotic system, in which a large effect typically requires a large control[44-46].

Stabilizing unstable orbits

One of the fundamental aspects of chaos is that many different possible motions are simultaneously present in the system. A particular manifestation of this is the fact that there are typically an infinite number of unstable periodic orbits that co-exist with the chaotic motion[47,48]. By a periodic orbit, we mean an orbit that repeats itself after some time (the period). If the system were precisely on an unstable periodic orbit, it would remain on that orbit forever. These orbits are unstable in the sense that the smallest deviation from the periodic orbit (for example due to noise) grows exponentially rapidly in time, and the system orbit quickly moves away from the periodic orbit. Thus, although these periodic orbits are present, they are not typically observed. Rather, what one sees is a chaotic trajectory which bounces around in an erratic, seemingly random fashion. Very rarely, the chaotic trajectory may, by chance, closely approach a particular unstable periodic orbit, in which case the chaotic trajectory would approximately follow the periodic cycle for a few periods, but it would then rapidly move away because of the instability of the periodic orbit. In addition to periodic orbits, it is common for continuous time dynamical systems to have unstable steady states embedded in chaotic motion (see our subsequent discussion of the Lorenz attractor). A ball placed exactly at the top of a hill is an example of an unstable steady state.

Although the existence of steady states and an infinity of different unstable periodic orbits embedded in chaotic motion is not usually obvious in free-running chaotic evolution, these orbits offer a great potential advantage if one wants to control a chaotic system. To demonstrate this, we adopt the following strategy[2]. First we examine the unstable steady states and low-period unstable periodic orbits embedded in the chaotic motion. For each of those unstable orbits, we ask whether the system performance would be improved if that orbit were actually followed. We then select one of the unstable orbits that yields improved performance. Assuming the motion of the free-running chaotic orbit to be ergodic, eventually the chaotic wandering of an orbit trajectory will bring it close to the chosen unstable periodic orbit or steady state. When this occurs, we can apply our small controlling perturbations to direct the orbit to the desired periodic motion or steady state. Moreover, if a small amount of noise is present, we can repeatedly apply the perturba-

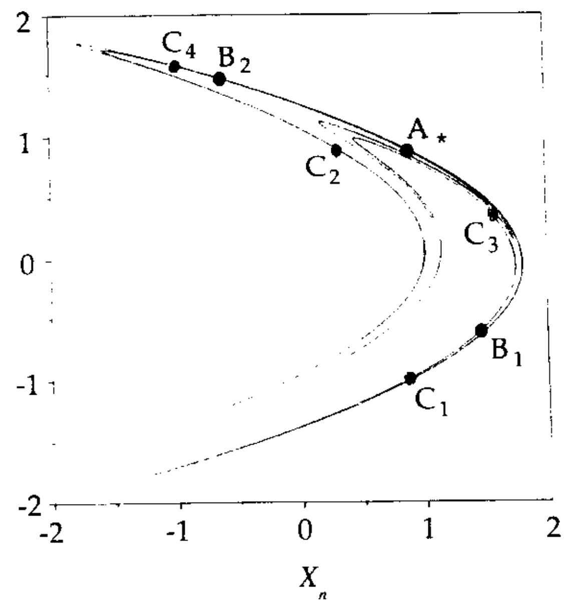


FIG. 1 Hénon attractor, with period-1 point, A_*; period-2 points, B_1 and B_2; and period-4 points, C_1, C_2, C_3 and C_4.

29

tions to keep the trajectory on the desired orbit. Thus small, carefully chosen perturbations are able to effect a large beneficial change in the long-term system behaviour. If, on the other hand, the system dynamics were not chaotic but were, say, stable periodic, then small controls could only change the orbit slightly. We would then be closely restricted to whatever system performance the stable periodic orbit gave, and we would have no option for improvement using small controls.

Furthermore, one may want a system to be used for different purposes or under different conditions at different times. If the system is chaotic, this requirement might be accommodated without altering the gross configuration of the system. In particular, depending on the use desired, the system could be optimized by switching the temporal programming of the small controls to stabilize different orbits. By contrast, in the absence of chaos, completely separate systems might be required for each use. Thus, when designing a system, it may be advantageous to build chaos in, so as to achieve flexibility. For an experimental example, see Box 1. This experiment couples a magnetic field with a gravitationally buckling ribbon. Other recent relevant experiments have involved laser systems[10-14], electrical circuits[15], thermal convection[16] and arhythmically oscillating cardiac tissue (controlled using a small electrical stimulus)[18].

We will now consider examples of dynamical processes that we wish to control. A simple example might be a metre stick balanced on one's palm. This system is not chaotic and, provided that the stick does not stray too far from the vertical, it can be stabilized in its normally unstable, vertical state by making small motions of one's palm. This is an example of stabilizing an unstable steady state.

As a simple illustrative example of controlling a chaotic system we consider the 'Hénon map'[49]. Here the word 'map' refers to the fact that the time variable is discrete and integer-valued. The Hénon map is also described as a 'two-dimensional map'

BOX 1 Experimental confirmation

THE control of chaos by the application of tiny perturbations has been experimentally applied in several laboratories[14,17,29] (a modification, effective in stabilizing high-period orbits, appears in refs 10, 15). The first such application[9] used a nonlinear, inverted, magneto-elastic ribbon. This ribbon, sketched in Fig. B1, was clamped at its base but was otherwise free to move. The stiffness of the material was nonlinearly dependent on applied magnetic field, and in an oscillating field, applied by external field coils (not shown), the ribbon alternately buckled and stiffened in a complicated, chaotic pattern[76]. The position of the ribbon, $X(t)$, was measured at a point near its base using an optical sensor. By making small adjustments to the amplitude of the oscillating field (variations were less than 9% of its nominal value), the ribbon could be trapped in a variety of periodic motions. Figure B2 shows the ribbon displacement sampled once every oscillation of the applied magnetic field as a function of the number of periods, n. The sampling of the orbit at the drive period can be regarded as a "stroboscopic" surface of section. Initially no control was applied, leading to chaotic variation in the range of ribbon displacements between 2.7 and 4.4 on the vertical axis. Then, at about time $n \cong 2{,}200$, the control algorithm was activated to attempt to stabilize a period-1 orbit in the stroboscopic surface of section (compare this with Fig. 2*b*). After this orbit had been stabilized for over 2,000 oscillations, the control algorithm was switched at time $n \cong 4{,}900$ to try to stabilize a period-2 orbit (see Fig. 2*c*), and was later switched back to the period-1 orbit at $n \cong 7{,}100$.

The nonlinear ribbon was subsequently used to show that chaotic sensitivity can be used to actually direct trajectories in a chaotic system. By making small adjustments to the applied magnetic field (less than 5% of its nominal value), any trajectory could be quickly directed to a small target state. Figure B3 shows several successive trajectories which are rapidly brought from a variety of initial states, indicated by grey circles, to the target neighbourhood, $X = 2.5 \pm 0.01$. This target was chosen because it is ordinarily seldom visited: without control, the time to reach the chosen target was once in 500 iterations on average. By applying small perturbations, on the other hand, the target could typically be reached in less than once every 20 iterations.

We reiterate that the control was accomplished in real time and in the presence of experimental noise and modelling errors (to accommodate this, the targeting algorithm must be periodically reapplied; see refs 33, 35 for details). Additionally, the computational models used both for stabilization and for targeting were constructed entirely from available experimental data and without an *a priori* analytic model.

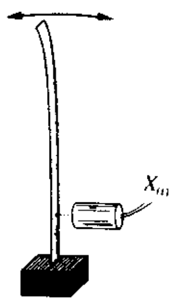


FIG. B1 Chaotic ribbon.

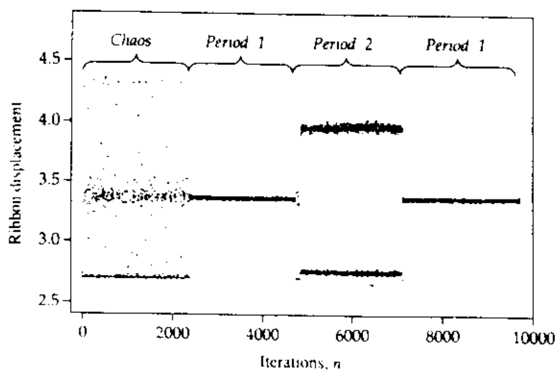


FIG. B2 Experimental stabilization of chaotic ribbon (reprinted with permission from ref. 9).

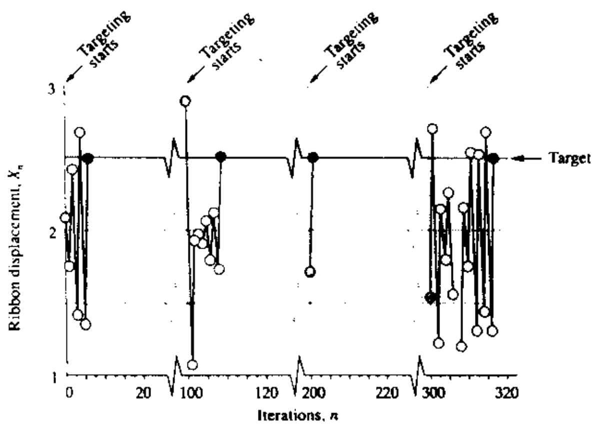


FIG. B3 Several successive realizations of targeting of a small neighbourhood of $X = 2.5$. Ribbon is allowed to wander chaotically between targeting attempts: targeting is initiated at $n = 0, 100, 200, 300$. With targeting, the neighbourhood of $X = 2.5$ is reached within 20 iterations; without targeting, the same neighbourhood is visited less than once every 500 iterations. The origin of the ordinate scale here differs from that of Fig. B2 because the optical sensor was moved between the times of the two experiments.

because the state of the system at time n (where $n=0, 1, 2, \ldots$) is given by two scalar variables, x_n and y_n. The map specifies a rule for evolving the state of the system at time n to the state at time $(n+1)$. For the Hénon map, this rule is

$$x_{n+1} = p + 0.3y_n - x_n^2 \quad (1a)$$

$$y_{n+1} = x_n \quad (1b)$$

where the parameter p is set to a nominal value of $p_0 = 1.4$. Thus, given an initial state (x_1, y_1), the map allows us to calculate (x_2, y_2), which, when again inserted into the map, yields (x_3, y_3), and so on. Iterations of this map eventually converge to a strange attractor, as shown in Fig. 1. This structure is called an attractor because distant points are drawn toward it under successive iterations of the map, and it is called strange because it is infinitely intricate on every distance scale we might choose to examine. In the terminology of Mandelbrot, it is a fractal and has a fractal dimension of ~1.3. In the figure, we show just a few of the infinite number of embedded unstable periodic orbits; a period-1 point, A_*, which is revisited every map iteration, period-2 points, B_1 and B_2, each of which are revisited every other map iteration ($B_1 \to B_2 \to B_1 \to B_2 \ldots$), and period-4 points, $C_1 \ldots C_4$, which are cycled through every four map iterations.

Although the above example is a map (the time, n, is discrete), many problems in science and engineering involve continuous time, such as a system of M first-order, autonomous ordinary differential equations, $d\boldsymbol{\xi}/dt = \mathbf{G}(\boldsymbol{\xi})$, where $\boldsymbol{\xi}(t) = (\xi^{(1)}(t), \xi^{(2)}(t), \xi^{(3)}(t), \ldots, \xi^{(M)}(t))$ is an M-vector, and the continuous variable t denotes time. In such a case, discrete time systems are still of interest, as the M-dimensional continuous time system can be reduced to an $M-1$ dimensional map by the Poincaré surface-of-section technique illustrated in Fig. 2a for $M=3$. Here we associate the continuous time trajectory with a discrete time trajectory, $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$, where $\mathbf{Z}_n$ denotes an $M-1$ coordinate vector specifying the position on the surface of section of the nth upward piercing of the surface. Given a $\mathbf{Z}_n$, we can integrate the equation $d\boldsymbol{\xi}/dt = \mathbf{G}(\boldsymbol{\xi})$ forward in time from that point, until the next upward piercing of the surface of section, at the surface coordinates $\mathbf{Z}_{n+1}$. Thus $\mathbf{Z}_{n+1}$ is uniquely determined by $\mathbf{Z}_n$, and there must exist a map, $\mathbf{Z}_{n+1} = \mathbf{F}(\mathbf{Z}_n)$, from one trajectory point on the surface of section to the next. Although we may not be able to write down $\mathbf{F}$ explicitly, the knowledge that it exists is still useful. Figure 2b shows a periodic orbit of the continuous time system which results in a period-1 orbit of the associated Poincaré surface of section map: $\mathbf{Z}_* = \mathbf{F}(\mathbf{Z}_*)$. Figure 2$c$ shows a periodic orbit of the continuous time system resulting in a period-2 orbit of the map: $\mathbf{Z}_2 = \mathbf{F}(\mathbf{Z}_1)$, $\mathbf{Z}_1 = \mathbf{F}(\mathbf{Z}_2)$.

Let us say that we have selected one of these unstable periodic orbits as providing the best performance of some hypothetical system. For simplicity, we consider the case where the desired orbit is a period-1 orbit of some N-dimensional map

$$\mathbf{Z}_{n+1} = \mathbf{F}(\mathbf{Z}_n, p) \quad (2)$$

where $\mathbf{Z}$ is an N-dimensional vector, and p denotes a system parameter. We first approximate the dynamics near the period-1 point, denoted $\mathbf{Z}_*$ (where $\mathbf{Z}_* = \mathbf{F}(\mathbf{Z}_*, p_0)$), for values of the parameter, p, close to the nominal value, p_0, by the linear map

$$(\mathbf{Z}_{n+1} - \mathbf{Z}_*) = \mathbf{A} \cdot (\mathbf{Z}_n - \mathbf{Z}_*) + \mathbf{B}\ (p - p_0) \quad (3)$$

Here $\mathbf{A}$ is an $N \times N$ dimensional jacobian matrix and $\mathbf{B}$ is an N-dimensional column vector, where $\mathbf{A} = \partial\mathbf{F}/\partial\mathbf{Z}$, $\mathbf{B} = \partial\mathbf{F}/\partial p$, and these partial derivatives are evaluated at $\mathbf{Z} = \mathbf{Z}_*$ and $p = p_0$. We now assume that we can adjust the parameter p on each iteration. That is, we determine $\mathbf{Z}_n$ and on that basis make an appropriate small change in p from the nominal value p_0. Thus we replace p by p_n. Taking the control law to be linear, we have

$$(p_n - p_0) = -\mathbf{K}^T \cdot (\mathbf{Z}_n - \mathbf{Z}_*) \quad (4)$$

Here $\mathbf{K}$ is a constant N-dimensional column vector and $\mathbf{K}^T$ is its transpose. Choice of the vector $\mathbf{K}$ specifies the control law specifying p_n on each iteration. Substituting equation (4) into equation (3), we obtain

$$\delta\mathbf{Z}_{n+1} = (\mathbf{A} - \mathbf{B} \cdot \mathbf{K}^T)\delta\mathbf{Z}_n \quad (5)$$

where $\delta\mathbf{Z}_n = \mathbf{Z}_n - \mathbf{Z}_*$. Thus the period-1 point, $\mathbf{Z}_*$, will be stable if one can choose $\mathbf{K}$ so that the matrix $(\mathbf{A} - \mathbf{B} \cdot \mathbf{K}^T)$ only has eigenvalues with modulus smaller than unity. In this case, $\delta\mathbf{Z}_n \to 0$ (that is, $\mathbf{Z}_n \to \mathbf{Z}_*$) as $n \to \infty$. The choice of $\mathbf{K}$ represents a standard problem in control theory (for example, see ref. 50). (In fact, the matrix $(\mathbf{A} - \mathbf{B} \cdot \mathbf{K}^T)$ can be made to have any desired set of N eigenvalues provided that $\mathbf{B}$ satisfies a certain condition (called controllability[50]).) This procedure gives the time-dependent parameter values, p_n, required to stabilize an unstable period-1 point in a chaotic system. For the case of higher-period orbits, see ref. 4. We imagine that, because of practical constraints, we cannot make the deviations of p_n from p_0 too large. Thus we imagine that $|p_n - p_0|$ is bounded by some allowed maximum value, $\delta p_{\max}$. Hence by equation (4), $\delta p_{\max} > |\mathbf{K}^T \cdot (\mathbf{Z}_n - \mathbf{Z}_*)|$. If the system state is outside this region, we apply no perturbation and wait until the state falls within the

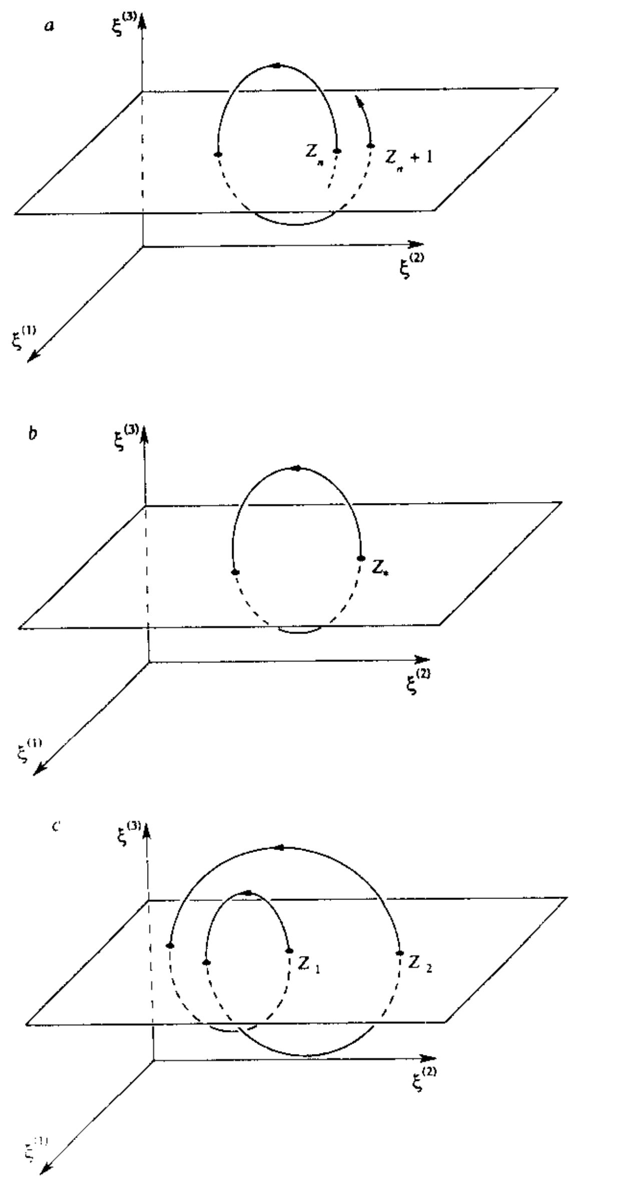


FIG. 2 a, The Poincaré surface-of-section technique. The surface of section, here shown for the case ξ_3 = constant, can, in principle, be chosen in any convenient way. b, Period-1 orbit; c, period-2 orbit.

31

given region. Once this occurs, we trap the system near the desired periodic state by applying the appropriate small nudges given by equation (4). Stabilization of the periodic points of the Hénon map by small perturbations is similar to stabilizing a vertical metre stick insofar as neither can be accomplished if the current state is far from the desired unstable state. We will discuss later how to direct a chaotic orbit to a neighbourhood of a desired state.

To illustrate stabilization of an unstable period-1 orbit in a more intuitive and geometrical manner, we consider the special case of a two-dimensional map where the desired period-1 orbit has one unstable and one stable direction. That is, there are two curves through $\mathbf{Z}_*$, called the stable manifold and the unstable manifold, as shown in Fig. 3*a*. We consider a small neighbourhood of $\mathbf{Z}_*$, so that the lines shown in Fig. 3 are roughly straight. An orbit starting from a point on the stable manifold remains on the stable manifold and moves exponentially toward $\mathbf{Z}_*$, whereas orbits of points on the unstable manifold move exponentially away from $\mathbf{Z}_*$. This corresponds to the case in which the matrix **A** has two real eigenvalues, one with magnitude less than one (stable), and one with magnitude greater than one (unstable). Also shown in Fig. 3*a* is a dashed line along which $\mathbf{Z}_*$ can be shifted by a change of the parameter p (this line is parallel to the vector **B**).

Imagine that the point $\mathbf{Z}_n$ falls close to $\mathbf{Z}_*(p_0)$, as shown in Fig. 3*a*. We now perturb the value of p from p_0 to $p_0+\delta p$, as shown in Fig. 3*b*. On the next iteration, the orbit is attracted towards $\mathbf{Z}_*(p_0+\delta p)$, in a direction parallel to its stable manifold, and, at the same time, is repelled from $\mathbf{Z}_*(p_0+\delta p)$ in a direction parallel to its unstable manifold. Thus if δp is chosen properly, we can cause $\mathbf{Z}_{n+1}$ to fall precisely on the stable manifold of $\mathbf{Z}_*(p_0)$. Thereafter, we can return the parameter to its nominal value, p_0, and the orbit will remain on the stable manifold of $\mathbf{Z}_*(p_0)$ and will approach $\mathbf{Z}_*(p_0)$. In terms of the more general discussion leading to equations (5), the above corresponds to a special choice of the control vector, **K**, such that one of the eigenvalues of $(\mathbf{A}-\mathbf{B}\cdot\mathbf{K}^T)$ is zero and the other is the original stable eigenvalue of **A**. As shown in ref. 4, this choice is optimal in that it minimizes the average time during which the orbit wanders chaotically before it can be stabilized.

As an example, in Fig. 4, we show the results of stabilizing the periodic orbit, $\mathbf{A}_*$, on the Hénon attractor by adjusting p by less than 1% of its nominal value. Starting at a random initial point on the attractor, we see that for the first 86 iterations, the trajectory moves chaotically on the attractor, never falling within the desired small region about $\mathbf{A}_*$. Then, on the 87th iteration, the state falls within the desired region, and thereafter is held near $\mathbf{A}_*$. In the presence of noise, the stabilization procedure would have to be regularly reapplied (see ref. 3 for details).

The control strategy need not rely on an *a priori* analytical model. Just as one can balance a metre stick on the palm without knowing anything at all about Newton's equations of motion for the stick, one can produce effective stabilization for more complicated systems without an explicit set of differential equations. By empirical study of the effects of small parameter changes on orbits near a desired periodic state, an arbitrary periodic state in a chaotic attractor can be stabilized using only small controls (see Box 1). This is important because in experimental situations one may often not have on hand an accurate analytical model.

Figure 5 illustrates how knowledge of a periodic orbit, the matrix **A**, and the vector **B** can all be extracted purely from observations of the trajectory on the strange attractor. Imagine that we collect a long data string of observed surface-of-section piercings, $\mathbf{Z}_1$, $\mathbf{Z}_2$ and so on. If two successive **Z**s are close to each other, say $\mathbf{Z}_{100}$ and $\mathbf{Z}_{101}$, then there will typically be a period-1 orbit $\mathbf{Z}_*$ nearby[47,48] (Fig. 5). Having observed a first such close return, we then search the succeeding data for other close-return pairs $(\mathbf{Z}_n, \mathbf{Z}_{n+1})$ restricted to the small region of the original close return (shown shaded in Fig. 5). Because orbits on a strange attractor are ergodic, we will have many such pairs if the data string is long enough. Assuming the shaded region to be small, we then try to fit these close returning pairs with a linear relation

$$\mathbf{Z}_{n+1}=\hat{\mathbf{A}}\cdot\mathbf{Z}_n+\hat{\mathbf{C}} \tag{6}$$

Generally, if there is noise in the data, we would want to use as many pairs as possible and fit the matrix $\hat{\mathbf{A}}$ and the vector $\hat{\mathbf{C}}$ using a least-squares fit to the data. Thus $\hat{\mathbf{A}}$, the least-squares fit matrix, is an approximation to the jacobian matrix **A** of equation (3), and the period-1 point, $\mathbf{Z}_*(p)$, is approximated by $(1-\hat{\mathbf{A}})^{-1}\cdot\hat{\mathbf{C}}$. To find **B** of equation (3), we change p slightly, $p\to p+\Delta p$, redetermine the period-1 point $\mathbf{Z}_*(p+\Delta p)$ as before, and approximate **B** as $[\mathbf{Z}_*(p+\Delta p)-\mathbf{Z}_*(p)]/\Delta p$. To find period-2 orbits, we proceed in the same way, but for pairs $(\mathbf{Z}_n, \mathbf{Z}_{n+2})$ that are close after two surface-of-section piercings (and similarly for higher periods).

In experimental studies of chaotic dynamics, it is often helpful to use 'delay coordinate embedding'[51,52]. These allow one to obtain information on the topological phase space structure of an attractor even when one cannot directly measure the instantaneous system state vector, $\boldsymbol{\xi}$, which we take to be M-dimensional. As a typical example, the time history of only a single scalar variable, say $\phi(t)$, might be the only thing that one

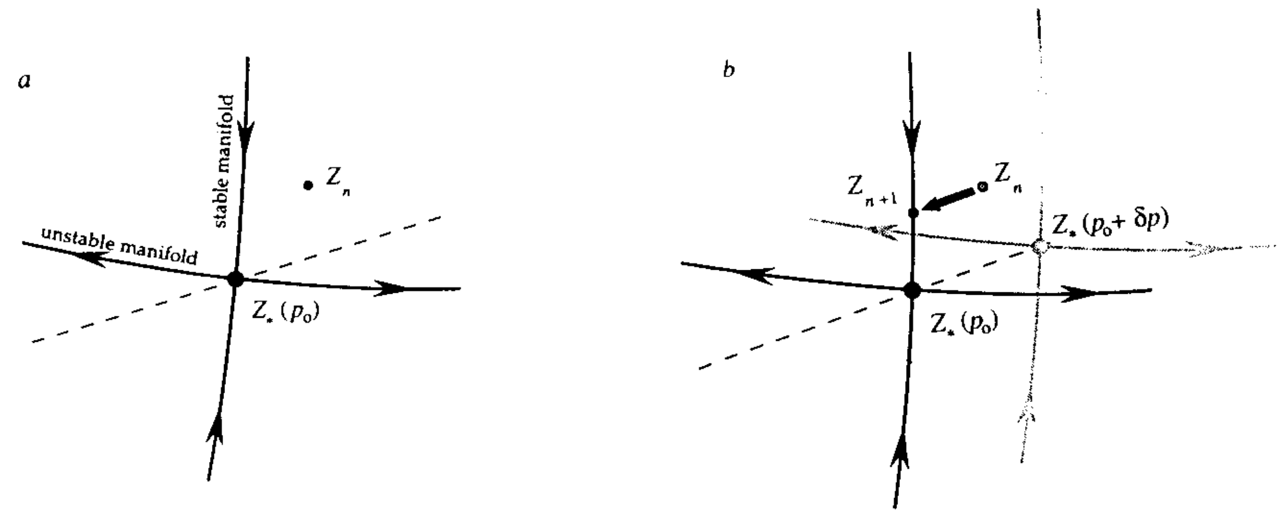


FIG. 3 *a*, The period-1 point, $\mathbf{Z}_*(p_0)$, its stable and unstable manifolds (solid lines), and the line (dashed) along which $\mathbf{Z}_*$ can be shifted by perturbation of the parameter p. *b*, Result of perturbing p to $p_0+\delta p$; the stable and unstable manifolds of $\mathbf{Z}_*(p_0+\delta p)$ are shown as grey lines through $\mathbf{Z}_*(p_0+\delta p)$.

32

can measure experimentally. To proceed, one forms the Q-dimensional 'delay coordinate vector'

$$\mathbf{\Phi}(t) = (\phi(t), \phi(t - T_D), \phi(t - 2T_D), \ldots, \phi(t - (Q-1)T_D)) \quad (7)$$

where T_D is some conveniently chosen delay time. Embedding theorems[53,54] guarantee that for $Q \geq 2M + 1$, the vector $\mathbf{\Phi}(t)$ is generically a global one-to-one representation of the system state. (Actually, for our purposes we do not require a global embedding; we only require a one-to-one correspondence in the small region near the periodic orbit, and this can typically be achieved with $Q = M$.) To obtain a discrete time series from $\mathbf{\Phi}(t)$, we can again use a Poincaré surface of section, but this time in $\mathbf{\Phi}$-space. Let $\tilde{\mathbf{Z}}_n$ denote points in the $\mathbf{\Phi}$-space surface of section, and let the corresponding surface-of-section map for a constant value of the parameter p be denoted $\tilde{\mathbf{Z}}_{n+1} = \tilde{F}(\tilde{\mathbf{Z}}_n, p)$. As pointed out in ref. 7, in the presence of parameter variation ($p \to p_n$), delay coordinates lead to a map of a form other than $\tilde{\mathbf{Z}}_{n+1} = \tilde{F}(\tilde{\mathbf{Z}}_n, p_n)$, which is the form assumed for equations (3) and (4). As an example, say that the time interval T_n between the surface-of-section piercing at $\tilde{\mathbf{Z}}_n$ and at $\tilde{\mathbf{Z}}_{n+1}$ is such that

$$rT_n \geq (Q+1)T_D > (r-1)T_n \quad (8)$$

where r is an integer. Then the relevant map is of the form

$$\tilde{\mathbf{Z}}_{n+1} = \tilde{F}(\tilde{\mathbf{Z}}_n, p_n, p_{n-1}, \ldots, p_{n-r}). \quad (9)$$

This follows because $\mathbf{\Phi}(t_n)$, where t_n denotes the nth piercing of the surface of section (corresponding to $\tilde{\mathbf{Z}}_n$), has components $\phi(t_n), \ldots, \phi(t_n - (Q-1)T_D)$ and hence $\tilde{\mathbf{Z}}_n$ must depend, not only on p_n, but also on all the other parameter values in effect during the time interval $t_n \leq t \leq [t_n - (Q-1)T_D]$. For discussion of how the analysis of equations (3) to (5) can be extended to the case of delay coordinates, see refs 4 and 7.

Although the approach just described provides a systematic means of choosing a feedback algorithm, it will often be the case (particularly when the dynamics are low-dimensional) that a trial and error procedure will succeed[10,15]: simply choose some feedback law (arbitrarily choose $\mathbf{K}$ in equation (4)) and vary it until it is observed that the desired orbit stabilizes. (Knowledge of the periodic orbit location in phase space is still required, so a procedure such as that of Fig. 5 may still be necessary.)

Further discussion of stabilization

We now briefly discuss some other considerations relevant to the stabilization of unstable periodic orbits and steady states embedded in strange attractors. One issue is related to bifurcations. The word bifurcation, when applied to a periodic orbit or steady state, refers to a change in the character of the orbit from stable to unstable while the system is continuously changed. Often the nset of chaos comes about as a result of bifurcations of periodic orbits or steady states. Two well known cases are the commonly observed cascade of period-doubling bifurcations preceding chaos, and the onset of chaos when a steady state bifurcates from stable to unstable. Thus one way of controlling chaos is to prevent or change the character of these bifurcations. This has been discussed in refs 18 and 19, where control methods that are insensitive to errors in the knowledge of the system are used. In this regard, we also mention the laser experiments of refs 12 and 13.

Control can also be used as a means of tracking the location of unstable orbits as the system is changed[21,22]. In a recent experiment, this technique was used to increase the stable steady power output of a laser by an order of magnitude[17]. Another issue is that of prescribing controls that are assured of bringing the trajectory to the desired periodic orbit or steady state. This has been addressed using the Lyapunov function method[16,23].

Other interesting work in this general area includes the use of the describing function technique of control theory for finding and controlling unstable periodic orbits[24]; control of unstable chaotic sets (as opposed to chaotic attractors)[25]; the control of homoclinic orbits[26]; and control of aperiodic orbits[55].

Directing chaotic trajectories

Stabilizing a system by small perturbations is extremely effective once the system at hand comes close to the desired state. But if it starts far from the desired state, it might take an unacceptably long time before a typical orbit comes close enough to the desired state to be captured. We now discuss how small perturbations, applied when the orbit is far from the desired state, can be used to steer the system to this state. For example, chaotic rhythms in cardiac tissue can be stabilized as outlined above[17]. What if one wants to stabilize a regular rhythm in the heart without having the luxury of simply waiting for the system to fall near a desired state? It is, moreover, intrinsically desirable to be able to steer a system to a general target in phase space (not necessarily a periodic orbit). NASA's mission specialists demonstrated this when, as mentioned earlier, they achieved the first scientific cometary encounter by steering the spacecraft ISEE-3/ICE in a complex trajectory using only small nudges from the spacecraft's dwindling fuel supply[39-43].

The application of chaotic sensitivity to steer trajectories to targets in chaotic systems using small controls is discussed in refs 33-37; related work appears in refs 38 and 55. To understand the idea in its simplest form, consider the well known logistic map (see for example ref. 56):

$$X_{n+1} = pX_n(1 - X_n) \quad (10)$$

where we take a nominal value of 3.9 for p. This map is used to describe the behaviour of a population of organisms after successive years, where n denotes the year. In this case, X represents the (normalized) population, and p defines its growth rate per year when the population is small. For any X between 0 and 1, it is easy, using only a pocket calculator, to adjust the growth rate slightly so that any given target, again between 0 and 1, is quickly reached.

Suppose for example that the current state is $X_1 = 0.4$ and we want to reach the vicinity of $X_n = 0.8$. If we can adjust p during the first year by a small amount, say between 3.8 and 4.0, then after one year, the population can range between $X_2 = 0.91$ and $X_2 = 0.96$. We then return to the nominal parameter value, $p = 3.9$, and after a second year the range grows to cover from $X_3 = 0.15$ to $X_3 = 0.31$. After a third year, the range grows even more to extend from $X_4 = 0.50$ to $X_4 = 0.84$. As our target, $X = 0.8$, is in this range, there must be some value of p_1 in the

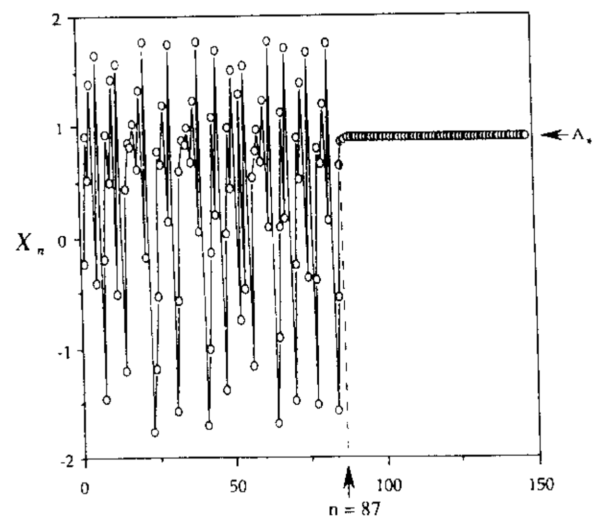


FIG. 4 Stabilization of the period-1 state for the Hénon map.

range $3.8 \le p_1 \le 4.0$ such that when p is shifted to that value in the first year, X falls on the target in only three years. Indeed, a little work with our pocket calculator reveals that we can reach our target by setting $p_1 = 3.83189\ldots$. We emphasize that it is possible to accomplish this only because the logistic system is chaotic, and because chaotic systems are characterized by the exponential growth of small disturbances. This exponential growth implies that we can reach any accessible target extremely quickly (that is, in a time of the order of the logarithm of the maximum allowed size of the small parameter perturbation), using only a small perturbation.

As another example, consider the equations

$$\frac{dX}{dt} = \sigma(Y - X) \tag{11a}$$

$$\frac{dY}{dt} = -XY - Y + r + p(t) \tag{11b}$$

$$\frac{dZ}{dt} = XY - bZ \tag{11c}$$

which (for $p(t) \equiv 0$) were introduced[57] by Lorenz as a simplified model of chaotic fluid thermal convection. The Lorenz equations provide a leading-order description of the dynamics of a fluid contained in a thin vertically oriented torus with a heat source applied at the bottom[58]. The equations with $p(t) \equiv 0$ have a strange attractor for the parameter values $\sigma = 10$, $r = 28$, $b = 8/3$, and an orbit on this chaotic attractor is shown in blue in Fig. 6. The steady state, $X(t) = Y(t) = Z(t) = 0$, representing no fluid convection, is a solution of the Lorenz equations. Furthermore, this solution is contained in the chaotic attractor. Note from Fig. 6 that the blue finite-duration orbit does not reach anywhere near this steady state (the open circle in the figure). If the orbit were followed long enough, it would eventually come arbitrarily near the stationary state. We estimate that of the order of one in 10^{10} orbits around either of the lobes of the attractor ever passes through a sphere of radius 0.1 centred at $X = Y = Z = 0$. (For comparison, the blue trajectory in Fig. 6 shows only about 20 orbits around either lobe of the attractor.) For the purpose of demonstrating control of the Lorenz system we have added the term $p(t)$ to the right-hand side of equation (11b). For the physical situation of a vertically oriented fluid-filled torus, the term $p(t)$ represents a perturbation of the position of the heat source slightly to the left or right in the plane of the torus ($p = 0$ corresponds to heating exactly at the bottom of the torus). Restricting the perturbations, $p(t)$, to be small, $|p(t)| < 0.01$, it typically takes of the order of 10 orbits around a lobe to reach the origin (as compared, again, with 10^{10} orbits to reach $\sqrt{X^2 + Y^2 + Z^2} < 0.1$ when no control is applied). Figure 6 shows such an orbit in red. The method for programming the perturbations is described in ref. 36 and is similar in principle to that which we have described for the logistic map, equation (10).

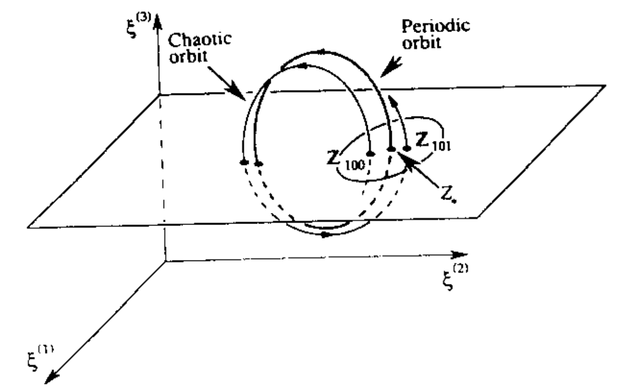


FIG. 5 Determining unstable periodic point, **Z** from data, $\mathbf{Z}_{100}$, $\mathbf{Z}_{101}$ . . .

The preceding example illustrates the idea of targeting. More detailed descriptions and methods can be found in refs 33–38. One technique is to patch together different trajectories to reach a desired target state. These trajectories can originate in a single chaotic system with differing parameter values[37,38], or from a systematic examination of small changes in the system state[39–43]. NASA's manoeuvre of the spacecraft ISEE-3/ICE is such a case. Five separate swings past the Moon were combined to produce the final spacecraft trajectory. By inventive choices of intermediate trajectories, researchers have shown this technique to be effective in matching greatly differing initial and final state vectors.

At this point, we must make two remarks. First, the sensitivity of chaotic systems allows us to produce large changes in the orbit of the system after some time using tiny perturbations, but the same sensitivity makes the final state depend on ubiquitous noise. The noise, if small enough, can be compensated for, however, by periodically reapplying the targeting algorithm, thereby obtaining mid-course corrections of the parameter perturbation (see Box 1 and refs 33 and 35). Second, one needs a global model of the system in order to direct trajectories. This differs from the stabilization of periodic orbits or steady states, where one only needs local information, near the desired periodic orbit. In general, this makes the use of experimental delay coordinate techniques for application to targeting more difficult. Nevertheless, in some cases it may be possible (Box 1).

The control of chaos by small perturbations has been developed into a proposed application for communication[59]. Consider the 'double scroll' electrical oscillator[60], which yields a chaotic signal consisting of a seemingly random sequence of positive and negative peaks. If we associate a positive peak with a one and a negative peak with a zero, the signal yields a binary sequence. With small control perturbations, we can cause the signal to follow an orbit whose binary sequence represents the information we wish to communicate. (Here our 'target' is a particular binary sequence rather than a point in phase space.) Hence the chaotic power stage that generates the waveform for transmission can remain simple and efficient (complex chaotic behaviour occurs in simple systems), while all the complex electronics controlling the output can remain at the low-power microelectronic level.

Other ways to alter chaotic dynamics

We have focused here on work using the sensitivity of chaotic systems to stabilize existing chosen periodic orbits and steady states and to steer trajectories with only small controls. Several authors have developed other techniques to alter the dynamics of chaotic systems without explicitly using this sensitivity, and in this section, we summarize some of their key contributions.

One body of research[61–68] seeks to control a nonlinear system to follow a prescribed goal dynamics. If we denote the system by

$$\frac{d\boldsymbol{\xi}}{dt} = \mathbf{F}(\boldsymbol{\xi}) + \mathbf{U}(t) \tag{12}$$

where $\mathbf{U}(t)$ is an additive controlling term, then the object is to choose $\mathbf{U}(t)$ so that $|\boldsymbol{\xi}(t) - \mathbf{g}(t)| \to 0$ as $t \to \infty$, where $\mathbf{g}(t)$ is the goal dynamics. To accomplish this, the simple choice

$$\mathbf{U}(t) = \frac{d\mathbf{g}}{dt} - \mathbf{F}(\mathbf{g}(t)) \tag{13}$$

is made. Thus $\boldsymbol{\xi}(t) = \mathbf{g}(t)$ is clearly a solution of the controlled equations. What is not so clear is that convergence to this goal will often occur ($|\boldsymbol{\xi}(t) - \mathbf{g}(t)| \to 0$ as $t \to \infty$). Whether it does so depends on the particular $\mathbf{F}$ and the initial condition, $\boldsymbol{\xi}(0)$. The regions of $\boldsymbol{\xi}$-space such that controlled orbits originating in them converge to the goal, $\mathbf{g}$, are called entrainment regions[64–66]. This method potentially works for nonlinear systems in general (not necessarily chaotic) and has the advantage of not requiring

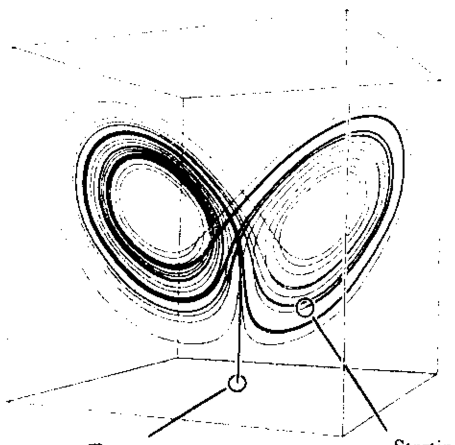


FIG. 6 Targeting of steady state in Lorenz attractor. Blue; trajectory without control; red: trajectory with control.

feedback. On the other hand, the applied controls are not typically small and convergence to the goal is not assured.

Another body of research addresses the effects of periodic[69-73] and stochastic[74,75] perturbations on chaotic systems. As one might expect, the effects of such perturbations can be quite difficult to predict in general, and indeed these studies are not 'goal oriented', in that a desired behaviour is not specified in advance and a generic technique for achieving such a goal has not been developed. Nevertheless, dramatic changes in the dynamics of chaotic systems have been recorded using these methods; for example, periodic or nearly periodic behaviour can sometimes be produced from originally chaotic dynamical systems.

The advantages of chaos

The presence of chaos may be a great advantage for control in a variety of situations. In a nonchaotic system, small controls typically can only change the system dynamics slightly. Short of applying large controls or greatly modifying the system, we are stuck with whatever system performance already exists. In a chaotic system, on the other hand, we are free to choose between a rich variety of dynamical behaviours. Thus we anticipate that it may be advantageous to design chaos into systems, allowing such variety without requiring large controls or the design of separate systems for each desired behaviour.

The general problem of controlling chaotic systems is very rich, and may help solve technologically important problems in widely diverse fields of study. In communications, it has been proposed that chaotic fluctuations can be put to use to send controlled, pre-planned signals[59]. In physiology, applications have been proposed for controlling chaos in the heart[17] and in neural information processing[27]. In fluid mechanics, it has been demonstrated in a simple configuration that chaotic convection[16] can be controlled. Chemical researchers have developed mechanisms for controlling chaotic autocatalytic reactions[28]. Chaotic lasers[10,11] have been controlled, as has the chaotic diode circuit[15]. The wealth of results such as these encourage us to look forward to a fruitful future for the control of chaotic systems. □

Troy Shinbrot is at the Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA; present address for Troy Shinbrot is Laboratory for Fluid Mechanics, Chaos and Mixing, Northwestern University, Evanston, Illinois 60208, USA. Celso Grebogi and James A. Yorke are at the Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. Edward Ott is at the Department of Electrical Engineering, the Department of Physics and the Institute for Systems Research, University of Maryland, College Park, Maryland 20742, USA.

1. Dyson, F. *Infinite in All Directions*, 183-184 (Harper and Row, New York, 1988).
2. Ott, E., Grebogi, C. & Yorke J. A. *Phys. Rev. Lett.* **64,** 1196-1199 (1990).
3. Ott, E., Grebogi, C. & Yorke, J. A. in *Chaos: Soviet-American Perspectives on Nonlinear Science* (ed. Campbell, D. K.) 153-172 (Am. Inst. Phys., New York, 1990).
4. Ott, E., Romeiras, F. J., Grebogi, C. & Dayawansa, W. P. *Physica* D **58,** 165-192 (1992).
5. Auerbach, D., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev. Lett.* **69,** 3479-3482 (1992).
6. Fowler, T. B. *IEEE Trans. Auto. Control* **34,** 201-205 (1989)
7. Dressler, U. & Nitsche, G. *Phys. Rev. Lett.* **68,** 1-4 (1992).
8. Pyragas, K. *Phys. Lett.* A**170,** 421-428 (1992).
9. Ditto, W. L., Rauseo, S. N. & Spano, M. L. *Phys. Rev. Lett.* **65,** 3211-3214 (1990).
10. Roy, R., Murphy, T. W., Maier, T. D., Gills, A. & Hunt, E. R. *Phys. Rev. Lett.* **68,** 1259-1262 (1992).
11. Gills, Z., Iwata, C., Roy, R., Schwartz, I. B. & Triandaf, I. *Phys. Rev. Lett.* **69,** 3169-3172 (1992).
12. Bielawski, S., Bouazaoui, M., Derozier, D. & Glorieux, P. *Proc. Nonlinear Dynam. Optical Syst. Topical Mtg. Alpbach, Austria, June 22-26* (Optical Society of America, Washington, DC 1992).
13. Bielawski, S., Bouazaoui, M., Derozier, D. & Glorieux, P. *Stabilization and Characterization of Unstable Steady States in a Laser* (preprint, 1992).
14. Reyl, C., Flepp, L., Badii, R. & Brun, E. *Control of NMR-laser Chaos in High-dimensional Embedding Space* (preprint, 1992).
15. Hunt, E. R. *Phys. Rev. Lett.* **67,** 1953-1955 (1991).
16. Singer, J., Wang, Y. Z. & Bau, H. H. *Phys. Rev. Lett.* **66,** 1123-1125 (1991).
17. Garfinkel, A., Spano, M. L., Ditto, W. L. & Weiss, J. N. *Science* **257,** 1230-1235 (1992).
18. Abed, E. H., Wang, H. O. & Lee, H.-C. *Proc. 1992 Am. Control Conf.*, 2236-2237 (Chicago, 1992)
19. Wang, H. O. & Abed, E. H. *Proc. 2nd IFAC Workshop Syst. Struct. Control*, 494-497 (Prague, 1992).
20. Wang, H. O. & Abed, E. H. *Proc. 2nd IFAC Nonlinear Control Sys. Design Symp.*, 57-62 (Bordeaux, France, June 1992).
21. Schwartz, I. B. & Triandaf, I. *Phys. Rev.* A**46,** 7439-7444 (1992).
22. Carroll, T., Triandaf, I., Schwartz, I. B. & Pecora, L. *Phys. Rev.* A**46,** 6189-6192 (1992).
23. Chen, G. & Dong, X. *J. Circ. Syst. Comput.* **3** (in the press).
24. Genesio, R. & Tesi, A. *Automatica* **28,** 531-548 (1992)
25. Tél, T. *J. Phys.* A**24,** L1359-L1368 (1992).
26. Bloch, A. M. & Marsden, J. E. *Theor. comput. Fluid Dynam.* **1,** 179-190 (1989).
27. Ding, M. & Kelso, J. A. S. in *Measuring Chaos in the Human Brain* (eds Duke, D. & Pritchard, W.) 17-31 (World Scientific, Singapore, 1991).
28. Petrov, V., Gáspár, V., Masere, J. & Showalter, K. *Nature* **361,** 240-243 (1993).
29. Hübinger, B., Doerner, R. & Martienssen, W. *Controlling Chaotic Motion in Noisy Systems* (preprint, Phys. Inst. der J. W. Goethe-Universität, Frankfurt-am-Main, 1992).
30. Vassiliadis, D. *Physica* D (in the press).
31. Huberman, B. A. & Lumer, E. *IEEE Trans. Circ. Syst.* **37,** 547-550 (1990).
32. Sinha, S. & Ramaswamy, R. *Physica* D**43,** 118-128 (1990).
33. Shinbrot, T., Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **65,** 3215-3218 (1990).
34. Shinbrot, T. Ott, E., Grebogi, C. & Torke, J. A. *Phys. Rev.* A**45,** 4165-4168 (1992).
35. Shinbrot, T., Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **68,** 2863-2866 (1992).
36. Shinbrot, T., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Lett.* A**169,** 349-354 (1992).
37. Kostelich, E., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev.* E**47,** 305-310 (1993).
38. Bradley, E. in *Lecture Notes Control and Information Sciences*, No. 165 (eds Jacob, G. & Lamnabhi-Lagarrigue, F.) 307-325 (Springer, Berlin, 1991)
39. Farquhar, R., Muhonen, D. & Church, L. C. *J. astronaut. Sci.* **33,** 235-254 (1985).
40. Muhonen, D., Davis, S. & Dunham, D. *J. astronaut. Sci.* **33,** 255-273 (1985).
41. Dunham, D. W. & Davis, S. A. *J. astronaut. Sci.* **33,** 275-288 (1985).
42. Muhonen, D. & Folta, D. *J. astronaut. Sci.* **33,** 289-300 (1985).
43. Efron, L., Yeomans, D. K. & Schanzle, A. F. *J. astronaut. Sci.* **33,** 301-323 (1985).
44. So, P., Ott, E. & Dayawansa, W. P. *Phys. Lett.* A. (in the press).
45. Chen, G. & Dong, X. *From Chaos to Order—Perspectives and Methodologies in Controlling Chaotic Dynamical Systems Tech. Rep. 92-07* (Univ. Houston, Texas, 1992).
46. Vincent, T. L. & Yu, J. *Dynam. Control* **1,** 35-52 (1991).
47. Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev.* A**37,** 1711 (1988).
48. Auerbach, D., Cvitanović, P., Eckmann, J.-P., Gunaratne, G. & Procaccia, I. *Phys. Rev. Lett.* **58,** 2387 (1987).
49. Hénon, M. *Commun. Math. Phys.* **50,** 69-77 (1976).
50. Ogata, K. *Modern Control Engineering*, 2nd Ed, 347-889 (Prentice-Hall, Englewood Cliffs, 1990).
51. Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. *Phys. Rev. Lett.* **45,** 712-716 (1980).
52. Eckmann, J.-P. & Ruelle, D. *Rev. mod. Phys.* **57,** 617-656 (1985).
53. Whitney, H. *Ann. Math.* **37,** 645-671 (1936).
54. Takens, F. in *Lecture Notes in Mathematics 898* (eds Rand, D. A. & Young, L. S.) 366-381 (Springer, Berlin, 1981).
55. Mehta, N. J. & Henderson, R. M. *Phys. Rev.* A**44,** 4861-4865 (1991).
56. Schuster, H. G. *Deterministic Chaos*, 2nd ed., 37-38 (VCH, Weinheim, 1989).
57. Lorenz, E. *J. atmos. Sci.* **20,** 130-141 (1963).
58. Yorke, J. A., Yorke, E. D. & Mallet-Paret, J. *Physica* D**24,** 279-291 (1987).
59. Hayes, S., Grebogi, C. & Ott, E. *Phys. Rev. Lett.* **70,** 3031-3034 (1993).
60. Chua, L. O., Komura, M. & Matsumoto, T. *IEEE Trans. Circ. Syst.* **11,** 1073-1118 (1986).
61. Hübler, A. & Lüscher, E. *Naturwissenschaften* **76,** 67-69 (1989).
62. Hübler, A. W. *Helv. phys. Acta* **62,** 343-346 (1989).
63. Plapp, B. B. & Hübler, A. W. *Phys. Rev. Lett.* **65,** 2302-2305 (1990).
64. Jackson, E. A. & Hübler, A. W. *Physica* D**44,** 407-420 (1990).
65. Jackson, E. A. *Phys. Lett.* A**151,** 478-484 (1990).
66. Jackson, E. A. *Phys. Rev.* A**44,** 4839-4853 (1991).
67. Lüscher, E. & Hübler, A. W. *Helv. phys. Acta* **62,** 544-551 (1989).
68. Georgii, R., Eberl, W. & Lüscher, E. *Helv. phys. Acta* **62,** 290-293 (1989).
69. Alekseev, V. V. & Loskutov, A. Y. *Sov. Phys. Dokl.* **32,** 1346-1348 (1987).
70. Pettini, M. in *Dynamics and Stochastic Processes* (ed. Lima, R., Streit, L. & Vilela Mendes, R. 242-250 (Springer, Berlin, 1990).
71. Lima, R. & Pettini, M. *Phys. Rev.* A**41,** 726-733 (1990).
72. Braiman, Y. & Goldhirsch, I. *Phys. Rev. Lett.* **66,** 2545-2548 (1991).
73. Azevedo, A. & Rezende, S. M. *Phys. Rev. Lett.* **66,** 1342-1345 (1991).
74. Herzel, H. *A. Angew. Math. Mech.* **68,** 11-12 (1988).
75. Fahy, S. & Hamann, D. R. *Phys. Rev. Lett.* **69,** 761-764 (1992).
76. Ditto, W. L. *et al. Phys. Rev. Lett.* **63,** 923-966 (1989).
77. Lai, Y.-C., Ding, M.-Z. & Grebogi, C. *Phys. Rev.* E**47,** 86-92 (1992).
78. Abed, E. H., Wang, H. O. & Chen, R. *Physica* D (in the press).

ACKNOWLEDGEMENTS This work was supported by the United States Department of Energy (Office of Scientific Computing, Office of Energy Research). We thank D. Auerbach, G. R. Chen, A. Hübler and K. Scarbrough for helpful comments. W. Ditto, S. Rauseo and M. Spano are gratefully acknowledged for providing data for this article.

35

Physica D 58 (1992) 165–192
North-Holland

PHYSICA D

Controlling chaotic dynamical systems

Filipe J. Romeiras[a,b], Celso Grebogi[a,c,d], Edward Ott[a,e,f] and W.P. Dayawansa[f,g]

[a]Laboratory for Plasma Research, University of Maryland, College Park, MD 20742, USA
[b]Centro de Electrodinâmica (INIC) and Departamento de Matemática, Instituto Superior Técnico, 1096 Lisbon Codex, Portugal
[c]Department of Mathematics, University of Maryland, College Park, MD 20742, USA
[d]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA
[e]Department of Physics, University of Maryland, College Park, MD 20742, USA
[f]Department of Electrical Engineering, University of Maryland, College Park, MD 20742, USA
[g]Systems Research Center, University of Maryland, College Park, MD 20742, USA

Received 18 November 1991
Revised manuscript received 16 January 1992
Accepted 16 January 1992

We describe a method that converts the motion on a chaotic attractor to a desired attracting time periodic motion by making only small time dependent perturbations of a control parameter. The time periodic motion results from the stabilization of one of the infinite number of previously unstable periodic orbits embedded in the attractor. The present paper extends that of Ott, Grebogi and Yorke [Phys. Rev. Lett. 64 (1990) 1196], allowing for a more general choice of the feedback matrix and implementation to higher-dimensional systems. The method is illustrated by an application to the control of a periodically impulsively kicked dissipative mechanical system with two degrees of freedom resulting in a four-dimensional map (the "double rotor map"). A key issue addressed is that of the dependence of the average time to achieve control on the size of the perturbations and on the choice of the feedback matrix.

1. Introduction

It is common for systems to evolve with time in a chaotic way. In practice, however, it is often desired that chaos be avoided and/or that the system be optimized with respect to some performance criterion. Given a system which behaves chaotically, one approach might be to make some large (and possibly costly) alteration in the system which completely changes its dynamics in such a way as to achieve the desired objectives. Here we assume that this avenue is not available. Thus we address the following question: Given a chaotic system, how can we obtain improved performance and achieve a desired attracting time-periodic motion by making only *small* controlling temporal perturbations in an accessible system parameter.

The key observation is that a chaotic attractor typically has embedded densely within it an infinite number of unstable periodic orbits [1–5]. In addition, chaotic attractors can also sometimes contain unstable steady states (e.g., the Lorenz attractor has such an embedded steady state). Since we wish to make only small controlling perturbations to the system, we do not envision creating new orbits with very different properties from the already existing orbits. Thus we seek to exploit the already existing unstable periodic orbits and unstable steady states. Our approach is as follows: We first determine some of the unstable low-period periodic orbits and unstable steady states that are embedded in the chaotic attractor. We then examine these orbits and choose one which yields improved system performance. Finally, we apply small controls so as

36

to stabilize this already existing orbit.

Some comments concerning this method are the following:

(1) Before settling into the desired controlled orbit the trajectory experiences a chaotic transient whose expected duration diverges as the maximum allowed size of the control approaches zero.

(2) Small noise can result in occasional bursts in which the orbit wanders far from the controlled orbit.

(3) Controlled chaotic systems offer an advantage in flexibility in that any one of a number of different orbits can be stabilized by the small control, and the choice can be switched from one to another depending on the current desired system performance.

Although we describe the details only in the case of discrete time systems, this method is applicable in the continuous time case as well by considering the discrete time system obtained from the induced dynamics on a Poincaré section.

In order to illustrate the method we apply it to a periodically forced mechanical system (the kicked double rotor), which results in a four-dimensional map. Amongst the examples considered, we study cases where the unstable orbit of the uncontrolled system has two unstable eigenvalues and two stable eigenvalues, and the stabilization is achieved by variation of one control parameter characterizing the strength of the periodic forcing. The present paper generalizes our previous work [6] to the case of higher-dimensional systems [7] and also includes new material illustrating the effect of the choice of stabilization on the length of the chaotic transient experienced by the orbit before control is achieved. Other relevant references on the feedback stabilization of periodic or steady orbits embedded in chaotic attractors are the experiments of Ditto et al. [8], Singer et al. [9], and the paper of Fowler [10]. (Other works in the general field are listed in ref. [11].)

The plan of the paper is as follows. In section 2, we give an implementation of the method, initially developed in ref. [6], by using the "pole placement technique" [7, 12]. In particular, we address the problem of stabilization of periodic orbits with more than one unstable eigenvalue. We also discuss experimental implementation in the absence of an a priori mathematical system model and generalization of the method to deal with cases where delay coordinates embedding is used. In section 3 we present some results for the control of the Hénon map [13], a two-dimensional system that is used as a paradigm in the study of dynamical systems; these results extend those given in ref. [6] in directions relevant to our present study. In section 4 we present results for the control of the double rotor map [14], a four-dimensional system that describes a particular impulsively periodically forced mechanical system. Finally, in section 5 we present the main conclusions of the work.

2. Description of the method

2.1. Formulation

For the sake of simplicity we consider a discrete time dynamical system,

$$\mathbf{Z}_{i+1} = \boldsymbol{F}(\mathbf{Z}_i, p)\,, \tag{2.1}$$

where $\mathbf{Z}_i \in \mathbb{R}^n$, $p \in \mathbb{R}$ and $\boldsymbol{F}$ is sufficiently smooth in both variables. Here, p is considered a real parameter which is available for external adjustment but is restricted to lie in some small interval,

$$|p - \bar{p}| < \delta\,, \tag{2.2}$$

around a nominal value $\bar{p}$. We assume that the nominal system (i.e., for $p = \bar{p}$) contains a chaotic attractor. Our objective is to vary the parameter p with time i in such a way that for almost all initial conditions in the basin of the chaotic attractor, the dynamics of the system converge onto a desired time periodic orbit contained in the attractor. The control strategy is the following. We will find a stabilizing local feedback control law which is defined on a neighborhood of the desired periodic orbit. This is done by considering the first order approximation of the system at the chosen unstable periodic orbit. Here we assume that this approximation is stabilizable. Since stabilizability is a generic property of linear systems, this assumption is quite reasonable. The ergodic nature of the chaotic dynamics ensures that the state trajectory eventually enters into the neighborhood. Once inside, we apply the stabilizing feedback control law in order to steer the trajectory towards the desired orbit.

For simplicity we shall describe the method as applied to the stabilization of fixed points (i.e., period one orbits) of the map $\boldsymbol{F}$. The consideration of periodic orbits of period larger than one is straightforward and is discussed in section 2.5. Let $\boldsymbol{Z}_*(p)$ denote an unstable fixed point on the attractor. For values of p close to $\bar{p}$ and in the neighborhood of the fixed point $\boldsymbol{Z}_*(\bar{p})$ the map (2.1) can be approximated by the linear map

$$\boldsymbol{Z}_{i+1} - \boldsymbol{Z}_*(\bar{p}) = \mathbf{A}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})] + \boldsymbol{B}(p - \bar{p}), \tag{2.3}$$

where $\mathbf{A}$ is an $n \times n$ Jacobian matrix and $\boldsymbol{B}$ is an n-dimensional column vector,

$$\mathbf{A} = \mathbf{D}_Z \boldsymbol{F}(\boldsymbol{Z}, p), \tag{2.4}$$

$$\boldsymbol{B} = \mathbf{D}_p \boldsymbol{F}(\boldsymbol{Z}, p), \tag{2.5}$$

and these partial derivatives are evaluated at $\boldsymbol{Z} = \boldsymbol{Z}_*(\bar{p})$ and $p = \bar{p}$. We now introduce the time-dependence of the parameter p by assuming that it is a linear function of the variable $\boldsymbol{Z}_i$ of the form

$$p - \bar{p} = -\boldsymbol{K}^{\mathrm{T}}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})]. \tag{2.6}$$

The $1 \times n$ matrix $\boldsymbol{K}^{\mathrm{T}}$ is to be determined so that the fixed point $\boldsymbol{Z}_*(\bar{p})$ becomes stable. Substituting (2.6) into (2.3) we obtain

$$\boldsymbol{Z}_{i+1} - \boldsymbol{Z}_*(\bar{p}) = (\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}})[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})], \tag{2.7}$$

which shows that the fixed point will be stable provided the matrix $\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}}$ is asymptotically stable; that is, all its eigenvalues have modulus smaller than unity.

The solution to the problem of the determination of $\boldsymbol{K}^{\mathrm{T}}$, such that the eigenvalues of the matrix $\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}}$ have specified values, is well known from control systems theory and is called "pole placement technique" (see, for example, Ogata [12]). We summarize the relevant results.

2.2. Review of the pole placement technique

The eigenvalues of the matrix $\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}}$ are called the "regulator poles", and the problem of placing these poles at the desired locations by choosing $\boldsymbol{K}^{\mathrm{T}}$ with $\mathbf{A}$ and $\boldsymbol{B}$ given is the "pole placement problem".

Pole placement problem. Determine the matrix $\boldsymbol{K}^{\mathrm{T}}$ in such a way that the eigenvalues of the matrix $\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}}$ have specified (complex) values $\{\mu_1, \ldots, \mu_n\}$.

The following results [12] give a necessary and sufficient condition for a unique solution of the pole placement problem to exist, and also a method for obtaining it (Ackermann's method).

(1) The pole placement problem has a unique solution if and only if the $n \times n$ matrix

$$\mathbf{C} = (\boldsymbol{B} \,\vdots\, \mathbf{A}\boldsymbol{B} \,\vdots\, \mathbf{A}^2\boldsymbol{B} \,\vdots\, \ldots \,\vdots\, \mathbf{A}^{n-1}\boldsymbol{B}),$$

is of rank n. ($\mathbf{C}$ is called the controllability matrix).

(2) The solution of the pole placement problem is given by

$$\boldsymbol{K}^{\mathrm{T}} = (\alpha_n - a_n \ldots \alpha_1 - a_1)\mathbf{T}^{-1},$$

where $\mathbf{T} = \mathbf{CW}$, and

$$\mathbf{W} = \begin{pmatrix} a_{n-1} & a_{n-2} & \cdots & a_1 & 1 \\ a_{n-2} & a_{n-3} & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a_1 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Here $\{a_1, \ldots, a_n\}$ are the coefficients of the characteristic polynomial of $\mathbf{A}$,

$$|s\mathbf{I} - \mathbf{A}| = s^n + a_1 s^{n-1} + \cdots + a_n ,$$

and $\{\alpha_1, \ldots, \alpha_n\}$ are the coefficients of the desired characteristic polynomial of $\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}}$,

$$\Pi_{j=1}^{n}(s - \mu_j) = s^n + \alpha_1 s^{n-1} + \cdots + \alpha_n .$$

2.3. Control parameter

Our considerations so far are based on the linear eq. (2.7) and therefore only apply in the local region near $\boldsymbol{Z}_*(\bar{p})$. On the other hand, the limitation in the size of the parameter perturbations given by (2.2), when combined with (2.6), yields

$$|\boldsymbol{K}^{\mathrm{T}}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})]| < \delta . \tag{2.8}$$

This defines a slab of width $2\delta/|\boldsymbol{K}^{\mathrm{T}}|$. We choose to activate the control according to (2.6) only for values of $\boldsymbol{Z}_i$ inside this slab, and we choose to leave the control parameter at its nominal value (i.e., $p = \bar{p}$) when $\boldsymbol{Z}_i$ is outside this slab. Other choices are possible.

In summary, the control is determined by

$$p - \bar{p} = -\boldsymbol{K}^{\mathrm{T}}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})] \times u(\delta - |\boldsymbol{K}^{\mathrm{T}}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})]|) , \tag{2.9}$$

for arbitrary $\boldsymbol{Z}_i$ [not necessarily close to $\boldsymbol{Z}_*(\bar{p})$], where u is the unit step function defined by

$$u(\alpha) = \begin{cases} 0, & \alpha < 0 , \\ 1, & \alpha > 0 . \end{cases}$$

At this stage it should be pointed out that the matrix $\boldsymbol{K}^{\mathrm{T}}$ can be chosen in many different ways. In principle, any choice of regulator poles inside the unit circle serves our purpose. In ref. [6], the authors made a very special, though quite natural, choice of the gain matrix $\boldsymbol{K}^{\mathrm{T}}$: the resulting value of $p - \bar{p}$ forces the orbit onto the (linear) stable manifold of the fixed point at each iteration. In terms of regulator poles this choice corresponds to setting n_s of these poles equal to the n_s stable eigenvalues of matrix $\mathbf{A}$ and the remaining $n - n_s$ to 0. In terms of the slab (2.8) this choice corresponds not only to orientating it parallel to the stable manifold but also taking an appropriate width.

The choice of the matrix $\boldsymbol{K}^{\mathrm{T}}$ will be discussed at some length in our applications of the method in sections 3 and 4.

2.4. Time to achieve control

The control is activated (i.e., $p \neq \bar{p}$) only if $\boldsymbol{Z}_i$ falls in the narrow slab (2.8). Thus, for small δ, a typical initial condition will execute a chaotic orbit, unchanged from the uncontrolled case, until $\boldsymbol{Z}_i$ falls in this slab. Even then, because of nonlinearity not included in the linearized eq. (2.7), the control may not be able to bring the orbit to the fixed point. In this case the orbit will leave the slab and continue to wander chaotically as if there was no control. Since the orbit on the uncontrolled chaotic attractor is ergodic, at some time it will eventually satisfy (2.8) and also be sufficiently close to the desired fixed point so that control is achieved.

Thus, we create a stable orbit, which, for a typical initial condition, is preceded by a chaotic transient [15–18] in which the orbit is similar to orbits on the uncontrolled chaotic attractor. The length τ of such chaotic transient depends sensitively on the initial condition of the particular orbit. For initial conditions randomly chosen in the basin of attraction the distribution of chaotic transient lengths is exponential [15, 16],

$$\phi(\tau) \simeq \frac{1}{\langle\tau\rangle} \exp\left(-\frac{\tau}{\langle\tau\rangle}\right) , \tag{2.10}$$

for large τ. The quantity $\langle \tau \rangle$ is the characteristic length of the chaotic transient, called in the present case the *average time to achieve control*. Estimates of the scaling of $\langle \tau \rangle$ with δ for small δ are given in Appendix A for the case of two-dimensional maps.

2.5. Control of periodic orbits of period greater than one

The analysis of periodic orbits given in sections 2.1–2.3 can be extended to nontrivial periodic orbits (i.e., orbits with period greater than one). The most direct way is to take the Tth iterate of the map, where T denotes the period of the orbit to be stabilized. For the T times iterated map, any point on the periodic orbit is a fixed point, and we can then apply the discussion sections 2.1–2.3. This method is, however, overly sensitive to noise, especially when long period periodic orbits are involved. Next we outline another method which we believe should, in general, be better. In terms of the treatment of section 2.2, the prescription we give below corresponds to placing the unstable eigenvalues of the uncontrolled problem at zero, while leaving the stable eigenvalues unchanged. (This is only one of many possibilities that could be given.)

We denote the periodic orbit by $\mathbf{Z}_{i*}(p)$, where $\mathbf{Z}_{(i+T)*}(p) = \mathbf{Z}_{i*}(p)$. In addition, we introduce the set of T matrices $\mathbf{A}_i$ which are $n \times n$ and the set of T column vectors $\boldsymbol{B}_i$ which are of dimension n, where

$$\mathbf{A}_i = \mathbf{A}_{i+T} = \mathbf{D}_Z \boldsymbol{F}(\mathbf{Z}, p)\,,$$

$$\boldsymbol{B}_i = \boldsymbol{B}_{i+T} = \mathbf{D}_p \boldsymbol{F}(\mathbf{Z}, p)\,,$$

and the partial derivatives are evaluated at $\mathbf{Z} = \mathbf{Z}_{i*}(\bar{p})$ and $p = \bar{p}$.

Linearizing as in eq. (2.3), we have

$$\begin{aligned}&\mathbf{Z}_{i+1} - \mathbf{Z}_{(i+1)*}(\bar{p}) \\ &\quad = \mathbf{A}_i[\mathbf{Z}_i - \mathbf{Z}_{i*}(\bar{p})] + \boldsymbol{B}_i(p_i + \bar{p})\,. \end{aligned} \tag{2.11}$$

Say that the periodic orbit has u unstable eigenvalues (i.e., u eigenvalues with magnitude greater than one) and s stable eigenvalues, where $u + s = n$. At each point $\mathbf{Z}_{i*}(\bar{p})$ on the $p = \bar{p}$ periodic orbit, determine vectors $\{\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,s}\}$ which span the linearized stable subspace. Now let

$$\boldsymbol{\Phi}_{i,j} = \mathbf{A}_{i+u-1}\mathbf{A}_{i+u-2}\cdots\mathbf{A}_{i+j+1}\mathbf{A}_{i+j}\,,$$

for $j = 1, 2, \ldots, (u-1)$ and

$$\begin{aligned}\mathbf{C}_i = (&\boldsymbol{\Phi}_{i,1}\boldsymbol{B}_i \,\vdots\, \boldsymbol{\Phi}_{i,2}\boldsymbol{B}_{i+1} \,\vdots\, \cdots \,\vdots\, \boldsymbol{\Phi}_{i,u-1}\boldsymbol{B}_{i+u-2} \,\vdots\, \boldsymbol{B}_{i+u-1} \\ &\vdots\, \boldsymbol{v}_{i+u,1} \,\vdots\, \boldsymbol{v}_{i+u,2} \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_{i+u,s})\end{aligned}$$

(One choice of the vectors $\{\boldsymbol{v}_{i,1}, \boldsymbol{v}_{1,2}, \ldots, \boldsymbol{v}_{i,s}\}$ is the stable eigenvectors of $\mathbf{A}_i\mathbf{A}_{i-1}\ldots\mathbf{A}_{i-T-1}$.) The controllability condition (analogous to that in section 2.2) is that $\mathbf{C}_i$ be nonsingular. The desired result for the control is then specified by

$$p_i - \bar{p} = -\boldsymbol{K}_i^{\mathrm{T}}[\mathbf{Z}_i - \mathbf{Z}_{i*}(\bar{p})]\,, \tag{2.12a}$$

where

$$\boldsymbol{K}_i^{\mathrm{T}} = \boldsymbol{\kappa}\mathbf{C}_i^{-1}\boldsymbol{\Phi}_{i,0}\,, \tag{2.12b}$$

and $\boldsymbol{\kappa}$ denotes an n-dimensional row vector whose first entry is one and all of whose remaining entries are zeros.

To derive eqs. (2.12) we iterate (2.11) u times,

$$\begin{aligned}\mathbf{Z}_{i+u} - \mathbf{Z}_{(i+u)*}(\bar{p}) &= \boldsymbol{\Phi}_{i,0}[\mathbf{Z}_i - \mathbf{Z}_{i*}(\bar{p})] \\ &+ \boldsymbol{\Phi}_{i,1}\boldsymbol{B}_i(p_i - \bar{p}) + \boldsymbol{\Phi}_{i,2}\boldsymbol{B}_{i+1}(p_{i+1} - \bar{p}) \\ &+ \cdots + \boldsymbol{B}_{i+u-1}(p_{i+u-1} - \bar{p})\,. \end{aligned} \tag{2.13a}$$

We then demand that $\mathbf{Z}_{i+u}$ land on the linearized stable manifold of the periodic orbit through the point $\mathbf{Z}_{(i+u)*}(\bar{p})$. That is, we choose the p's such that there exists s coefficients $\alpha_1, \alpha_2, \ldots, \alpha_s$ such that

$$\begin{aligned}\mathbf{Z}_{i+u} - \mathbf{Z}_{(i+u)*}(\bar{p}) &= \alpha_1\boldsymbol{v}_{i+u,1} + \alpha_2\boldsymbol{v}_{i+u,2} \\ &+ \cdots + \alpha_s\boldsymbol{v}_{i+u,s}\,. \end{aligned} \tag{2.13b}$$

Regarding (2.13a) and (2.13b) as $n = u + s$ equations in the n unknowns, $p_i, p_{i+1}, \ldots, p_{i+u-1}$, $\alpha_1, \alpha_2, \ldots, \alpha_s$, we then solve for p_i to obtain (2.12).

[Note from the above that at time i we could, once and for all, calculate all the control parameter values to be applied in the next u iterates, p_i, $p_{i+1}, \ldots, p_{i+u-1}$. In the presence of noise, however, this is not a good idea (assuming $u > 1$), since it does not take advantage of the opportunity to correct for the noise on each iterate. Therefore, we believe that, in the presence of noise, it is best to perform the calculation of p_i via eq. (2.12) on each iterate.]

2.6. *Use of delay coordinates*

In experimental studies of chaotic dynamical systems, delay coordinates are often used to represent the system state. This is sometimes useful because it only requires measurement of the time series of a *single* scalar state variable which, we denote $\xi(t)$. A delay coordinate vector can be formed as follows:

$$\mathbf{Z}(t) = (\xi(t), \xi(t - T_D), \xi(t - 2T_D), \ldots,$$
$$\times \xi(t - MT_D)),$$

where T_D is some conveniently chosen delay time, and the time variable t is assumed continuous. Embedding theorems guarantee that for $M \geq 2n$, where n is the system dimensionality, the vector $\mathbf{Z}$ is generically a global one-to-one representation of the system state. (Actually, for our purposes, we do not require a global embedding; we only require $\mathbf{Z}$ to be one-to-one in the small region near the periodic orbit, and this can typically be achieved with $M = n - 1$.) To obtain a map, one can take a Poincaré surface of section. For the often encountered case of a system which is periodically forced at a period T_F, one can define a "stroboscopic surface of section" by sampling the state at discrete times $t_i = iT_F + t_0$. In this case we have the discrete state variable

$$\mathbf{Z}_i = \mathbf{Z}(t_i).$$

As pointed out by Dressler and Nitsche [19], in the presence of parameter variation, delay coordinates lead to a map of a different form than

$$\mathbf{Z}_{i+1} = \mathbf{F}(\mathbf{Z}_i, p_i),$$

which is the form assumed in sections 2.1–2.5. For example, in the periodically forced case, since the components of $\mathbf{Z}_i$ are $\xi(t_i - mT_D)$ for $m = 0, 1, \ldots, M$, the vector $\mathbf{Z}_{i+1}$ must depend not only on p_i, but also on all previous values of the parameter that were in effect during the time interval $t_i \leq t \leq t_i - MT_D$. In particular, let r be the smallest integer such that $MT_D < rT_F$. Then the relevant map is in general of the form

$$\mathbf{Z}_{i+1} = \mathbf{G}(\mathbf{Z}_i, p_i, p_{i-1}, \ldots, p_{i-r}). \tag{2.14a}$$

For $r = 1$ we have

$$\mathbf{Z}_{i+1} = \mathbf{G}(\mathbf{Z}_i, p_i, p_{i-1}). \tag{2.14b}$$

We now discuss how the technique of section 2.2 can be applied in the case of delay coordinates, and, for simplicity, we limit the discussion to $r = 1$, eq. (2.14b). Linearizing as in eq. (2.3) and again restricting our attention to the case of a fixed point orbit, we have

$$\mathbf{Z}_{i+1} - \mathbf{Z}_*(\bar{p}) = \mathbf{A}[\mathbf{Z}_i - \mathbf{Z}_*(\bar{p})]$$
$$+ \boldsymbol{B}_a(p_i - \bar{p}) + \boldsymbol{B}_b(p_{i-1} - \bar{p}), \tag{2.15}$$

where $\mathbf{A} = \mathbf{D}_Z\mathbf{G}(\mathbf{Z}, p, p')$, $\boldsymbol{B}_a = \mathbf{D}_p\mathbf{G}(\mathbf{Z}, p, p')$, and $\boldsymbol{B}_b = \mathbf{D}_{p'}\mathbf{G}(\mathbf{Z}, p, p')$, and all partial derivatives are evaluated at $\mathbf{Z} = \mathbf{Z}_*(\bar{p})$ and $p = \bar{p} = p'$.

Now define a new state variable with one extra component by

$$\tilde{\mathbf{Z}}_{i+1} = \begin{pmatrix} \mathbf{Z}_{i+1} \\ p_i \end{pmatrix}, \tag{2.16}$$

and introduce the linear control law,

$$p_i - \bar{p} = -\boldsymbol{K}^{\mathrm{T}}[\boldsymbol{Z}_i - \boldsymbol{Z}_*(\bar{p})] - k(p_{i-1} - \bar{p}) . \tag{2.17}$$

Combining these equations, we obtain

$$\tilde{\boldsymbol{Z}}_{i+1} - \tilde{\boldsymbol{Z}}_*(\bar{p}) = (\tilde{\mathbf{A}} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{K}}^{\mathrm{T}})[\tilde{\boldsymbol{Z}}_i - \tilde{\boldsymbol{Z}}_*(\bar{p})] , \tag{2.18}$$

where

$$\tilde{\boldsymbol{Z}}_*(\bar{p}) = \begin{pmatrix} \boldsymbol{Z}_*(\bar{p}) \\ \bar{p} \end{pmatrix}, \quad \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \boldsymbol{B}_{\mathrm{b}} \\ \mathbf{0} & 0 \end{pmatrix},$$

$$\tilde{\boldsymbol{B}} = \begin{pmatrix} \boldsymbol{B}_{\mathrm{a}} \\ 1 \end{pmatrix}, \quad \tilde{\boldsymbol{K}} = \begin{pmatrix} \boldsymbol{K} \\ k \end{pmatrix}.$$

Since (2.18) is now of the same form as (2.3), the method of section 2.2 can be applied. (A similar result for any $r > 1$ also clearly holds.)

Another method of control for delay coordinates is to reduce (2.14b) directly to the form $\boldsymbol{Z}_{i+1} = \boldsymbol{F}(\boldsymbol{Z}_i, p_i)$ and then proceed as in sections 2.1 and 2.2. This reduction can be done by setting $p_i \equiv \bar{p}$ for every other time step. For example, say $p_i = 0$ for i odd, and $j = \frac{1}{2}i$ for even i. Then making the replacements $\boldsymbol{Z}_i \to \hat{\boldsymbol{Z}}_j$, $p_i \to \hat{p}_j$ for even i, and iterating (2.14b) twice we have

$$\begin{aligned} \hat{\boldsymbol{Z}}_{j+1} &= \boldsymbol{G}[\boldsymbol{G}(\hat{\boldsymbol{Z}}_j, \bar{p}, \hat{p}_j), \hat{p}_j, \bar{p}] \\ &\equiv \hat{\boldsymbol{F}}(\hat{\boldsymbol{Z}}_j, \hat{p}_j) , \end{aligned} \tag{2.19}$$

which is of the required form. We believe, however, that the first method we have given [i.e., that based on eq. (2.18)] should usually be capable of yielding superior results to the method based on (2.19) with respect to noise sensitivity and time to achieve control. This is because our second method does not take advantage of the opportunity to control on each time iterate while our first method does.

3. Controlling the Hénon map

In ref. [6], the authors used the Hénon map to illustrate the control method and, in particular, to test their theoretical predictions concerning the average time to achieve control. As already pointed out, their work is based on a particular choice of the gain matrix $\boldsymbol{K}^{\mathrm{T}}$. In this section we consider how different choices of $\boldsymbol{K}^{\mathrm{T}}$ affect the average time to achieve control for the Hénon map.

The Hénon map [13] is the two-dimensional map

$$\boldsymbol{Z} \mapsto \boldsymbol{Z}' = \boldsymbol{F}(\boldsymbol{Z}) ,$$

defined by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a - x^2 + by \\ x \end{pmatrix},$$

where $(x, y) \in \mathbb{R} \times \mathbb{R}$. We keep the parameter b fixed throughout ($b = 0.3$) and allow the control parameter a to vary around a nominal value $\bar{a}$ ($\bar{a} = 1.4$) for which the map has a chaotic attractor.

For $a = \bar{a} = 1.4$ there is an unstable saddle fixed point contained in the chaotic attractor. This fixed point is located at

$$\boldsymbol{Z}_*(a) = x_*(a)\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$x_*(a) = -c + (c^2 + a)^{1/2} , \quad c = \tfrac{1}{2}(1 - b) ,$$

for $a \geq -c^2$. Noting that the Jacobian matrix of partial derivatives of the map is

$$\mathbf{D}_{\boldsymbol{Z}}\boldsymbol{F}(\boldsymbol{Z}) = \begin{pmatrix} -2x & b \\ 1 & 0 \end{pmatrix},$$

and that the stability of the fixed point is determined by the roots of the characteristic equation

$$|\mathbf{D}_{\boldsymbol{Z}}\boldsymbol{F}[\boldsymbol{Z}_*(a)] - s\mathbf{I}| = 0 ,$$

one can easily check that the fixed point is stable for $-c^2 < a < 3c^2$ and unstable for $a > 3c^2$. (Hence the fixed point is unstable for $b = 0.3$, $a = \bar{a} = 1.4$ since $c = 0.35$.)

The quantities that appear in section 2.2 are as follows:

$$\mathbf{A} = \begin{pmatrix} -2\bar{x}_* & b \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\mathbf{C} = (\boldsymbol{B} : \mathbf{A}\boldsymbol{B}) = \begin{pmatrix} 1 & -2\bar{x}_* \\ 0 & 1 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 2\bar{x}_* & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{K}^{\mathrm{T}} = (\alpha_1 - a_1 \quad \alpha_2 - a_2),$$

where

$$a_1 = 2\bar{x}_* = -(\lambda_{\mathrm{u}} + \lambda_{\mathrm{s}}), \quad a_2 = -b = \lambda_{\mathrm{u}}\lambda_{\mathrm{s}},$$

and

$$\alpha_1 = -(\mu_1 + \mu_2), \quad \alpha_2 = \mu_1\mu_2 .$$

Here $\bar{x}_* = x_*(\bar{a})$, and λ_{u} and λ_{s} are the eigenvalues of matrix $\mathbf{A}$,

$$\left.\begin{matrix}\lambda_{\mathrm{s}} \\ \lambda_{\mathrm{u}}\end{matrix}\right\} = -\bar{x}_* \pm (\bar{x}_*^2 + b)^{1/2} .$$

The quantities μ_1 and μ_2 are the regulator poles [i.e., the eigenvalues of $(\mathbf{A} - \boldsymbol{B}\boldsymbol{K}^{\mathrm{T}})$].

In order to better illustrate the different choices of regulator poles or, equivalently, of the matrix $\boldsymbol{K}^{\mathrm{T}}$, we have used the plane (α_1, α_2) [cf. fig. 1]. In this plane we have plotted the lines of marginal stability $\mu_1 = \pm 1$ $(1 \pm \alpha_1 + \alpha_2 = 0)$ and $\mu_1\mu_2 = 1$ $(\alpha_2 = 1)$; the bounded triangular region delimited by these lines (shown shaded in the figure) is the region where the regulator poles are stable. In addition, we have plotted as dashed lines the axes $(k_1, k_2) = \boldsymbol{K}^{\mathrm{T}}$ which are related to (α_1, α_2) by the translations

$$k_1 = \alpha_1 - a_1, \quad k_2 = \alpha_2 - a_2 .$$

The straight solid line in the figure going through the origin of the (k_1, k_2) plane has slope $-\lambda_{\mathrm{s}}$ and intersects the line $\alpha_2 = 0$ at the point Q with coordinates $(\alpha_1, \alpha_2) = (-\lambda_{\mathrm{s}}, 0)$. To this point

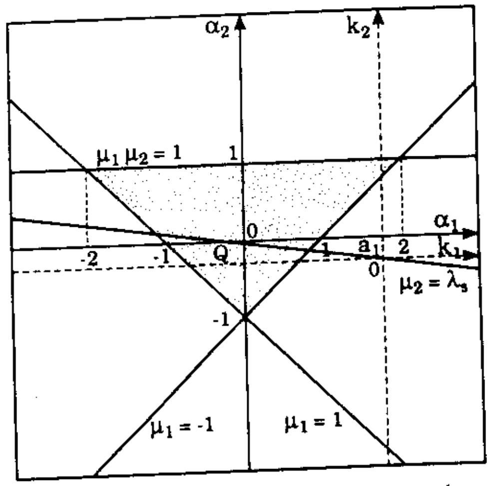


Fig. 1. Hénon map: choice of regulator poles.

corresponds the regulator poles

$$\mu_1 = 0, \quad \mu_2 = \lambda_{\mathrm{s}},$$

and the matrix

$$\boldsymbol{K}^{\mathrm{T}} = \lambda_{\mathrm{u}}(1 \quad -\lambda_{\mathrm{s}}) \equiv \boldsymbol{K}_Q^{\mathrm{T}} .$$

$\boldsymbol{K}_Q^{\mathrm{T}}$ is the special choice of matrix $\boldsymbol{K}^{\mathrm{T}}$ made in ref. [6].

Before proceeding with the discussion, it is convenient to express the vector $\boldsymbol{K}^{\mathrm{T}}$ in polar coordinates

$$\boldsymbol{K}^{\mathrm{T}} = |\boldsymbol{K}^{\mathrm{T}}|(\cos\theta, \sin\theta) .$$

We consider the following two ways of varying the vector $\boldsymbol{K}^{\mathrm{T}}$ (inside the triangular region of stability):

(I) θ fixed, $|\boldsymbol{K}^{\mathrm{T}}|$ variable.

(II) $|\boldsymbol{K}^{\mathrm{T}}|$ fixed, θ variable.

In terms of the control slab defined by eq. (2.8) we have that in situation (I) the slab is kept orientated in a fixed direction while its width $w = 2\delta/|\boldsymbol{K}^{\mathrm{T}}|$ varies, whereas in situation (II) the direction of the slab is rotated while its width is kept fixed at $w = 2\delta/|\boldsymbol{K}_Q^{\mathrm{T}}|$. The choice of the $\boldsymbol{K}^{\mathrm{T}}$ in ref. [6] has $\theta = \theta_Q \equiv \tan^{-1}(-\lambda_{\mathrm{s}})$ and, as we shall see, this choice is optimal from the point of view of the time to achieve control. (To see that the choice of ref. [6] corresponds to $\theta = \theta_Q$, we note that with this choice one obtains a convergence rate to the periodic orbit of μ_s as in ref. [6].)

In the numerical experiments we calculated the average time to achieve control by the method described in Appendix B. We also allowed for different values of the maximum amplitude of the parameter perturbations, δ.

First we consider the case where θ is fixed (case I) at the value

$$\theta = \theta_Q .$$

This case has a simple interpretation in terms of regulator poles: $\mu_2 = \lambda_s$ is kept fixed while μ_1 is allowed to vary between -1 and $+1$. μ_1 and $|\boldsymbol{K}^{\mathrm{T}}|$ are related by

$$|\boldsymbol{K}^{\mathrm{T}}| = |\mu_1 - \lambda_u|(1 + \lambda_s^2)^{1/2} .$$

Fig. 2 shows results for $\langle \tau \rangle$ for this case. We see that the average time to achieve control increases with μ_1, although only moderately.

Fig. 3 shows results for $\langle \tau \rangle$ versus θ for $|\boldsymbol{K}|^{\mathrm{T}}$ held fixed (case II) at

$$|\boldsymbol{K}^{\mathrm{T}}| = |\boldsymbol{K}_Q^{\mathrm{T}}| = |\lambda_u|(1 + \lambda_s^2)^{1/2} .$$

We see that the average time to achieve control has a strong minimum at $\theta = \theta_Q$.

Fig. 4 shows $\langle \tau \rangle$ versus $|\boldsymbol{K}^{\mathrm{T}}|$ for three values of θ, $\theta = \theta_0$, $\theta = \theta_Q$, and $\theta = \theta_1$, where $\theta_0 < \theta_Q < \theta_1$ and θ_0 and θ_1 are close to θ_Q ($\theta_0 = 170.4°$, $\theta_Q = 171.1°$, $\theta_1 = 172.0°$). We observe that the $\theta = \theta_Q$ result is always below the results for $\theta = \theta_0$ and $\theta = \theta_1$ indicating that the average time

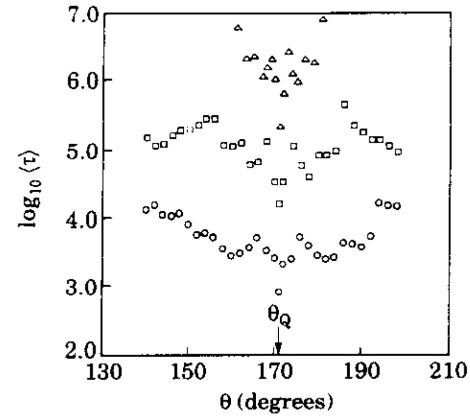


Fig. 3. Hénon map: $\log_{10}\langle \tau \rangle$ versus θ, with $|\mathbf{K}^{\mathrm{T}}| = |\mathbf{K}_Q^{\mathrm{T}}|$, for (○) $\delta = 10^{-2}$, (□) $\delta = 10^{-3}$, (△) $\delta = 10^{-4}$ ($\bar{a} = 1.4$, $b = 0.3$).

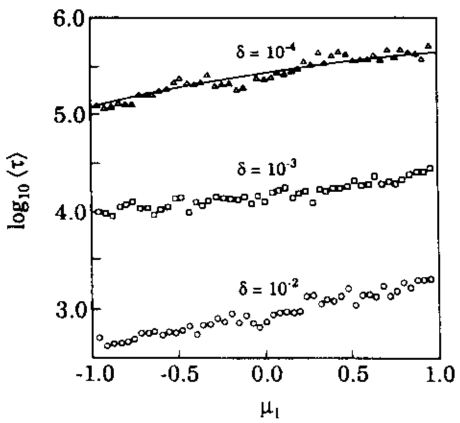


Fig. 2. Hénon map: $\log_{10}\langle \tau \rangle$ versus μ_1, with $\mu_2 = \lambda_s$, for (○) $\delta = 10^{-2}$, (□) $\delta = 10^{-3}$, (△) $\delta = 10^{-4}$. The theoretical curve was calculated using eq. (A.9) of appendix A. ($\bar{a} = 1.4$, $b = 0.3$).

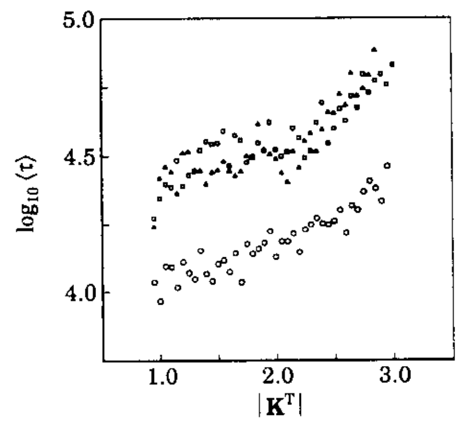


Fig. 4. Hénon map: $\log_{10}\langle \tau \rangle$ versus $|\boldsymbol{K}^{\mathrm{T}}|$ for (○) $\theta = \theta_Q$, (□) $\theta = \theta_0$, (△) $\theta = \theta_1$ ($\theta_0 < \theta_Q < \theta_1$; $\theta_0 = \tan^{-1}[-\lambda_u\lambda_s/(\lambda_u + \lambda_s)] = 170.4°$, $\theta_Q = \tan^{-1}(-\lambda_s) = 171.1°$, $\theta_1 = 172.0°$), and $\delta = 10^{-3}$.

to achieve control has a strong minimum at $\theta = \theta_Q$ not only for $|K^{\mathrm{T}}| = |K_Q^{\mathrm{T}}|$ but for all values of $|K^{\mathrm{T}}|$. Thus the condition $\theta = \theta_Q$ is optimal.

In Appendix A we show how the average time to achieve control can be obtained theoretically in the case of two-dimensional maps and verify that there is excellent agreement between the theoretical and experimental results in the case of the Hénon map.

4. Controlling the double rotor

In this section we apply the control method described in section 2 to a dynamical system known as the double rotor map. We start by deriving the map (section 4.1 and Appendix B), then study its fixed points (section 4.2) and its attractors (section 4.3), including chaotic ones, and finally proceed to control some of the fixed points embedded in one of the chaotic attractors (sections 4.4 and 4.5).

4.1. The double rotor map

The double rotor map is a four-dimensional map which describes the time evolution of a mechanical system known as the kicked double rotor [14]. This system is a four-dimensional extension of the kicked (single) rotor, a two-dimensional system that is described by the well-known dissipative standard map [20].

The double rotor is composed of two thin, massless rods connected as shown in fig. 5. The first rod, of length l_1, pivots about P_1 (which is fixed), and the second rod, of length $2l_2$, pivots about P_2 (which moves). The angles $\theta_1(t)$, $\theta_2(t)$ specify the orientations at time t of the first and second rods, respectively. A mass m_1 is attached at P_2, and masses $\frac{1}{2}m_2$ are attached to each end of the second rod (P_3 and P_4). Friction at P_1 (with coefficient ν_1) slows the first rod at a rate proportional to its angular velocity $\dot{\theta}_1(t) \equiv d\theta_1(t)/dt$; friction at P_2 (with coefficient ν_2) slows the second rod (and simultaneously accelerates the

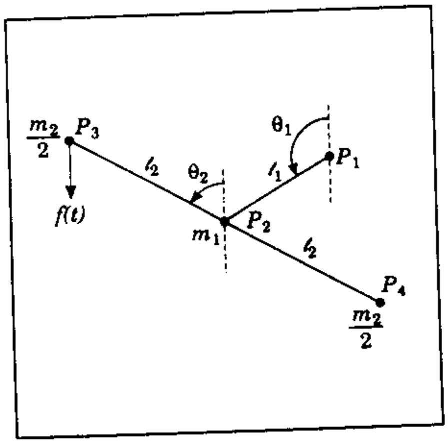


Fig. 5. The double rotor.

first rod) at a rate proportional to $\dot{\theta}_2(t) - \dot{\theta}_1(t)$. The end of the second rod marked P_3 receives periodic impulse kicks at times $t = T, 2T, \ldots$, always from the same direction and with constant strength f_0. There is no gravity.

In Appendix C we write the differential equations that describe the kicked double rotor and proceed to derive from them the *double rotor map* relating the state of the system just after consecutive kicks. We obtain the four-dimensional map

$$\mathbf{Z} \mapsto \mathbf{Z}' = \boldsymbol{F}(\mathbf{Z}),$$

defined by

$$\begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} \mathbf{M}Y + X \\ \mathbf{L}Y + G(X') \end{pmatrix}, \tag{4.1}$$

where

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathrm{S}^1 \times \mathrm{S}^1, \quad Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R} \times \mathbb{R},$$

and

$$G(X') = \begin{pmatrix} c_1 \sin x_1' \\ c_2 \sin x_2' \end{pmatrix}. \tag{4.2}$$

x_1, x_2 are the angular positions of the rods at the instant of the kth kick, $x_j = \theta_j(kT)$, while y_1, y_2

are the angular velocities of the rods immediately after the kth kick, $y_i = \dot{\theta}_i(kT^+)$. S^1 is the circle $\mathbb{R}(\mathrm{mod}\, 2\pi)$. $\mathbf{L}$ and $\mathbf{M}$ are constant 2×2 matrices defined by

$$\mathbf{L} = \sum_{j=1}^{2} \mathbf{W}_j \, e^{\lambda_j T}, \quad \mathbf{M} = \sum_{j=1}^{2} \mathbf{W}_j \, \frac{e^{\lambda_j T} - 1}{\lambda_j},$$

$$\mathbf{W}_1 = \begin{pmatrix} a & b \\ b & d \end{pmatrix}, \quad \mathbf{W}_2 = \begin{pmatrix} d & -b \\ -b & a \end{pmatrix},$$

$$a = \frac{1}{2}\left(1 + \frac{\nu_1}{\Delta}\right), \quad d = \frac{1}{2}\left(1 - \frac{\nu_1}{\Delta}\right), \quad b = -\frac{\nu_2}{\Delta},$$

$$\left.\begin{matrix}\lambda_1\\ \lambda_2\end{matrix}\right\} = -\tfrac{1}{2}(\nu_1 + 2\nu_2 \pm \Delta),$$

$$\Delta = (\nu_1^2 + 4\nu_2^2)^{1/2}.$$

Finally, c_1 and c_2 are given by

$$c_j = \frac{f_0}{I} l_j, \quad j = 1, 2,$$

where

$$I = (m_1 + m_2) l_1^2 = m_2 l_2^2.$$

The following relation between matrices $\mathbf{L}$ and $\mathbf{M}$ will be useful below:

$$\mathbf{L} = \mathbf{I} + \mathbf{A}_\nu \mathbf{M}, \tag{4.3}$$

where

$$\mathbf{A}_\nu = \begin{pmatrix} -(\nu_1 + \nu_2) & \nu_2 \\ \nu_2 & -\nu_2 \end{pmatrix}.$$

(λ_1, λ_2 are precisely the eigenvalues of $\mathbf{A}_\nu$.) Note also that

$$|\mathbf{L}| = e^{(\lambda_1 + \lambda_2)T},$$

$$|\mathbf{M}| = \frac{e^{\lambda_1 T} - 1}{\lambda_1} \, \frac{e^{\lambda_2 T} - 1}{\lambda_2},$$

$$|\mathbf{A}_\nu| = \nu_1 \nu_2.$$

From now on we assume that $\nu_1 = \nu_2 \equiv \nu$. This leads to

$$\left.\begin{matrix}\lambda_1\\ \lambda_2\end{matrix}\right\} = -\tfrac{1}{2}\nu(3 \pm \sqrt{5}),$$

$$\left.\begin{matrix}a\\ d\end{matrix}\right\} = \tfrac{1}{2}(1 \pm \tfrac{1}{5}\sqrt{5}),$$

$$b = -\tfrac{1}{5}\sqrt{5}.$$

In all the numerical work described in the rest of this section the parameters ν, T, I, m_1, m_2, l_1, and l_2 were kept fixed at the values

$$\nu = T = I = m_1 = m_2 = l_2 = 1, \quad l_1 = 1/\sqrt{2}. \tag{4.4}$$

The only parameter which we shall vary is the forcing f_0 used as the control parameter.

4.2. Fixed points of the double rotor map

The fixed points $\boldsymbol{Z}_* = (\boldsymbol{X}_*, \boldsymbol{Y}_*)$ of the map (4.1) are solutions of the system

$$\boldsymbol{X}_* = \mathbf{M}\boldsymbol{Y}_* + \boldsymbol{X}_* - 2\pi \boldsymbol{N}, \quad \boldsymbol{Y}_* = \mathbf{L}\boldsymbol{Y}_* + \boldsymbol{G}(\boldsymbol{X}_*), \tag{4.5}$$

where the components of the vector $\boldsymbol{N} = (n_1, n_2)$ are integer and are the rotation numbers in the x_1, x_2 variables. The rotation numbers n_1, n_2 are defined as the multiples of 2π by which x_{1*}, x_{2*} are increased in one iteration of the map before being brought to the interval $[0, 2\pi]$. From eqs. (4.5) we obtain, using (4.3),

$$\boldsymbol{Y}_* = 2\pi \mathbf{M}^{-1} \boldsymbol{N}, \quad \boldsymbol{G}(\boldsymbol{X}_*) = -2\pi \mathbf{A}_\nu \boldsymbol{N}. \tag{4.6}$$

Using the definitions of the matrices $\boldsymbol{G}$ and $\mathbf{A}_\nu$ we rewrite the second of the eqs. (4.6) in the form

$$\begin{pmatrix} \sin x_{1*} \\ \sin x_{2*} \end{pmatrix} = -\frac{2\pi\nu I}{f_0} \begin{pmatrix} (1/l_1)(-2n_1 + n_2) \\ (1/l_2)(n_1 - n_2) \end{pmatrix} \equiv \frac{1}{f_0} \begin{pmatrix} f_{01} \\ f_{02} \end{pmatrix}. \tag{4.7}$$

where the identity on the right defines the two new quantities f_{01} and f_{02}. These equations show that for each pair of rotation numbers (n_1, n_2) a set of four possible solutions for (x_{1*}, x_{2*}) exists if $|f_0| \geq |f_{0c}|$, where $|f_{0c}| = \max(|f_{01}|, |f_{02}|)$. The four fixed points correspond to the four combinations of values of (x_{1*}, x_{2*}) that have the same pair of values of $(\sin x_{1*}, \sin x_{2*})$. When necessary we will use the notation

$$Z_*^{[N;q]} = (X_*^{[N;q]}, Y_*^{[N]}),$$

or more simply $[N; q]$, to identify the fixed points, where the index q labels the four possible solutions of (4.7) ($q = 1, 2, 3, 4$) and, as shown in fig. 6, corresponds to the ordering

$$x_{1*}^{[N;1]} = x_{1*}^{[N;2]} < x_{1*}^{[N;3]} = x_{1*}^{[N;4]},$$
$$x_{2*}^{[N;1]} = x_{2*}^{[N;3]} < x_{2*}^{[N;2]} = x_{2*}^{[N;4]}$$

Note that $Y_*^{[N]} = (y_{1*}^{[N]}, y_{2*}^{[N]})$ is the same for the four fixed points (i.e., it does not depend on q). Eqs. (4.7) also show that for $|f_0| \geq |f_{0c}|$, $(n_1, n_2) \neq (0, 0)$, another set of four fixed points exists with rotation numbers $(-n_1, -n_2)$. It is easy to see that to each point $(x_{1*}, x_{2*}, y_{1*}, y_{2*})$ of the first set corresponds a point of the second set given by $(2\pi - x_{1*}, 2\pi - x_{2*}, -y_{1*}, -y_{2*})$. This is a reflection of the fact that the double rotor map (4.1) itself is invariant under the change of variables $(x_1, x_2, y_1, y_2) \mapsto (2\pi - x_1, 2\pi - x_2, -y_1, -y_2)$.

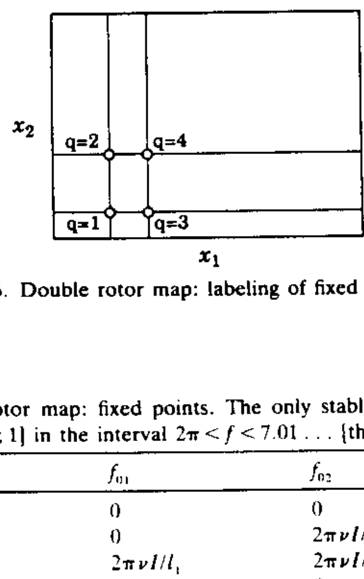


Fig. 6. Double rotor map: labeling of fixed points.

In table 1 we summarize the properties of the five sets of fixed points (36 fixed points) with smaller values of f_{0c} (when the other parameters of the map take the values specified by eqs. (4.4)), with rotation numbers $N = (0,0)$, $\pm(1,2)$, $\pm(0,1)$, $\pm(1,1)$, $\pm(2,3)$. Note that the last three sets have the same value of f_{0c}. In fig. 7 we have plotted these fixed points in the plane (x_1, x_2). Their (y_1, y_2) coordinates are given by the first of eqs. (4.6).

Let us now turn our attention to the stability of the fixed points. The basic element of the analysis is the Jacobian (4×4) matrix of partial derivatives of the map (4.1),

$$\mathbf{D}_Z F(Z) = \begin{pmatrix} \mathbf{I}_2 & \mathbf{M} \\ \mathbf{H}(X') & \mathbf{L} + \mathbf{H}(X')\mathbf{M} \end{pmatrix},$$

where

$$\mathbf{H}(X') = \mathbf{D}_{x'} G(X') = \begin{pmatrix} c_1 \cos x_1' & 0 \\ 0 & c_2 \cos x_2' \end{pmatrix},$$

and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The characteristic polynomial of $\mathbf{D}_Z F(Z_*)$ is

$$\begin{aligned} P(s) &= |\mathbf{D}_Z F(Z_*) - s\mathbf{I}_4| \\ &= |s^2 \mathbf{I}_2 - s(\mathbf{I}_2 + \mathbf{L} + \mathbf{HM}) + \mathbf{L}|, \end{aligned} \tag{4.8a}$$

Table 1
Double rotor map: fixed points. The only stable fixed points are: $[(0,0); 4]$ in the interval $0 < f_0 < 4.27\ldots$; $[(1,2); 4]$ and $[(-1,-2); 1]$ in the interval $2\pi < f < 7.01\ldots$ [the other parameters are given by eq. (4.4)].

(n_1, n_2)	f_{01}	f_{02}	f_{0c}	h_{11c}	h_{22c}
$(0,0)$	0	0	0	0	0
$\pm(1,2)$	0	$2\pi\nu I/l_2$	$2\pi\nu I/l_2$	$\pm 2\pi\nu l_1/l_2$	0
$\pm(0,1)$	$2\pi\nu I/l_1$	$2\pi\nu I/l_2$	$2\pi\nu I/l_1$	0	$\pm 2\pi\nu(l_2^2/l_1^2 - 1)^{1/2}$
$\pm(1,1)$	$2\pi\nu I/l_1$	0	$2\pi\nu I/l_1$	0	$\pm 2\pi\nu l_2/l_1$
$\pm(2,3)$	$2\pi\nu I/l_1$	$2\pi\nu I/l_2$	$2\pi\nu I/l_1$	0	$\pm 2\pi\nu(l_2^2/l_1^2 - 1)^{1/2}$

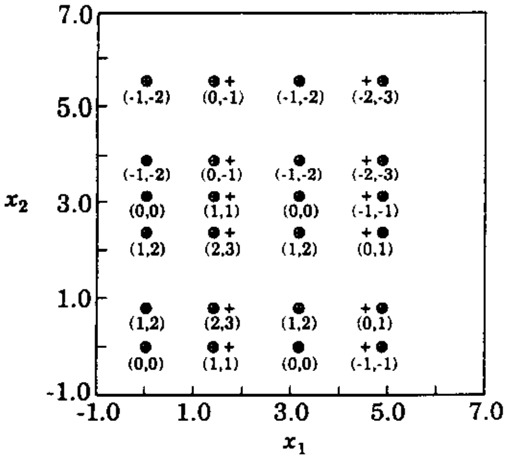


Fig. 7. Double rotor map: fixed points with rotation numbers (n_1, n_2). The symbol (+) denotes fixed points with one unstable eigendirection, while the symbol ⊕ denotes fixed points with two unstable eigendirections. [$f_0 = 9.0$, other parameters given by eq. (4.4)].

where, for simplicity, we have set $\mathbf{H} \equiv \mathbf{H}(X_*)$. The characteristic equation

$$P(s) = 0\,, \tag{4.8b}$$

determines the stability of the fixed points: if all the four roots have modulus smaller than one, the fixed point is stable. The stability of the fixed points as determined from eqs. (4.8) is discussed in appendix D.

For $f_0 = 9.0$, the nominal value in the control experiments of sections 4.4 and 4.5, all the fixed points are unstable. We have indicated in fig. 7 the number of unstable eigendirections at each fixed point.

We observe, from eq. (4.7), that as the forcing f_0 increases, the number of fixed points increases without bound. Not all these fixed points are necessarily embedded in the chaotic attractor, but those that are embedded in it are necessarily unstable. Furthermore, we find that the fixed points are roughly spread throughout the attractor, suggesting that there can be substantial flexibility to select among a variety of asymptotic behaviors by selecting different fixed points for control. (Even more flexibility can be achieved if we also consider periodic orbits of period greater than one.)

4.3. Bifurcation diagram

A bifurcation diagram shows how the attractors of a dynamical system change with a system parameter.

In figs. 8a, 8b we present a bifurcation diagram for the double rotor map, which was obtained in the following way. For each value of the parameter we took a large number of initial

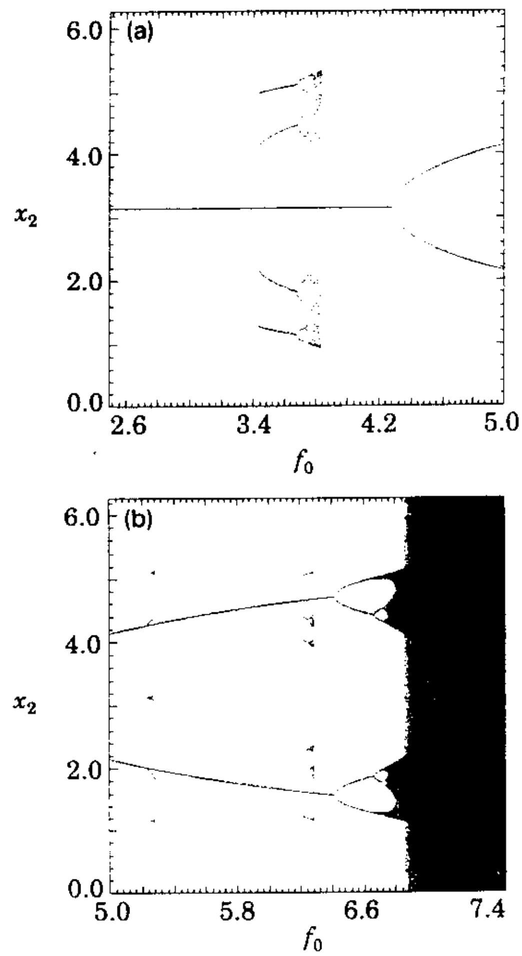


Fig. 8. (a), (b) Double rotor map: bifurcation diagram [parameters given by eq. (4.4): number of values of f_0 in each figure: 251; number of initial conditions for each f_0: 625; snapshot taken after 6000 iterations].

58

angles (x_1, x_2) with both x_1 and x_2 distributed uniformly in $[0, 2\pi]$ and iterated them starting with zero angular velocity [i.e., $(y_1, y_2) = (0, 0)$]. After iterating a sufficient number of times so that the orbits are essentially on the attractor, we plotted the x_2 component of all the orbits.

The diagram clearly exhibits a main branch that develops continuously for all values of the parameter. This main branch illustrates a period-doubling bifurcation sequence to chaos: a period-1 periodic orbit bifurcates (at $f_0 \simeq 4.27$) to a period-2 periodic orbit which then bifurcates (at $f_0 \simeq 6.42$) to a period-4 periodic orbit which then bifurcates (at $f_0 \simeq 6.67$) to a period-8 periodic orbit, and so on, with an accumulation point of period doublings at $f_0 \simeq 6.75$ beyond which chaos appears. The period-2 periodic orbit in the sequence results from the bifurcation of the stable orbit $\mathbf{Z}_*^{[(0,0);4]} = (\pi, \pi, 0, 0)$ discussed in section 4.2 which exists for $f_0 \geq 0$; at the value $f_{0u}^{[(0,0);4]}$ at which this orbit becomes unstable the stable period-2 periodic orbit is born.

Although it cannot be seen in the diagram, this period doubling sequence is peculiar in the following sense: what appears to be a period-2^m periodic orbit, $m \geq 2$, is in fact 2 period-2^{m-1} periodic orbits. This is a consequence of the symmetry of the double rotor map that forces the period-1 orbit to become unstable (at $f_0 \cong 4.2$) through an eigenvalue 1 instead through -1 as occurs in the normal period doubling bifurcation (an example of which is the bifurcation of the period-1 periodic orbit).

Besides this main branch, there are other period doubling sequences, one of which starts with a period-4 periodic orbit (at $f_0 \simeq 3.42$) and ends with a crisis (at $f_0 \simeq 3.84$). (A *crisis* is the sudden disappearance of a chaotic attractor by collision with an unstable periodic orbit [15, 16].)

It is convenient to have some quantitative characterization of the chaotic attractors revealed by the bifurcation diagram. For this purpose we introduce the spectrum of *Lyapunov exponents*, defined as follows [21, 22].

Consider an n-dimensional map $\mathbf{Z} \mapsto \mathbf{F}(\mathbf{Z})$ and its Jacobian matrix of partial derivatives $\mathbf{J}(\mathbf{Z}) = \mathbf{D}_Z \mathbf{F}(\mathbf{Z})$. Consider also the sequence $\{\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{k-1}\}$ generated by successive iteration of the initial condition $\mathbf{Z}_0$. For this sequence introduce the matrix

$$\mathbf{J}_k = \mathbf{J}(\mathbf{Z}_{k-1})\mathbf{J}(\mathbf{Z}_{k-2}) \ldots \mathbf{J}(\mathbf{Z}_1)\mathbf{J}(\mathbf{Z}_0) .$$

Now let

$$\zeta_1(k) \geq \zeta_2(k) \geq \cdots \geq \zeta_n(k) ,$$

denote the n eigenvalues of $(\mathbf{J}_k^T \mathbf{J}_k)^{1/2}$, where $\mathbf{J}_k^T$ is the transpose of $\mathbf{J}_k$. The Lyapunov numbers of the map are then defined by

$$\eta_j = \lim_{k \to \infty} [\zeta_j(k)]^{1/k} , \quad j = 1, \ldots, n ,$$

where the positive real kth root is taken. They satisfy the same ordering as the $\zeta_j(k)$, $j = 1, \ldots, n$. The Lyapunov exponents are the logarithms of the Lyapunov numbers,

$$L_j = \log_e \eta_j , \quad j = 1, \ldots, n ,$$

satisfying the same ordering

$$L_1 \geq L_2 \geq \cdots \geq L_n .$$

Hence, for chaotic attractors of an n-dimensional map there are n Lyapunov exponents, L_j, $j = 1, \ldots, n$. A chaotic attractor is defined to be one which possesses a positive Lyapunov exponent, $L_1 > 0$.

For typical dynamical systems the Lyapunov exponents are the same for almost all initial conditions on the basin of attraction of the attractor. (This is true in particular for the chaotic attractors of the double rotor map for which we calculated Lyapunov exponents; these results are reported below.) Thus the spectrum of Lyapunov exponents may be indeed considered to be a property of the attractor. For maps such that the determinant of the Jacobian matrix is independent of the variable $\mathbf{Z}$ the Lyapunov exponents satisfy the identity

$$\sum_{j=1}^{n} L_j = \log_e |\mathbf{J}| .$$

This is true in the case of the double rotor map for which we have

$$\sum_{j=1}^{4} L_j = \log_e |\mathbf{L}| = (\lambda_1 + \lambda_2)T = -3\nu ,$$

the last equality applying when $\nu_1 = \nu_2 \equiv \nu$ and $T = 1$.

From the spectrum of Lyapunov exponents define the *Lyapunov dimension*.

$$d_L = k_L + \frac{\sum_{j=1}^{k_L} L_j}{|L_{k_L+1}|} ,$$

where $1 \le k_L \le n-1$ is the largest integer for which $\sum_{j=1}^{k_L} L_j \ge 0$. If $L_1 < 0$, we define $d_L = 0$; if $\sum_{j=1}^{n} L_j \ge 0$, we define $d_L = n$. (Note that $d_L = n$ is not possible in the case of dissipative systems for which $\log_e |\mathbf{J}| < 0$.) Kaplan and Yorke [23, 24] conjecture that d_L, as given above in terms of the Lyapunov exponents, is typically equal to the fractal dimension of the support of the measure of the attractor (the information dimension).

We have numerically calculated the Lyapunov exponents and the Lyapunov dimension of the chaotic attractor in the main branch of the bifurcation diagram as a function of the forcing f_0. We used the method described in refs. [21, 22] to calculate the exponents of a large number of orbits in the basin of attraction and then took the average of these values. The results of the calculation at evenly spaced values along the f_0 axis are shown in fig. 9. The Lyapunov dimension first becomes positive at the onset of chaos ($f_0 \simeq 6.75$). The attractor dimension goes through the integer values $d_L = 2$ and 3 at $f_0 \simeq 6.88$ and 12.7, respectively.

In the numerical experiments on control that we describe in sections 4.4 and 4.5 we took $\bar{f}_0 = 9.0$ as the nominal value of the control parameter. In Table 2 we list the corresponding values of the four Lyapunov exponents and the Lyapunov dimension. In order to illustrate the point made above regarding the fact that the Lyapunov exponents are the same for almost all initial conditions on the basin of attraction of the

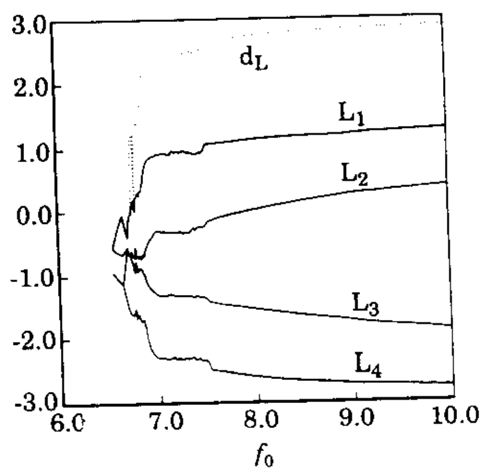


Fig. 9. Double rotor map: spectrum of Lyapunov exponents and Lyapunov dimension of chaotic attractors versus f_0 [eq. (4.4)].

Table 2
Double rotor map: calculation of Lyapunov exponents and Lyapunov dimension of chaotic attractor [$f_0 = 9.0$, other parameters given by eq. (4.4); number of initial conditions $= N_0 = 256$; number of iterations = 10 000]. $d_L = 2 + (L_1 + L_2)/|L_3| = 2.838$.

	j 1	2	3	4
$L_j = (1/N_0)\sum_{i=1}^{N_0} L_j^{(i)}$	1.205	0.256	−1.744	−2.717
$\min_{i=1,N_0} L_j^{(i)}$	1.182	0.228	−1.771	−2.734
$\max_{i=1,N_0} L_j^{(i)}$	1.229	0.284	−1.719	−2.693
$[(1/N_0)\sum_{i=1}^{N_0}(L_j^{(i)} - L_j)^2]^{1/2}$	0.00816	0.0102	0.00910	0.00724

attractor, we also give some details on the numerical calculation of these exponents.

We have now described in sufficient detail the two ingredients necessary to the application of the control method to the double rotor map: chaotic attractors and fixed points. It remains to be checked if the fixed points determined in section 4.2 are embedded in the chaotic attractor. By this we mean that any neighborhood of the fixed point contains an infinite number of points of the chaotic attractor. In order to check this, we consider the intersection of the attractor in its four-dimensional phase space with a three-dimensional hyperplane containing the fixed points $\mathbf{Z}_*$ that we wish to check. Numerically we approximate the hyperplane by a very narrow slab through each fixed point of the form

$$|\hat{\boldsymbol{K}}^{\mathrm{T}}(\mathbf{Z}-\mathbf{Z}_*)|<w\,. \tag{4.9}$$

Actually we took the slabs parallel to the plane (x_1, x_2) which implies that each slab contains the four fixed points with the same rotation number. We then examine a very long orbit and plot only those points satisfying (4.9). The intersection of our 2.8-dimensional attractor with a three dimensional hyperplane is a 1.8-dimensional cross-section. The small scale structure of this 1.8-dimensional intersection is somewhat fuzzed out due to the finite slab thickness. The results, for $f_0 = 9.0$, are given in figs. 10a–10e, which refer to the rotation numbers $N=(0,0)$, $(1,2)$, $(0,1)$, $(1,1)$ and $(2,3)$, respectively. In these figures the relevant fixed points are denoted by a + symbol. The results indicate, with different degrees of certitude, that the first four sets of fixed points are indeed embedded in the attractor while the fifth is not. Note that fig. 10a nicely reveals the symmetry of the map with respect to the point $(\pi, \pi, 0, 0)$. Note also the fractal-like structure in this figure.

We conclude this discussion by mentioning what seems to be an interesting issue: the loss of hyperbolicity due to the existence of fixed points embedded in the attractor that have a number of unstable directions (that is, eigenvalues with magnitude bigger than one) different from the number of unstable directions of the attractor (that is, positive Lyapunov exponents). In fact, from the observation of fig. 7 and table 2, we see that while the chaotic attractor for $f_0 = 9.0$ has two positive Lyapunov exponents some of the unstable fixed points embedded in the attractor have only one unstable eigenvalue.

4.4. Control

We now proceed to apply the method developed in section 2 to control the fixed points of the double rotor map with control parameter f_0. Let us denote by $\bar{\boldsymbol{Z}}_*$ the fixed point to be controlled at the nominal value $\bar{f}_0$ of the parameter. The quantities that were introduced in section 2 now take the following particular form:

$$\mathbf{A}=\begin{pmatrix}\mathbf{I}_2 & \mathbf{M}\\ \mathbf{H}(\bar{\boldsymbol{X}}_*) & \mathbf{L}+\mathbf{H}(\bar{X}_*)\mathbf{M}\end{pmatrix},$$

$$\mathbf{H}(\bar{\boldsymbol{X}}_*)=\frac{\bar{f}_0}{I}\begin{pmatrix}l_1\cos\bar{x}_{1*} & 0\\ 0 & l_2\cos\bar{x}_{2*}\end{pmatrix},$$

$$\boldsymbol{B}^{\mathrm{T}}=\begin{pmatrix}0 & 0 & \frac{l_1}{I}\sin\bar{x}_{1*} & \frac{l_2}{I}\sin\bar{x}_{2*}\end{pmatrix},$$

$$\mathbf{C}=(\boldsymbol{B}\,\vdots\,\mathbf{A}\boldsymbol{B}\,\vdots\,\mathbf{A}^2\boldsymbol{B}\,\vdots\,\mathbf{A}^3\boldsymbol{B}),$$

$$\mathbf{T}=\mathbf{CW},$$

$$\mathbf{W}=\begin{pmatrix}a_3 & a_2 & a_1 & 1\\ a_2 & a_1 & 1 & 0\\ a_1 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\end{pmatrix},$$

$$\boldsymbol{K}^{\mathrm{T}}=(\alpha_4-a_4\quad \alpha_3-a_3\quad \alpha_2-a_2\quad \alpha_1-a_1)\mathbf{T}^{-1}\,.$$

One immediate conclusion that can be drawn from these results is that the controllability matrix $\mathbf{C}$ is identically zero in the case of the fixed points with rotation numbers $N=(0,0)$ for which $\sin\bar{x}_{1*}=\sin\bar{x}_{2*}=0$. Hence these points are uncontrollable, at least when the control parameter is f_0. We will show in the next subsection that this set of fixed points can be controlled if we modify the double rotor map to allow for

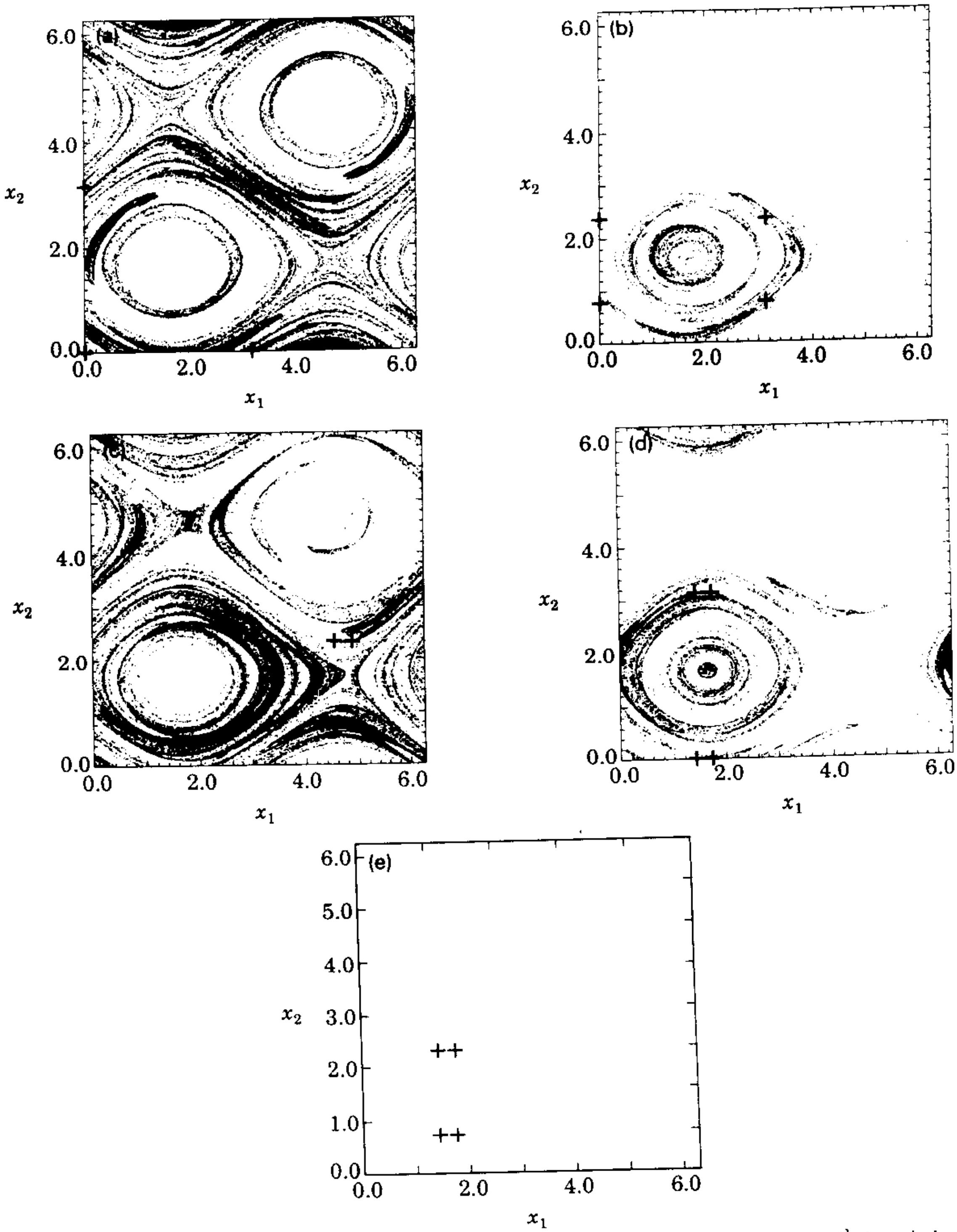


Fig. 10. Double rotor map: sections of chaotic attractor by slab $|\hat{K}^{T}(Z - Z_{*})| < w$, $\hat{K}^{T} = (0, 0, 1, 1)$, $w = 10^{-2}$, through the fixed points (+) with rotation numbers (a) $N = (0, 0)$, (b) $N = (1, 2)$, (c) $N = (0, 1)$, (d) $N = (1, 1)$, (e) $N = (2, 3)$. The map was iterated 10^{8} times [$f_0 = 9.0$, eq. (4.4)].

kicks with variable direction and then take as control parameter the angle the kicks make with the vertical direction in fig. 5.

The method is illustrated in figs. 11a, 11b. The control of the first fixed point was turned on at $i = 0$ with switches to control other fixed points occurring at later times. We plot the x_1 and x_2 coordinates of an orbit as a function of (discrete) time. The parameter perturbations were programmed to control successively four different fixed points of the set with rotation numbers $N = \pm(0, 1)$. The times at which we switched the control from stabilizing one fixed point to stabilizing another are labeled by the arrows in the figure. The figure clearly illustrates the flexibility offered by the method in controlling different periodic motions embedded in the attractor. The figure also shows that the time to achieve control is different from case to case.

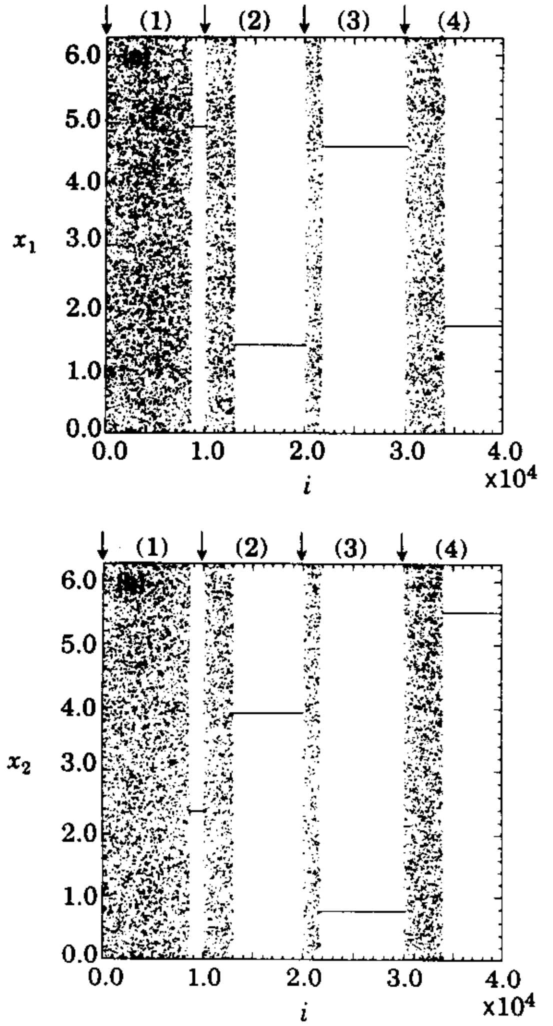


Fig. 11. Double rotor map: successive control of fixed points (1) $[(0, 1); 4]$, (2) $[(0, -1); 1]$, (3) $[(0, 1); 1]$, (4) $[(0, -1); 4]$. The arrows indicate the times of switching. The regulator poles correspond to projection onto the stable manifold [$\delta = 1.0$, $\bar{f}_0 = 9.0$, eq. (4.4)].

We now report the results of several numerical experiments that were carried out with the purpose of understanding the behavior of the time to achieve control.

The first experiment was intended to confirm that the time to achieve control indeed follows an exponential probability distribution as indicated in section 2.4. We proceeded to control the fixed point $[(0, 1); 4]$ by starting at a large number of different points on the attractor and measuring the time each orbit took to reach the fixed point. We then obtained the distribution function of the time to achieve control $\phi(\tau)$ by plotting a histogram of τ using bins of constant size. The results are presented as a semilog plot. in fig. 12 and show excellent agreement with the predicted fit to a straight line.

In our next experiment we looked at the dependence of the average time to achieve control on the size of the parameter perturbations, δ. The results are shown in fig. 13, where we have used logarithmic scales in both axes. The two fixed points $[(0, 1); 4]$ and $[(0, 1); 1]$ were controlled. (The first of these points has two unstable eigenvalues while the second has only one unstable eigenvalue.) We see that for the smaller values of δ the results closely follow straight lines indicating a power law dependence,

$$\langle \tau \rangle \sim \delta^{-\gamma} ,$$

in accord with the theoretical predictions of ref. [6] for two-dimensional maps.

In the experiments described until this point the choice of the regulator poles (eigenvalues of

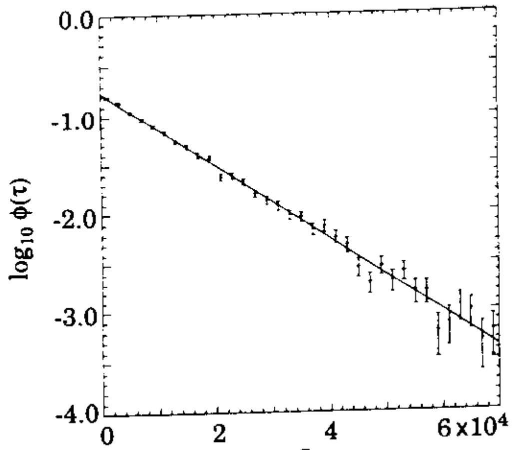


Fig. 12. Double rotor map: histogram of the time to achieve control τ of a sample of 8192 orbits. The fixed point controlled was [(0, 1); 4]. The regulator poles correspond to projection onto the stable manifold. [$\delta = 1.0$, $\bar{f}_0 = 9.0$, eq. (4.4)].

$\mathbf{A} - \boldsymbol{BK}^{\mathrm{T}}$) corresponded to projection onto the stable manifold of the fixed points. That is, the stable eigenvalues of matrix $\mathbf{A}$ were left unchanged, and the unstable eigenvalues were shifted to zero.

In our next set of experiments we looked at how different choices of regulator poles affect the average time to achieve control. We considered the fixed point [(0, 1); 4] with two unstable eigendirections and kept two of the regulator poles equal to the two stable eigenvalues of the fixed point. As regards the other two regulator poles, μ_1 and μ_2, three cases were considered:

(I) $\mu_2 = 0$,
(II) $\mu_2 = \mu_1$,
(III) $\mu_2 = -\mu_1$.

μ_1 was then allowed to vary in the interval $(-1, 1)$. The results of the experiments are shown in fig. 14. In cases (I) and (II) the average time to achieve control essentially increases with μ_1, indicating behavior similar to that found for the Hénon map in fig. 2. In case (III) the average time to achieve control passes through a broad minimum. (Note that the point $\mu_1 = \mu_2 = 0$, which is common to the three cases, corresponds to projection onto the stable manifold.)

4.5. f_0-uncontrollable fixed points

We saw that the set of four fixed points with rotation numbers $N = (0, 0)$ could not be con-

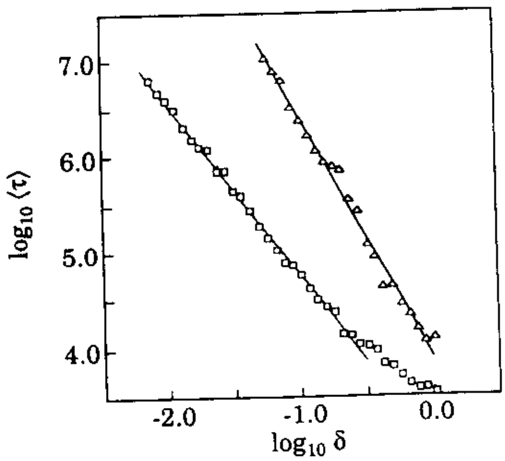


Fig. 13. Double rotor map: $\log_{10}\langle\tau\rangle$ versus $\log_{10}\delta$ for control of the fixed points (□) [(0, 1); 1], (△) [(0, 1); 4]. The regulator poles correspond to projection onto the stable manifold. The straight lines are least square fits to the data [excluding the last nine data points in the case of (□)], [$\bar{f}_0 = 9.0$, eq. (4.4)].

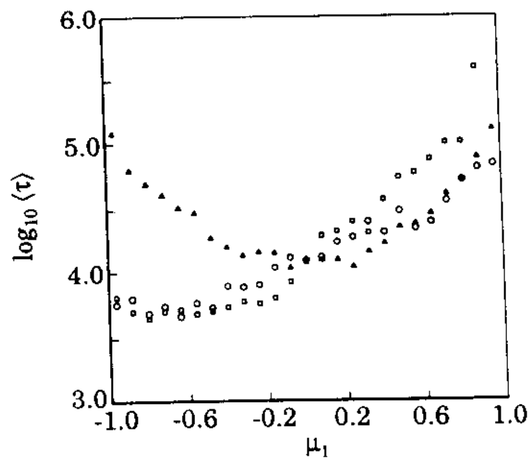


Fig. 14. Double rotor map: $\log_{10}\langle\tau\rangle$ versus μ_1 for (○) $\mu_2 = 0$, (□) $\mu_2 = \mu_1$, (△) $\mu_2 = -\mu_1$. The other two regulator poles were kept equal to the stable eigenvalues of the uncontrolled map. The fixed point controlled was [(0, 1); 4] [$\delta = 1.0$, $\bar{f}_0 = 9.0$, eq. (4.4)].

trolled by changes in the parameter f_0 because the controllability matrix at these points is identically zero.

We show now that these fixed points can be controlled by modifying the double rotor map to allow for kicks with variable direction and then taking the direction of the kicks to be the control parameter, with the nominal value corresponding to the previously fixed direction.

Let us assume that the direction of the kicks makes an angle ψ with the vertical downward direction. Going back to the derivation of the double rotor map in Appendix B, it is easy to verify that the introduction of kicks with variable direction can be taken into account by simply replacing the function G used in the definition of the map and given by eq. (4.2) by the new function

$$\tilde{G}(X) = \begin{pmatrix} c_1 \sin(x_1 - \psi) \\ c_2 \sin(x_2 - \psi) \end{pmatrix}.$$

Taking ψ to be the control parameter with variations around the nominal value $\bar{\psi} = 0$, the application of the method now involves the following quantities:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{M} \\ \tilde{\mathbf{H}}(\bar{X}_*) & \mathbf{L} + \tilde{\mathbf{H}}(\bar{X}_*)\mathbf{M} \end{pmatrix},$$

$$\tilde{\mathbf{H}}(\bar{X}_*) = \frac{f_0}{I} \begin{pmatrix} l_1 \cos \bar{x}_{1*} & 0 \\ 0 & l_2 \cos \bar{x}_{2*} \end{pmatrix},$$

$$\boldsymbol{B}^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & -\frac{f_0 l_1}{I} \cos \bar{x}_{1*} & -\frac{f_0 l_2}{I} \cos \bar{x}_{2*} \end{pmatrix}.$$

The fixed points are now all controllable by small perturbations of the parameter ψ around the nominal value $\bar{\psi} = 0$.

Figs. 15a, 15b illustrate the control of the fixed points $[(0,0);3]$ and $[(0,0);4]$ by kicks of variable direction. The parameter perturbations were programmed to control the first of these points from $i = 0$ to $i = 10^4$ and the second from $i = 10^4$ to $i = 2 \times 10^4$.

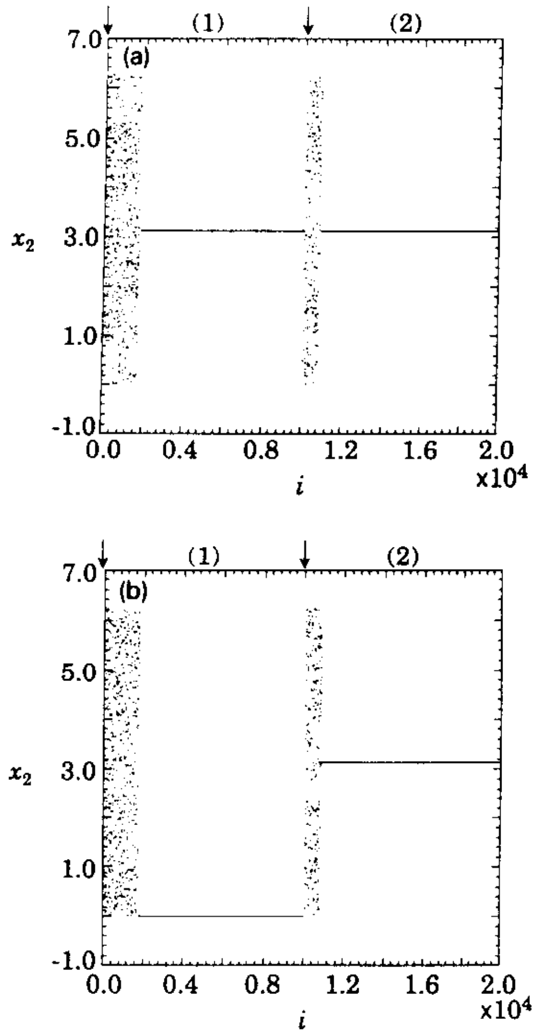


Fig. 15. Modified double rotor map: successive control of fixed points (1) $[(0,0);3]$, (2) $[(0,0);4]$ by kicks of variable direction. The arrows indicate the times of switching. The regulator poles correspond to projection onto the stable manifold $[\delta = 0.05,\ f_0 = 9.0$, eq. (4.4)].

5. Discussion

The transient phase where the orbit wanders chaotically before locking in to a controlled orbit can be greatly shortened by applying the technique discussed by Shinbrot et al. [25]. In the latter paper it was pointed out that orbits can be rapidly brought to a target region on the attractor (in the present case the neighborhood of the periodic orbit which we wish to stabilize) by using small control perturbations when the orbit is far from the neighborhood of the periodic orbit to be stabilized. The idea was that, since chaotic systems are exponentially sensitive to perturbations, careful choice of even small control perturbations can, after some time, have a large effect on the orbit location and can be used to guide it. Thus the time to achieve control can, in principle, be greatly shortened by properly applying small controls when the orbit is far from the neighborhood of the desired periodic orbit.

One issue which we have not addressed is the effect of noise. If the noise remains small, it may not be sufficient to kick the orbit out of the neighborhood of the chosen periodic orbit where the control is activated. In this case, the orbit remains near the desired periodic orbit indefinitely. However, it may be that the random noise is such that it may occasionally kick the orbit far enough away from the periodic orbit that the orbit falls outside the small controlled phase space region. In this case, after the orbit is kicked out of the controlled phase space region, it wanders chaotically over the attractor until it falls in the controlled region again. Thus there are epochs where the orbit is kept near the desired orbit interspersed with epochs wherein the orbit wanders chaotically far from the desired orbit. If the latter are, on average, relatively much shorter than the former, then one might still regard the control as being effective. See ref. [6] for numerical experiments on this effect using the Hénon map. We also remark that the procedure discussed in the previous paragraph [25] can be used to greatly reduce the duration of the noise induced epochs where the orbit bursts out of the controlled phase space region.

In this paper we have considered the case where there is only a single control parameter available for adjustment. While generically a single parameter is sufficient for stabilization of a desired periodic orbit, there may be some advantage to utilizing several control variables. Therefore, the single control parameter p becomes a vector (e.g., ref. [26] discusses the case where the number of control parameters is equal to the number of unstable eigenvalues). In particular, the added freedom in having several control parameters might allow better means of choosing the control so as to minimize the time to achieve control, as well as the effects of noise.

Finally we wish to point out that full knowledge of the system dynamics is not necessary in order to apply our technique (see also ref. [6]). In particular, we only require the location of the desired periodic orbit, the linearized dynamics about the periodic orbit, and the dependence of the location of the periodic orbit on small variation of the control parameter. Recently, delay coordinate embedding [19, 27] has been utilized in several experimental studies (refs. [8, 28–31]) to extract such information purely from observations of experimental chaotic orbits on the attractor without any a priori knowledge of the system of equations governing the dynamics, and such information has been utilized to control periodic orbits [9]. Hence, application of our method is not limited to cases where a complete knowledge of the system is available.

In conclusion, we have demonstrated that chaotic dynamics can often be converted, by using only a small feedback control, to motion on a desired periodic orbit. Furthermore, by switching the small control, one can switch the time asymptotic behavior from one periodic orbit to another. In some situations, where the flexibility offered by the ability to do such switching is desirable, it may be advantageous to design the system so that it is chaotic. In other situations, where one is presented with a chaotic system, the method may allow one to eliminate the chaos and achieve greatly improved behavior at relatively low cost.

Acknowledgements

This work was supported by the US Department of Energy (Scientific Computing Staff, Office of Energy Research), the Portuguese Junta Nacional de Investigação Científica e Tecnológica, and the National Science Foundation (Engineering Research Center Program). The computation was done at the National Energy Research Supercomputer Center.

Appendix A. Time to achieve control in the case of two-dimensional maps

We assume that control is achieved if the orbit remains in the slab (2.8) for two consecutive iterations of the map. The two conditions

$$|\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{Z}_*(\bar{p}))|<\delta\,,\quad |\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{Z}'-\boldsymbol{Z}_*(\bar{p}))|<\delta\,, \tag{A.1}$$

define a control "parallelepiped" P_c, where $\boldsymbol{Z}'=\boldsymbol{F}(\boldsymbol{Z},p)$. For small δ, an initial condition will bounce around on the set comprising the uncontrolled chaotic attractor for a long time before it falls in the control parallelepiped P_c. At any given iterate the probability of falling in P_c is approximately the natural measure (see, for example, [17, 18, 23]) of the uncontrolled chaotic attractor contained in P_c. If we follow many orbits this probability $\mu(P_c)$ also gives the rate at which these orbits fall into P_c. Thus $\mu(P_c)$ is the inverse of the average time for a typical orbit to first fall in P_c,

$$\langle\tau\rangle^{-1}=\mu(P_c)\,. \tag{A.2}$$

An estimate for $\mu(P_c)$ can be given in the two-dimensional case [23]:

$$\mu(P_c)\sim\int_{P_c}\rho|v_s|^{d_s-1}|v_u|^{d_u-1}\,\mathrm{d}v_s\,\mathrm{d}v_u\,, \tag{A.3}$$

where v_s and v_u denote linear coordinates in the stable and unstable directions. In here d_u and d_s are the pointwise dimensions [1] for the uncontrolled chaotic attractor at the fixed point in the unstable and the stable directions, respectively; ρ is a normalizing constant. Assuming that the attractor is smooth in the unstable direction we have $d_u=1$, while d_s is given in terms of the eigenvalues at the fixed point [1, 17, 18] by

$$d_s=\frac{\log_e|\lambda_u|}{\log_e(1/|\lambda_s|)}\,.$$

In order to determine the control parallelepiped, we need to obtain $\boldsymbol{Z}'$ in the neighborhood of the fixed point $\boldsymbol{Z}_*(\bar{p})$ with a better approximation than that provided by the linear map (2.3). We therefore take

$$\boldsymbol{V}'=\mathbf{A}\boldsymbol{V}+\boldsymbol{B}(p-\bar{p})+\tfrac{1}{2}\boldsymbol{Q}(\boldsymbol{V},\boldsymbol{V})+\tfrac{1}{2}\boldsymbol{D}(p-\bar{p})^2\,, \tag{A.4}$$

where

$$\boldsymbol{V}=\boldsymbol{Z}-\boldsymbol{Z}_*(\bar{p})\,,\quad \boldsymbol{V}'=\boldsymbol{Z}'-\boldsymbol{Z}_*(\bar{p})\,,$$

$\mathbf{A}$, $\boldsymbol{B}$ were defined by (2.4), (2.5) and $\boldsymbol{Q}$, $\boldsymbol{D}$ are two vectors with components q_k, d_k $(k=1,2)$ defined by

$$q_k=\sum_{i,j=1}^{2}\left.\frac{\partial^2 f_k}{\partial x_i\,\partial x_j}\right|_{(\boldsymbol{Z}_*(\bar{p}),\bar{p})}v_i v_j\,,$$

$$d_k=\left.\frac{\partial^2 f_k}{\partial p^2}\right|_{(\boldsymbol{Z}_*(\bar{p}),\bar{p})}\,;$$

in here x_k, f_k, and v_k $(k=1,2)$ denote the components of the vectors $\boldsymbol{X}$, $\boldsymbol{F}$ and $\boldsymbol{V}$. Using (2.6) to eliminate $p-\bar{p}$ from eq. (A.4), we obtain

$$\boldsymbol{V}'=\mathbf{A}\boldsymbol{V}+\tfrac{1}{2}\boldsymbol{Q}(\boldsymbol{V},\boldsymbol{V})-\boldsymbol{B}(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V})+\tfrac{1}{2}\boldsymbol{D}(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V})^2\,. \tag{A.5}$$

The control "parallelogram" P_c will therefore be defined by the two equations

$$|\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V}|<\delta\,,\quad |\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V}'|<\delta\,. \tag{A.6}$$

In order to compare with the numerical experimental results described in section 3 we have carried out the calculation of (A.3) in the case of the Hénon map. Writing this map in the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} a - x_1^2 + b x_2 \\ x_1 \end{pmatrix},$$

and taking a to be the control parameter while b is kept fixed, we obtain

$$\mathbf{A} = \begin{pmatrix} -2\bar{x}_* & b \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{Q}(\boldsymbol{V}, \boldsymbol{V}) = \begin{pmatrix} -2v_1^2 \\ 0 \end{pmatrix},$$

$$\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V} = k_1 v_1 + k_2 v_2,$$

$$\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V}' = -k_1 v_1^2 + (k_2 - k_1^2 - 2\bar{x}_* k_1) v_1 + k_1(b - k_2) v_2 .$$

Also we note that for the Hénon map the variables (v_1, v_2) and (v_s, v_u) are related by

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \lambda_s \gamma_s & \lambda_u \gamma_u \\ \gamma_s & \gamma_u \end{pmatrix} \begin{pmatrix} v_s \\ v_u \end{pmatrix}, \qquad \text{(A.7)}$$

where

$$\gamma_s = (1 + \lambda_s^2)^{-1/2}, \quad \gamma_u = (1 + \lambda_u^2)^{-1/2} .$$

Letting

$$\xi = v_1, \quad \eta = v_1 + t v_2, \quad t = k_2/k_1,$$

and using (A.7) to change the variables of integration, eq. (A.3) can be written in the form

$$\mu(P_c) \sim \rho_0 \frac{|r|}{|1 - r|^{d_s}} \int_{P_c} |\xi - r\eta|^{d_s - 1}\, \mathrm{d}\xi\, \mathrm{d}\eta, \qquad \text{(A.8)}$$

where

$$\frac{1}{r} = 1 + \frac{t}{\lambda_u}$$

and

$$\frac{1}{\rho_0} = \frac{1}{\rho} \gamma_u |\lambda_u| [\gamma_s(\lambda_s - \lambda_u)]^{d_s} .$$

The integration in the variable in the direction of the straight lines $\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V} = \pm\delta$ can be done exactly. On the contrary, except in the case $k_1 = 0$, the integration in the other variable does not seem to be possible in closed form. We have therefore resorted to numerical integration to obtain the results presented in fig. 16 (see below).

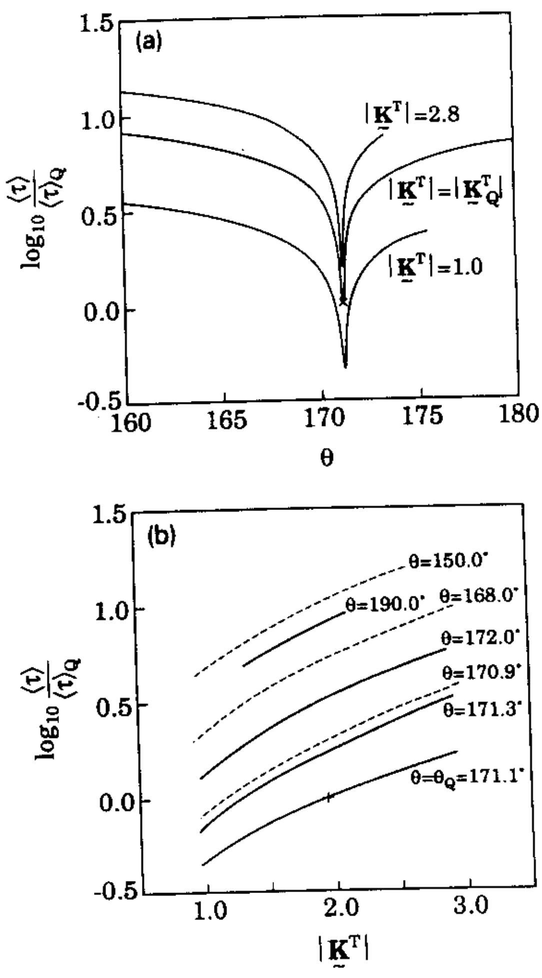


Fig. 16. Hénon map: (theoretical) curves of $\log_{10}(\langle\tau\rangle/\langle\tau\rangle_Q)$ versus $\theta = \arg(\boldsymbol{K}^{\mathrm{T}})$ with $|\boldsymbol{K}^{\mathrm{T}}|$ fixed (a) and of $\log_{10}(\langle\tau\rangle/\langle\tau\rangle_Q)$ versus $|\boldsymbol{K}^{\mathrm{T}}|$ with θ fixed (b). The × (+) denotes the reference value.

58

An accurate analytical approximation can be obtained in the case $t = -\lambda_s$ (slab parallel to the stable manifold) in the limit $\delta \to 0$:

$$\mu(P_c) \sim \rho_1 g(\mu_1)\, \delta^\gamma [1 + 0(\delta)]\,, \qquad (A.9)$$

where g is the function defined by

$$g(\mu_1) = \frac{2}{\gamma}\, \frac{(1+\mu_1)^\gamma - (1-\mu_1)^\gamma}{\mu_1(|\lambda_u| + \mu_1)^\gamma}\,,$$

$$\mu_{1m} < \mu_1 < 0 \quad 0 < \mu_1 < \mu_{1M}$$

$$g(0) = 4|\lambda_u|^{-\gamma}\,.$$

and

$$\gamma = \tfrac{1}{2} d_s + 1\,,$$

$$\rho_1 = \frac{\rho_0}{d_s}\, \frac{|r|}{|1-r|^{d_s}}\,, \quad \frac{1}{r} = 1 - \frac{\lambda_s}{\lambda_u}\,,$$

$$\mu_{1m} = -1 + \frac{r^2}{|\lambda_u| - 1}\, \delta + \mathcal{O}(\delta^1)\,,$$

$$\mu_{1M} = 1 - \frac{r^2}{|\lambda_u| + 1}\, \delta + \mathcal{O}(\delta^2)\,.$$

The dependence of $\mu(P_c)$ on δ given by (A.9) is precisely that predicted in ref. [6]. The dependence of $\mu(P_c)$ on μ_1 shows very good agreement with the experimental results – see fig. 2. Note that in plotting the theoretical curve we used as normalizing constant ρ_1 that obtained by least square fitting the theoretical curve to the experimental points.

We have used eq. (A.8) to study the dependence of $\langle\tau\rangle = 1/\mu(P_c)$ on the gain vector $\boldsymbol{K}^\mathrm{T}$. In fig. 16a we have plotted curves of $\langle\tau\rangle$ versus $\theta = \arg(\boldsymbol{K}^\mathrm{T})$ with $|\boldsymbol{K}^\mathrm{T}|$ kept fixed and in fig. 16b curves of $\langle\tau\rangle$ versus $|\boldsymbol{K}^\mathrm{T}|$ with θ kept fixed. $\langle\tau\rangle$ was normalized to its value at the point Q (see section 3 and fig. 1). The results show that $\langle\tau\rangle$ exhibits a strong minimum at $\theta = \theta_Q$ for all values of $|\boldsymbol{K}^\mathrm{T}|$ and increases slowly with $|\boldsymbol{K}^\mathrm{T}|$ for all values of θ, in agreement with the experimental results of figs. 3 and 4.

Appendix B. Numerical method for calculating $\langle\tau\rangle$

In this appendix we describe the procedure used in sections 3 and 4 to numerically obtain the average time to achieve control, $\langle\tau\rangle$.

From (2.10) we obtain the fraction of chaotic transients with length smaller than some value τ_{max},

$$\rho_{\tau_{max}} = \int_0^{\tau_{max}} \phi(\tau)\, d\tau = 1 - \exp\left(-\frac{\tau_{max}}{\langle\tau\rangle}\right),$$

and the average length of the chaotic transients with length smaller than τ_{max},

$$\langle\tau\rangle_{\tau_{max}} = \int_0^{\tau_{max}} \tau\phi(\tau)\, d\tau$$

$$= \langle\tau\rangle\left[1 - \left(1 + \frac{\tau_{max}}{\langle\tau\rangle}\right)\exp\left(-\frac{\tau_{max}}{\langle\tau\rangle}\right)\right].$$

Combining these two equations we obtain

$$\langle\tau\rangle = \frac{\langle\tau\rangle_{\tau_{max}}}{1 - (1 - \rho_{\tau_{max}})[1 - \log_e(1 - \rho_{\tau_{max}})]}\,, \qquad (B.1)$$

which is the required formula. Note that $\rho_\infty = 1$, $\langle\tau\rangle_\infty = \langle\tau\rangle$.

The numerical procedure to calculate the average time to achieve control is as follows. Take a large number N_0 of randomly chosen initial conditions and iterate each of them with the uncontrolled map [i.e., with $\boldsymbol{Z} \mapsto \boldsymbol{F}(\boldsymbol{Z}, \bar{p})$] a sufficient number of times until they are all distributed over the attractor according to its natural measure. Then switch on the control as specified by (2.9) and determine how many further iterates are necessary for $N_f \leq N_0$ orbits to fall within a circle of small radius centered at the fixed point. Letting τ_{max} be this number of iterates and $\{\tau_j\}$ with $j = 1, \ldots, N_f$ be the times required for the N_f orbits to fall within the small circle, we have

$$\rho_{\tau_{\max}} = \frac{N_f}{N_0}\,, \quad \langle\tau\rangle_{\tau_{\max}} = \frac{1}{N_f}\sum_{j=1}^{N_f} \tau_j\,.$$

Finally we use eq. (B.1) to obtain $\langle\tau\rangle$. In our numerical experiments described in sections 3 and 4, we took $N_0 = 192$, $N_f = 121$, values that led to a good compromise between accuracy and computation time.

Appendix C. Derivation of the double rotor map

The equations of motion of the kicked double rotor are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot\theta_j}\right) - \frac{\partial L}{\partial \theta_j} = -\frac{\partial F}{\partial \dot\theta_j}\,, \quad j = 1, 2\,, \tag{C.1}$$

where the Lagrangian function L is the difference between the kinetic energy,

$$K(\dot\theta_1, \dot\theta_2) = \tfrac{1}{2} I_1 \dot\theta_1^2 + \tfrac{1}{2} I_2 \dot\theta_2^2\,,$$

and the potential energy,

$$V(\theta_1, \theta_2, t) = (l_1 \cos\theta_1 + l_2 \cos\theta_2) f(t)\,,$$

i.e., $L = T - V$, and Rayleigh's dissipation function F is

$$F(\dot\theta_1, \dot\theta_2) = \tfrac{1}{2}\nu_1 I_1 \dot\theta_1^2 + \tfrac{1}{2}\nu_2 I_2 (\dot\theta_2 - \dot\theta_1)^2\,.$$

The sequence of forcing kicks is given by the semi-infinite comb of delta functions of period T and strength f_0,

$$f(t) = f_0 \sum_{k=1}^{\infty} \delta(t - kT)\,. \tag{C.2}$$

In here I_1 and I_2 are the moments of inertia,

$$I_1 = (m_1 + m_2) l_1^2\,, \quad I_2 = m_2 l_2^2\,,$$

and ν_1, ν_2 are the coefficients of friction.

Elimination of L and F from (C.1) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\dot\theta_1\\ \dot\theta_2\end{pmatrix} = \begin{pmatrix} -(\nu_1 + \nu_2 I_2/I_1) & \nu_2 I_2/I_1 \\ \nu_2 & -\nu_2 \end{pmatrix}\begin{pmatrix}\dot\theta_1\\ \dot\theta_2\end{pmatrix} + f(t)\begin{pmatrix}(l_1/I_1)\sin\theta_1\\ (l_2/I_2)\sin\theta_2\end{pmatrix}. \tag{C.3}$$

We now proceed to integrate eq. (C.3). For simplicity we take $I_1 = I_2 \equiv I$.

Since the effect of the kicks is instantaneous (i.e., $f(t) = 0$, for $t \neq kT$, $k = 1, 2, \ldots$) eqs. (C.3) are linear between successive kicks. In particular, for $0 < t < T$, eqs. (C.3) reduce to

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\dot\theta_1\\ \dot\theta_2\end{pmatrix} = \mathbf{A}_\nu \begin{pmatrix}\dot\theta_1\\ \dot\theta_2\end{pmatrix},$$

$$\mathbf{A}_\nu = \begin{pmatrix} -(\nu_1 + \nu_2) & \nu_2 \\ \nu_2 & -\nu_2 \end{pmatrix}. \tag{C.4}$$

This system can be easily solved by the usual methods for linear differential equations with constant coefficients. Denoting by $\dot\theta_1(0)$, $\dot\theta_2(0)$ the initial angular velocities this solution is

$$\begin{pmatrix}\dot\theta_1(t)\\ \dot\theta_2(t)\end{pmatrix} = \mathbf{L}(t)\begin{pmatrix}\dot\theta_1(0)\\ \dot\theta_2(0)\end{pmatrix}, \tag{C.5}$$

where

$$\mathbf{L}(t) = \sum_{j=1}^{2} \mathbf{W}_j \,\mathrm{e}^{\lambda_j t}\,.$$

λ_1, λ_2 are the eigenvalues of matrix $\mathbf{A}_\nu$,

$$\left.\begin{matrix}\lambda_1\\ \lambda_2\end{matrix}\right\} = -\tfrac{1}{2}(\nu_1 + 2\nu_2 \pm \Delta)\,,$$

$$\Delta = (\nu_1^2 + 4\nu_2^2)^{1/2}\,,$$

and $\mathbf{W}_1$, $\mathbf{W}_2$ are the constant matrices

$$\mathbf{W}_1 = \begin{pmatrix} a & b \\ b & d\end{pmatrix}, \quad \mathbf{W}_2 = \begin{pmatrix} d & -b \\ -b & a\end{pmatrix},$$

where

$$\left.\begin{matrix}a\\ d\end{matrix}\right\} = \tfrac{1}{2}\left(1 \pm \frac{\nu_1}{\Delta}\right), \quad b = -\frac{\nu_2}{\Delta}\,.$$

The position of the rods is obtained by integration of eq. (C.5). Denoting by $\theta_1(0)$, $\theta_2(0)$ the initial positions one obtains

$$\begin{pmatrix}\theta_1(t)\\ \theta_2(t)\end{pmatrix} = \mathbf{M}(t)\begin{pmatrix}\dot\theta_1(0)\\ \dot\theta_2(0)\end{pmatrix} + \begin{pmatrix}\theta_1(0)\\ \theta_2(0)\end{pmatrix}, \tag{C.6}$$

where

$$\mathbf{M}(t) = \int_0^t \mathbf{L}(\xi)\,d\xi = \sum_{j=1}^{2} \mathbf{W}_j \frac{e^{\lambda_j t} - 1}{\lambda_j}.$$

Eqs. (C.5), (C.6) completely describe the motion of the rotor for $0 < t < T$ (before the first kick).

At $t = T$ the kick instantaneously changes the angular velocity of each rod but not its position; that is, the angular velocity of each rod is discontinuous at $t = T$, while the position is continuous. Denoting by $\theta_j(T^\pm)$, $\dot\theta_j(T^\pm)$, $j = 1, 2$ the values of $\theta_j(t)$, $\dot\theta_j(t)$ just before and just after the kick at $t = T$, we therefore have

$$\theta_j(T^-) = \theta_j(T^+) = \theta_j(T), \tag{C.7}$$

$$\dot\theta_j(T^+) - \dot\theta_j(T^-) = \frac{f_0}{I}\, l_j \sin\theta_j(T), \tag{C.8}$$

for $j = 1, 2$.

The solution of eqs. (C.3) for $T < t < 2T$ is identical to the solution of the linear system eq. (C.4) for $0 < t < T$ except that the initial conditions $\theta_j(0)$, $\dot\theta_j(0)$, $j = 1, 2$ are replaced by $\theta_j(T)$, $\dot\theta_j(T^+)$, $j = 1, 2$.

The solution of eqs. (C.3) is a composition of the solution of eqs. (C.4) with the effect of the kicks at $t = T, 2T, \ldots$. To study the dynamics of the rotor it is natural to consider only the state of the system immediately after each kick. Thus we obtain from (C.5)–(C.8) the *double rotor map*,

$$\begin{pmatrix}\theta_1^{(k+1)}\\ \theta_2^{(k+1)}\end{pmatrix} = \mathbf{M}(T)\begin{pmatrix}\dot\theta_1^{(k)}\\ \dot\theta_2^{(k)}\end{pmatrix} + \begin{pmatrix}\theta_1^{(k)}\\ \theta_2^{(k)}\end{pmatrix}, \tag{C.9a}$$

$$\begin{pmatrix}\dot\theta_1^{(k+1)}\\ \dot\theta_2^{(k+1)}\end{pmatrix} = \mathbf{L}(T)\begin{pmatrix}\dot\theta_1^{(k)}\\ \dot\theta_2^{(k)}\end{pmatrix} + \frac{f_0}{I}\begin{pmatrix}l_1 \sin\theta_1^{(k+1)}\\ l_2 \sin\theta_2^{(k+1)}\end{pmatrix}, \tag{C.9b}$$

where

$$\theta_j^{(k)} = \theta_j(kT), \quad j = 1, 2$$

are the positions of the rods at the instant of the kth kick, and

$$\dot\theta_j^{(k)} = \dot\theta_j(kT^+), \quad j = 1, 2$$

are the angular velocities of the rods immediately after the kth kick.

Appendix D. Stability of fixed points for the double rotor map

The coefficients of the characteristic equation (4.8) depend on the fixed point and on the forcing f_0 only through the two non-zero elements of the matrix $\mathbf{H}$, which we are going to denote by h_{11} and h_{22}. The discussion of the stability of the fixed points is conveniently carried out in the plane (h_{11}, h_{22}) by considering the intersections between the lines of marginal stability of the characteristic equation (where one of the roots has modulus unity) and the "orbits" described by the paths followed by the fixed points as the forcing f_0 is varied.

The "orbits" of the fixed points can be obtained by first eliminating x_{j*} ($j = 1, 2$) between the two equations

$$f_0 \sin x_{j*} = f_{0j}, \quad h_{jj} = \frac{f_0}{I}\, l_j \cos x_{j*},$$

with the result

$$h_{jj} = \pm \frac{l_j}{I}\,(f_0^2 - f_{0j}^2)^{1/2}, \quad j = 1, 2,$$

and then eliminating f_0 between these two equations, with the result

$$\left(\frac{h_{22}}{l_2}\right)^2 - \left(\frac{h_{11}}{l_1}\right)^2 = \frac{1}{I^2}\,(f_{01}^2 - f_{02}^2),$$

which is the equation of the hyperbola described by each fixed point in the plane (h_{11}, h_{22}) when f_0 is varied. It should be pointed out that symmetric fixed points with respect to $(x_{1*}, x_{2*}, y_{1*}, y_{2*}) \mapsto (2\pi - x_{1*},\ 2\pi - x_{2*},\ -y_{1*},\ -y_{2*})$ describe the same "orbit." The lines of marginal stability are defined by the equation (see eq. (4.8b))

$$P(e^{i\alpha}) = 0\,, \tag{D.1}$$

where α can take values in the interval $[0, 2\pi)$. When $\alpha = 0$ or $\alpha = \pi$ this equation simplifies considerably. We obtain:

(i) $P(1) = 0 = |\mathbf{HM}| = |\mathbf{H}||\mathbf{M}|$; as $|\mathbf{M}| \neq 0$ this implies

$$h_{11}h_{22} = 0\,.$$

(ii) $P(-1) = 0 = |2(\mathbf{I} + \mathbf{L}) + \mathbf{HM}| = |\mathbf{H} + \mathbf{R}||\mathbf{M}|$, where

$$\mathbf{R} = 2(\mathbf{A}_\nu + 2\mathbf{M}^{-1})\,;$$

writing $\mathbf{R} = \{r_{ij}\}_{i,j=1,2}$ this leads to

$$h_{11}h_{22} + r_{22}h_{11} + r_{11}h_{22} + (r_{11}r_{22} - r_{12}r_{21}) = 0\,.$$

When $\alpha \neq 0, \pi$ it can be shown that the eq. (D.1) has no solutions in the (real) plane (h_{11}, h_{22}); that is, there are no lines of marginal stability with $\alpha \neq 0, \pi$.

In fig. 17 we have plotted in the plane (h_{11}, h_{22}) the lines of marginal stability $P(1) = P(-1) = 0$ (for the parameter values given by eq. (4.4)). The bounded region between these lines, which is the shaded region in the figure, is the only region of the plane where all the roots of the characteristic equation have modulus smaller than unity. We have also plotted the "orbits" of the first five sets of fixed points, the arrows indicating the direction the forcing increases; the critical values of (h_{11}, h_{22}) at $f_0 = f_{0c}$, which occur on the lines $P(1) = 0$, are given in table 1. We see that of all these orbits only two cross the shaded region: one corresponds to the fixed

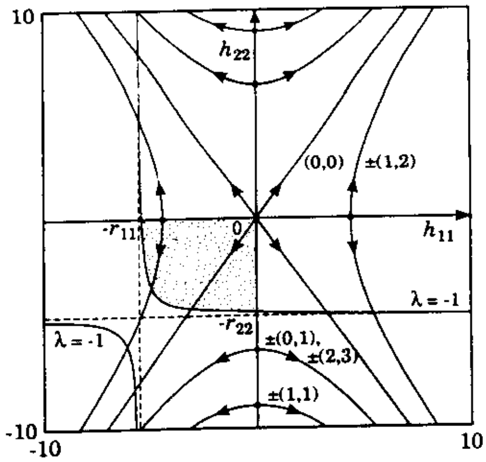


Fig. 17. Double rotor map: stability diagram of the fixed points with rotation numbers (n_1, n_2) [$f_0 = 9.0$, eq. (4.4)].

point $[(0,0); 4]$; the other to the fixed points $[(1,2); 4]$ and $[(-1,-2); 1]$. These fixed points are therefore stable while their "orbits" remain in the shaded region and become unstable when the "orbits" cross the line $P(-1) = 0$; that is, they are stable in a finite interval of values of f_0, $f_{0c} < f_0 < f_{0u}$. The other fixed points are unstable for all values of f_0. The values of f_{0u} are also given in table 2.

Fig. 17 only applies for the particular values of the parameters given by eqs. (4.4). For other values the relative positions of the lines of marginal stability and "orbits" of fixed points are different, and fixed points with other rotation numbers may be stable. In general, we can make the following statements regarding the stability of the fixed points: those with rotation numbers $(0,0)$ are stable over an interval $0 < f_0 < f_{0u}^{(0,0)}$; all the others are either stable over an interval $f_{0c}^{(n_1,n_2)} < f_0 < f_{0u}^{(n_1,n_2)}$ or are always unstable.

References

[1] C. Grebogi, E. Ott and J.A. Yorke, Phys. Rev. A 37 (1988) 1711.

[2] D. Auerbach, P. Cvitanovich, J.P. Eckmann, G. Gunaratne and I. Procaccia, Phys. Rev. Lett. 58 (1987) 2387.

[3] T. Morita, H. Hata, H. Mori, T. Horita and K. Tomita, Prog. Theor. Phys. 78 (1987) 511.
[4] A. Katok, Publ. Math. IHES 51 (1980) 137.
[5] R. Bowen, Trans. Am. Math. Soc. 154 (1971) 377.
[6] E. Ott, C. Grebogi and J.A. Yorke, Phys. Rev. Lett. 64 (1990) 1196.
[7] F.J. Romeiras, E. Ott, C. Grebogi and W.P. Dayawansa, Proc. 1991 Am. Control Conf. (American Automatic Control Council, IEEE Service Center, Piscataway, NJ, 1991), pp. 1112–1119.
[8] J. Singer, Y.-Z. Wang and H.H. Bau, Phys. Rev. Lett. 66 (1991) 1123.
[9] W.L. Ditto, S.N. Rauseo and M.L. Spano, Phys. Rev. Lett. 65 (1990) 3211.
[10] T.B. Fowler, IEEE Trans. on Automatic Control 34 (1989) 201.
[11] A.M. Block and J.E. Marsden, Theor. Comput. Fluid Dyn. 1 (1989) 179;
A. Hübler and E. Lüscher, Naturwissenshaften 76 (1989) 67;
E.A. Jackson, Phys. Lett. A 151 (1990) 478;
B. Huberman and E. Lumer, IEEE Trans. Circuits Syst. 37 (1990) 547;
S. Sinha, R. Ramaswamy and J.S. Rao, Physica D 43 (1990) 118;
B. Peng, V. Petrov and K. Showalter, J. Phys. Chem. 95 (1991) 4957;
T. Tél, Controlling transient chaos, preprint;
A. Azevedo and S.M. Rezende, Phys. Rev. Lett. 66 (1991) 1342;
Y. Braiman and J. Goldhirsch, Phys. Rev. Lett. 66 (1991) 2545.
[12] K. Ogata, Control engineering, Second Ed. (Prentice-Hall, Englewood Cliffs, NJ, 1990), pp. 782–784.
[13] M. Hénon, Commun. Math. Phys. 50 (1976) 69.
[14] C. Grebogi, E. Kostelich, E. Ott and J.A. Yorke, Physica D 25 (1987) 347; Phys. Lett. A 118 (1986) 448; A 120 (1987) 497 (E).
[15] C. Grebogi, E. Ott and J.A. Yorke, Chaotic attractors in crisis, Phys. Rev. Lett. 48 (1982) 1507.
[16] C. Grebogi, E. Ott and J.A. Yorke, Physica D 7 (1983) 181.
[17] C. Grebogi, E. Ott and J.A. Yorke, Phys. Rev. Lett. 57 (1986) 1284.
[18] C. Grebogi, E. Ott, F.J. Romeiras and J.A. Yorke, Phys. Rev. A 36 (1987) 5365.
[19] G. Nitsche and U. Dressler, Controlling chaotic dynamical systems using time delay coordinates, preprint.
[20] G.M. Zaslavsky, Phys. Lett. A 69 (1978) 145.
[21] G. Benettin, L. Galgani, A. Giorgilli and J. Strelcyn, Meccanica 15 (1980) 9.
[22] G. Benettin, L. Galgani, A. Giorgilli and J. Strelcyn, Meccanica 15 (1980) 21.
[23] J.D. Farmer, E. Ott and J.A. Yorke, Physica D 7 (1983) 153.
[24] J. Kaplan and J.A. Yorke, Chaotic behavior of multidimensional difference equations, in: Functional differential equations and the approximation of fixed points, Lecture Notes in Mathematics, eds. H.O. Peitgen and H.O. Walther, Vol. 730 (Springer, Berlin, 1978), p. 228–237.
[25] T. Shinbrot, E. Ott, C. Grebogi and J.A. Yorke, Phys. Rev. Lett. 65 (1990) 3215.
[26] E. Ott, C. Grebogi and J.A. Yorke, Controlling chaotic dynamical systems, in CHAOS/XAOC, Soviet-American perspective on nonlinear science, ed. D. Campbell (American Institute of Physics, New York, 1990), pp. 153–172.
[27] F. Takens, Detecting strange attractors in turbulence, in: Dynamical systems and turbulence, eds. D.A. Rand and L.-S. Young, Lecture Notes in Mathematics, Vol. 898 (Springer, New York, 1980), pp. 366–381.
[28] J.C. Sommerer, W.L. Ditto, C. Grebogi, E. Ott and M. Spano, Phys. Lett. A 153 (1991) 105.
[29] D.P. Lathrop and E.J. Kostelich, Phys. Rev. A 40 (1989) 4028.
[30] G.H. Gunaratne, P.S. Linsay and M.J. Vinson, Phys. Rev. Lett. 63 (1989) 1.
[31] J.C. Sommerer, W.L. Ditto, C. Grebogi, E. Ott and M.L. Spano, Phys. Rev. Lett. 66 (1991) 1947.

Plateau Onset for Correlation Dimension: When Does it Occur?

Mingzhou Ding,[1] Celso Grebogi,[2],[3] Edward Ott,[2],[4] Tim Sauer,[5] and James A. Yorke[3]
[1]*Center for Complex Systems and Department of Mathematics, Florida Atlantic University, Boca Raton, Florida 33431*
[2]*Laboratory for Plasma Research, University of Maryland, College Park, Maryland 20742*
[3]*Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742*
[4]*Departments of Physics and of Electrical Engineering, University of Maryland, College Park, Maryland 20742*
[5]*Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030*
(Received 22 February 1993)

Chaotic experimental systems are often investigated using delay coordinates. Estimated values of the correlation dimension in delay coordinate space typically increase with the number of delays and eventually reach a plateau (on which the dimension estimate is relatively constant) whose value is commonly taken as an estimate of the correlation dimension D_2 of the underlying chaotic attractor. We report a rigorous result which implies that, for long enough data sets, the plateau begins when the number of delay coordinates first exceeds D_2. Numerical experiments are presented. We also discuss how lack of sufficient data can produce results that seem to be inconsistent with the theoretical prediction.

PACS numbers: 05.45.+b

The estimation of the correlation dimension [1] of a presumed chaotic time series has been widely used by scientists to assess the nature of a variety of experimental as well as model systems, ranging from simple circuits to chemical reactions to the human brain. It is also known that many factors, such as noise and a lack of data, can hinder the successful application of the dimension extraction algorithm. In this paper, we address two issues related to the understanding of these difficulties, namely, what happens in an ideal situation (i.e., long data string with low noise) and what could be expected when the data set is small. In particular, we focus on the character of the dependence of the estimated correlation dimension on the dimension of the delay coordinate reconstruction space.

Consider an n-dimensional dynamical system that exhibits a chaotic attractor. A correlation integral $C(\epsilon)$ [1] is defined to be the probability that a pair of points chosen randomly on the attractor with respect to the natural measure ρ is separated by a distance less than ϵ on the attractor. The correlation dimension D_2 [1] of the attractor is then defined as $D_2=\lim_{\epsilon\to 0}\log C(\epsilon)/\log\epsilon$. Assume that we measure and record a trajectory of finite duration L on the attractor at N equally spaced discrete times, $\{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i\in\mathbf{R}^n$. The correlation integral $C(\epsilon)$ is then approximated by

$$C(N,\epsilon)=\frac{2}{N(N-1)}\sum_{j=1}^{N}\sum_{i=j+1}^{N}\Theta(\epsilon-|\mathbf{x}_i-\mathbf{x}_j|)\,, \quad (1)$$

where $\Theta(x)=1$ for $x>0$ and $\Theta(x)=0$ for $x\le 0$. In the limit $L,N\to\infty$, $C(N,\epsilon)\to C(\epsilon)$.

The dynamical information of a chaotic experimental system is often contained in a time series, $\{y_i=y(t_i)\}_{i=1}^N$, obtained by measuring a single scalar function $y=h(\mathbf{x})$ where $\mathbf{x}\in\mathbf{R}^n$ is the original phase space variable. From $\{y_i\}_{i=1}^N$ one reconstructs an m-dimensional vector $\mathbf{y}_i$ using the delay coordinates [2,3]

$$\mathbf{y}_i=\{y(t_i),y(t_i-T),\ldots,y(t_i-(m-1)T)\}\,, \quad (2)$$

where $T>0$ is the delay time and m is the dimension of the reconstruction space. We call the mapping from $\{\mathbf{x}_i\}$ in $\mathbf{R}^n$ to $\{\mathbf{y}_i\}$ in $\mathbf{R}^m$ the "delay coordinate map." Results in Ref. [4] show that, for typical $T>0$ and $m>2D_0$, this delay coordinate map is one to one. Here D_0 is the box-counting dimension of the original chaotic attractor.

Our main focus is to estimate correlation dimension from a time series using delay coordinates [Eq. (2)]. As a point of departure for subsequent discussions, we first report a theorem [5,6] which shows that, for estimating the correlation dimension, $m\ge D_2$ suffices. We emphasize that this result holds true irrespective of whether the delay coordinate map is one to one or not. This is contrary to the commonly accepted notion that an embedding (one to one and differentiable) is needed for dimension estimation, leading to the false surmise that m needs to be at least $2D_2+1$ to guarantee a correct dimension estimation (see [7] for further discussion).

Consider an n-dimensional map $G:\mathbf{R}^n\to\mathbf{R}^n$. Let A be an attractor of G in $\mathbf{R}^n$ with a natural probability measure ρ. For a function $h:\mathbf{R}^n\to\mathbf{R}$, define a delay coordinate map $F_h:\mathbf{R}^n\to\mathbf{R}^m$ as

$$F_h(\mathbf{x})=[h(\mathbf{x}),h(G^{-1}(\mathbf{x})),\ldots,h(G^{-(m-1)}(\mathbf{x}))]\,.$$

The projected image of the attractor A under F_h has an induced natural probability measure $F_h(\rho)$ in $\mathbf{R}^m$. Furthermore, assume that G has only a finite number of periodic points of period less than or equal to m in A. The following result then applies.

Theorem.—If $D_2(\rho)\le m$, then for almost every h, $D_2(F_h(\rho))=D_2(\rho)$.

The theorem says that the correlation dimension is preserved under the delay coordinate map with $m\ge D_2(\rho)$. Similar results hold for flows generated by ordinary differential equations. "Almost every" in the statement is understood in the sense of prevalence defined in Ref. [4]; roughly speaking, we can regard this "almost every" as meaning that the functions h that do not give

0031-9007/93/70(25)/3872(4)$06.00

64

the stated result are very scarce and are not expected to occur in practice. The above dimension preservation result also holds for almost all general projection maps meeting the condition in the theorem. To illustrate, consider a closed curve with a uniform measure in $\mathbf{R}^3$. The dimension of this curve is 1. The projected image of this curve onto the plane still has a dimension of 1 but is generally self-intersecting. Thus the map is not one to one but preserves dimension information. One can further project the image to $\mathbf{R}^1$ and obtain an interval, which again has a dimension of 1 but bears little resemblance to the original curve in $\mathbf{R}^3$.

In applications D_2 is commonly extracted from a time series as follows (see Refs. [8-11] for reviews). First, an m-dimensional trajectory is constructed using Eq. (2). Then, the correlation integral $C_m(N,\epsilon)$ is computed according to Eq. (1), where m indicates the dimensionality of the reconstruction space. From the curve $\log C_m(N,\epsilon)$ vs $\log\epsilon$ one then locates a linear scaling region for small ϵ and estimates the slope of the curve over the linear region. This slope, denoted $\bar{D}_2^{(m)}$, is then taken as an estimate of the correlation dimension $D_2^{(m)}$ of the projection of the attractor to the m-dimensional reconstruction space. If these estimates $\bar{D}_2^{(m)}$, plotted as a function of m, appear to reach a plateau for a range of large enough m values, then we denote the plateaued value $\bar{D}_2$ and take $\bar{D}_2$ an estimate of the true correlation dimension D_2 for the system. From the theorem it is clear that the onset of this plateau should ideally start at $m=\mathrm{Ceil}(D_2)$, where $\mathrm{Ceil}(D_2)$, standing for ceiling of D_2, denotes the smallest integer greater than or equal to D_2.

Our original interest in the current problem was motivated by published reports (see Refs. [12-17] for a sample) where $\bar{D}_2^{(m)}$ plateaus at m values that are considerably greater than $\bar{D}_2$. A particular concern is that, when this happens, what does it imply regarding the correctness of the assertion that $\bar{D}_2$ is an estimate of the true correlation dimension D_2 of the underlying chaotic process? In an attempt to answer this question we have obtained new results on the systematic behavior of the correlation integrals. Based on these behaviors we are able to explain how factors such as a lack of sufficient data can produce results, resembling those seen in the experimental reports cited above, which seem to be inconsistent with the theorem. Furthermore, we find that even in cases where the plateau onset of $\bar{D}_2^{(m)}$ occurs at m values considerably greater than $\mathrm{Ceil}(\bar{D}_2)$, there are situations where the plateaued $\bar{D}_2$ is a good estimate of the true correlation dimension D_2. See Refs. [18-25] for other relevant works addressing the issue of short data sets and noise.

To study the numerical aspects of dimension estimation we use chaotic time series generated by the Mackey-Glass equation [26] $dy(t)/dt=ay(t-\tau)/\{1+[y(t-\tau)]^c\}-by(t)$, where we fix $a=0.2$, $b=0.1$, $c=10.0$, and $\tau=100.0$, and take as the initial condition $y(t)=0.5$ for $t\in[-\tau,0]$. The numerical integration of this equation is done by the following iterative scheme [1]:

$$y(t+\delta t)=\frac{2-b\delta t}{2+b\delta t}y(t)+\frac{\delta t}{2+b\delta t}\left\{\frac{ay(t-\tau)}{1+[y(t-\tau)]^{10}}+\frac{ay(t-\tau+\delta t)}{1+[y(t-\tau+\delta t)]^{10}}\right\}, \tag{3}$$

where δt is the integration step size. We choose $\delta t=0.1$. Equation (3) is then a 1000-dimensional map, which, aside from being an approximation to the original equation, is itself a dynamical system. The time series, generated with a sampling time $t_s=10.0$, are normalized to the unit interval so that the reconstructed attractor lies in the unit hypercube in the reconstruction space. The norm we use to calculate distances in Eq. (1) is the max-norm in which the distance between two points is the largest of all the component differences. To reconstruct the attractor, we follow Eq. (2) and take the delay time to be $T=20.0$ The dimension of the reconstruction space is varied from $m=2$ to $m=25$.

The first time series, used to illustrate the theorem, consists of 50000 points. For each reconstructed attractor at a given m we calculate the correlation integral $C_m(N,\epsilon)$ according to Eq. (1). In Fig. 1 we display $\log_2 C_m(N,\epsilon)$ vs $\log_2\epsilon$ for $m=2$–8, 11, 15, 19, 23. For each m we identify a scaling region for small ϵ and fit a straight line through the region. The open circles in Fig. 2 show the values of $\bar{D}_2^{(m)}$ so estimated as a function of m. For $m\le 7$, $\bar{D}_2^{(m)}\approx m$. For $m\ge 8$, $\bar{D}_2^{(m)}$ plateaus at $\bar{D}_2$ which has a value of about 7.1. Identifying $\bar{D}_2$ with the true correlation dimension D_2 of the underlying attractor, this result is consistent with the prediction by the theorem that the onset of the plateau occurs at $m=\mathrm{Ceil}(D_2)$.

The second time series, used to illustrate the effect

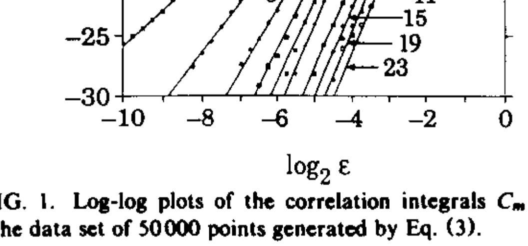


FIG. 1. Log-log plots of the correlation integrals $C_m(N,\epsilon)$ for the data set of 50000 points generated by Eq. (3).

65

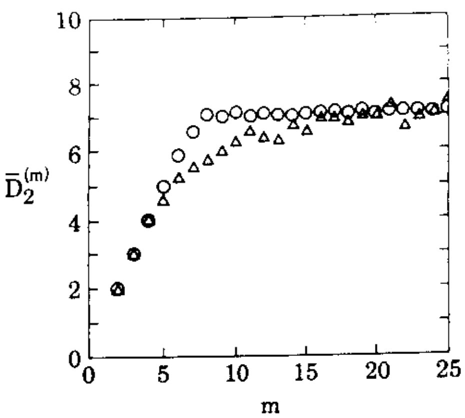


FIG. 2. $\bar{D}_2^{(m)}$ vs m plotted as open circles for the long data set ($N=50000$ and Fig. 1), $\bar{D}_2^{(m)}$ vs m plotted as triangles for the short data set ($N=2000$ and Fig. 3).

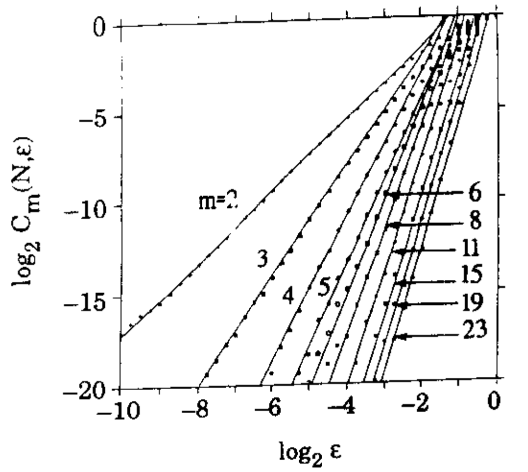


FIG. 3. $\log_2 C_m(N,\epsilon)$ vs $\log_2\epsilon$ for the data set of 2000 points.

due to a lack of data, consists of 2000 points. The $\log_2 C_m(N,\epsilon)$ vs $\log_2\epsilon$ curves are shown in Fig. 3 for $m=2$-6, 8, 11,15, 19, 23. The values of $\bar{D}_2^{(m)}$ in this case are plotted using triangles as a function of m in Fig. 2. This function attains an approximate plateau which begins at $m=16$ and extends beyond $m=25$. The slope averaged over the plateau is about 7.05 which is consistent with the value of 7.1 obtained using the long data set ($N=50000$) plotted as open circles in Fig. 2. But the D_2 estimates for the short data set fall systematically under that for the long data set for $5\leq m\leq 13$. Thus the plateau does not begin until m is substantially larger than $\mathrm{Ceil}(\bar{D}_2)$. This behavior has also been seen in many experimental studies. In what follows we explain the origin of this apparent inconsistency by exploring the systematic behavior of correlation integrals.

Figure 4 is a schematic diagram of a set of correlation integrals for $m=2$ to $m=13$. A dashed line is fit through the scaling region for each m. For $m\leq 5$, $\bar{D}_2^{(m)}\approx m$. For $m\geq 6$, $\bar{D}_2^{(m)}$ plateaus at $\bar{D}_2\approx 5.7$. This value is an estimate of the true D_2 for the system. This figure exhibits several features that are typical of correlation integrals for chaotic systems. The first feature we note is that the horizontal distance between $\log C_m(N,\epsilon)$ and $\log C_{m+1}(N,\epsilon)$ for $m\geq 6$ in the scaling regions is roughly a constant. This constant is predicted, for large m and small ϵ, to be $\Delta h=\Delta v/D_2$, where $\Delta v=TK_2$ with K_2 the correlation entropy [27] and T the delay time in Eq. (2). Two other significant features exhibited by Fig. 4 are as follows. For $m\leq 9$, $\log_2 C_m(N,\epsilon)$ increases with a gradually diminishing slope; while for $m\geq 11$, after exiting the linear region, the log-log plots in Fig. 4 first increase with a slope that is steeper than that in the linear scaling region and then level off to meet the point (0,0). These two different trends give rise to an uneven distribution in the extent of the scaling regions for different m with the most extended scaling region occurring at $m=10$.

In Refs. [8,28] arguments are presented to show that the trend observed for relatively small m is due to an "edge effect" resulting from the finite extent to the reconstructed attractor. Ding *et al.* [5] show that the steeper slope observed for relatively large m is caused by foldings occurring on the original attractor. This can be illustrated analytically [5] for the tent map [29]. For $m=1$, the correlation integral for the reconstructed tent map attractor is $C_1(\epsilon)=\epsilon(2-\epsilon)$. For $m=2$, $C_2(\epsilon)$ is written as $C_2(\epsilon)=C_1(\epsilon/2)+R(\epsilon)$. The first term arises because a pair of points y_j and y_l in the time series satisfying $|y_j-y_l|<\epsilon/2$ give rise to a pair of two-dimensional points, $\mathbf{y}_{j+1}=\{y_{j+1},y_j\}$ and $\mathbf{y}_{l+1}=\{y_{l+1},y_l\}$, satisfying $|\mathbf{y}_{j+1}-\mathbf{y}_{l+1}|<\epsilon$. The folding of the tent map at $y=\frac{1}{2}$ leads to situations in which $|y_j-y_l|>\epsilon/2$, but $|y_{j+1}-y_{l+1}|<\epsilon$. Thus the folding in the attractor underlies the correction term $R(\epsilon)$ which is calculated [5] to be $R(\epsilon)=\epsilon^2/2$ for $0\leq\epsilon<\frac{2}{3}$ and $R(\epsilon)=3\epsilon-7\epsilon^2/4-1$ for $\frac{2}{3}\leq\epsilon\leq 1$. For $0\leq\epsilon<\frac{2}{3}$, $d\log_2 C_2(\epsilon)/d\log_2\epsilon=1+\epsilon/(2+\epsilon)$. This derivative is 1 when $\epsilon=0$ ($D_2=1$ for the attractor) and increases due to the term $\epsilon/(2+\epsilon)$ whose presence reflects the influence of $R(\epsilon)$ which, in turn, is caused by the folding on the attractor.

From Eq. (1), the range of $C_m(N,\epsilon)$ is $\log_2 2/N^2$

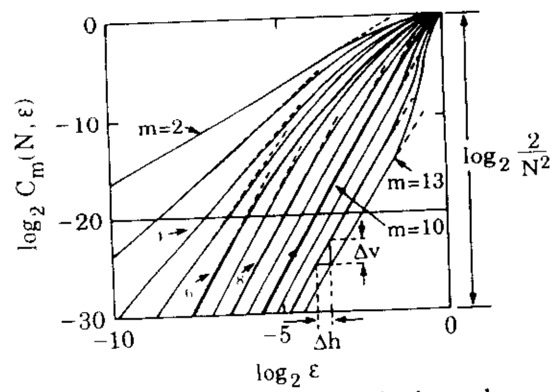


FIG. 4. Schematic diagram of correlation integrals.

66

$\leq \log_2 C_m(N,\epsilon) \leq 0$. Imagine a time series of $N=2000$ points generated by the system underlying Fig. 4. The plots of $\log_2 C_m(N,\epsilon)$ vs $\log_2 \epsilon$ for this data set roughly correspond to the portion of Fig. 4 above the horizontal line drawn at $\log_2 C_m(N,\epsilon) = \log_2[2/(2000)^2] \approx -20$. Since the upper boundary points of the scaling regions for $m=6$ and 7 are under this horizontal line, the correct dimension is not obtained for $m=6$ and 7. In fact, if we fit a straight line to an *apparent* linear region above $\log_2 C_m(N,\epsilon) = -20$ for $m=6$ we obtain a slope which is markedly smaller than the actual dimension. However, since the upper boundary points of the scaling regions for $m \geq 8$ are above the horizontal line, we can still expect to obtain the correct estimate of $D_2 = 5.7$ for $m \geq 8$. Thus the plateau onset is delayed due to a lack of data.

The same consideration applies to the short data set generated by Eq. (3). In particular, imagine that we restrict our attention to the region $\log_2 C_m(N,\epsilon) > -20$ in Fig. 1 and fit a line through an *apparent* linear range for the $m=8$ data in this region. The slope of this straight line is about 5.9, which is roughly the same as that of 5.8 estimated using 2000 points. Thus, by knowing the correlation integrals for a large data set, we can roughly predict the outcome of a dimension measurement based on a smaller subset of this data.

We remark that if one extends the range of m values beyond what is shown in Fig. 2, at large enough m, $\bar{D}_2^{(m)}$ will start to deviate from the plateau behavior and increase monotonically with m. This is caused by the finite length of the data set and can be understood from the systematic behavior of correlation integrals seen in Fig. 4. A lack of sufficient data will not only delay the plateau onset, but also make the deviation from the plateau behavior occur at smaller values of m, thus shortening the plateau length from both sides. This can again be understood with reference to Fig. 4.

This work was supported by the U.S. DOE (Office of Scientific Computing, Office of Energy Research) and the NSF (Divisions of Mathematical and Physical Sciences). M.D.'s research is also supported in part by a grant from the National Institute of Mental Health.

[1] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983); Physica (Amsterdam) **9D**, 189 (1983).

[2] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[3] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), p. 366; R. Mañe, *ibid.*, p. 230.

[4] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[5] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, Physica (Amsterdam) D (to be published).

[6] T. Sauer and J. A. Yorke (to be published).

[7] To get a rough idea of how widespread the misconception about the relevance of embedding to dimension estimation is, we conducted a literature search using the Science Citation Index for the years from 1987 to 1992 by looking for papers citing both Ref. [1] and Ref. [3]. We found 183 such papers. We then randomly selected a sample of 22 of these papers for closer examination. The following is what we found. Among the 22 papers there are 15 of them that calculate correlation dimension from time series. Five of these 15 papers make explicit connections between $2D+1$ and dimension estimation. The rest of these papers ignore this issue entirely. Based on this information we estimate that, during the period from 1987 to 1992, there are at least 42 papers (probably many more) where the authors implicitly or explicitly assumed that a one-to-one embedding is needed for dimension calculation. In addition, among the papers we researched for this work, only [8] and [9] imply that $m \geq D_2$ is sufficient for estimating D_2, although no justification is given.

[8] J. Theiler, J. Opt. Soc. Am. A **7**, 1055 (1990).

[9] P. Grassberger, in *Chaos*, edited by A. V. Holden (Manchester Univ. Press, Manchester, 1986).

[10] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos **1**, 521 (1991).

[11] *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress (Springer-Verlag, Berlin, 1986).

[12] J. Holzfuss and G. Mayer-Kress, in *Dimensions and Entropies in Chaotic Systems* (Ref. [11]).

[13] A. Branstater and H. L. Swinney, Phys. Rev. A **35**, 2207 (1987).

[14] U. Hübner, W. Klische, N. B. Abraham, and C. O. Weiss, in *Measures of Complexity and Chaos*, edited by N. B. Abraham *et al.* (Plenum, New York, 1990).

[15] K. R. Sreenivasan, in *Dimensions and Entropies in Chaotic Systems* (Ref. [11]).

[16] A. A. Tsonis and J. B. Elsner, Nature (London) **333**, 545 (1988).

[17] D. V. Vassiliadis, A. S. Sharma, T. T. Eastman, and K. Papadopoulos, Geophys. Res. Lett. **17**, 1841 (1990).

[18] A. Ben-Mizrachi, I. Procaccia, and P. Grassberger, Phys. Rev. A **29**, 975 (1984).

[19] J. G. Caputo, in *Measures of Complexity and Chaos*, edited by N. B. Abraham *et al.* (Plenum, New York, 1989).

[20] J. G. Caputo, B. Malraison, and P. Atten, in *Dimensions and Entropies in Chaotic Systems* (Ref. [11]).

[21] J.-P. Eckmann and D. Ruelle, Physica (Amsterdam) **56D**, 185 (1992).

[22] N. A. Gershenfeld, Physica (Amsterdam) **55D**, 155 (1992).

[23] J. W. Havstad and C. L. Ehlers, Phys. Rev. A **39**, 845 (1989).

[24] M. Möller, W. Lange, F. Mischke, N. B. Abraham, and U. Hübner, Phys. Lett. A **138**, 176 (1989).

[25] A. R. Osborne and A. Provenzale, Physica (Amsterdam) **35D**, 357 (1989).

[26] M. C. Mackey and L. Glass, Science **197**, 287 (1977).

[27] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983); Physica (Amsterdam) **13D**, 34 (1984).

[28] L. A. Smith, Phys. Lett. A **133**, 283 (1988).

[29] The tent map is defined as $y_{n+1} = T(y_n)$, where $T(y) = 2y$ if $0 \leq y < 1/2$ and $T(y) = 2 - 2y$ if $1/2 \leq y \leq 1$.

67