PHYSICS REPORTS (Review Section of Physics Letters) 193, No. 3 (1990) 137-163. North-Holland

POINCARÉ, CELESTIAL MECHANICS, DYNAMICAL-SYSTEMS THEORY AND "CHAOS"*

Philip HOLMES

Departments of Theoretical and Applied Mechanics, and Mathematics and Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA

Received October 1989

Contents:

1. Introduction 139
2. Two bodies, three bodies, reduction and Poincaré maps 140
3. Perturbation of integrable cases 148
4. The Smale–Birkhoff homoclinic theorem 151
5. Chaotic swinging 155
6. How attractive is chaos? 158
7. Discussion 160
References 162

Abstract:

As demonstrated by the success of James Gleick's recent book [1987], there is considerable interest in the scientific community and among the general public in "chaos" and the "new science" which is supposed to accompany it. However, as usual, it is not easy to separate hyperbole from fact. In an attempt to do this, I will offer a precise definition of chaos in the context of differential equations: mathematical models which, since Newton, have played a vital role in scientific discovery. I will show how the classical problems of celestial mechanics led Poincaré to ask fundamental questions on the qualitative behavior of differential equations, and to realize that chaotic orbits would provide obstructions to the conventional methods of solving them.

In a major paper which appeared almost exactly one hundred years ago, Poincaré studied mechanical systems with two degrees of freedom and identified an important class of solutions, now called transverse homoclinic orbits, the existence of which implies the system has no analytic integrals of motion other than the total (Hamiltonian) energy. I will explain these terms and outline the history of subsequent developments of these ideas by Birkhoff, Cartwright, Littlewood, Levinson and Smale, and describe how the ideas of Melnikov have made possible an "analytical algorithm" for the detection of chaos and proof of nonintegrability in wide classes of perturbed Hamitonian systems. I will discuss the physical implications of the mathematical statements that these methods afford. In the process, I will point out that, while there is a precise vocabulary and grammar of chaos, developed largely by mathematicians and stemming from Poincaré's work, it is not always easy to use it in speaking of the real world.

* An earlier version of this paper was delivered at a special A.A.A.S. session on the mathematical foundations of chaos at the Annual Meeting in San Francisco, January 15, 1989. The present version was delivered as the first part of the Mark Kac Memorial Lectures at Los Alamos National Laboratory on April 25, 1989. Parts of sections 3 and 4 are adapted from a paper in Chaos and Fractals (AMS, Providence, RI). The author thanks the Sherman Fairchild Foundation and the California Institute of Technology for their support during the preparation of this paper.

POINCARÉ, CELESTIAL MECHANICS, DYNAMICAL-SYSTEMS THEORY AND "CHAOS"

Philip HOLMES

Departments of Theoretical and Applied Mechanics, and Mathematics and Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA

NORTH-HOLLAND

"To doubt everything and to believe everything are two equally convenient solutions; each saves us from thinking."

H. Poincaré, *Science and Hypothesis* [Poincaré 1921]

1. Introduction

There is currently great excitement and much speculation about the "new science" of "chaos theory" and its potential role in our attempts to understand the world; yea, even the universe. The excitement is reflected in well over 5000 technical papers (cf. Shiraiwa [1985]), scores of reviews, monographs, proceedings, new journals and textbooks and now in the popular literature (cf. Gleick [1987], Stewart [1989]). In adding another piece of flotsam to this flood, I have taken a different viewpoint from many. Here you will find no computer simulations, not even in black and white, nor speculations about life, the universe and everything [Adams 1979]. Instead, what follows includes a brief (and biased) history of the mathematical foundations of the subject, a theorem of fundamental importance (with an outline of its proof), an analytical method which enables one to check the theorem's hypotheses, a simple but pretty example, and a discussion of some of the difficulties of extending the ideas to the study of mathematical models of physical interest.

For dynamical-systems theory, as it is more correctly if less spectacularly called, deals with the behavior of *mathematical* objects: primarily differential equations and their close relatives, iterated mappings. As such, it has little to say *directly* about the "real" world. Its ideas and methods help bridge the gulf between equation and solution, but they do not immediately help us build the equations from physical principles. Of course, they may *suggest* general strategies for the formulation of models: for example, the more we know about nonlinear analysis, the less we are tempted to linearize at the outset. It is usually better to be honest as long as possible, so that when we come to tell lies, they remain fresh in our memories. At the same time, we should beware of the tyranny of technique which afflicts many scientists: having ignored the work described below for 90 years, a certain type of researcher now sees chaos and strange attractors wherever (s)he looks.

Before we can import the insights of dynamical systems into model building and analysis, it is necessary to know what they are. Here I shall focus on a small but central part of the theory and initially I shall take a historical viewpoint. It is a good story.

One hundred years ago, Poincaré published a memoir [1890] describing the work for which he had been awarded one of the several mathematical prizes offered by King Oscar II of Sweden. This work addressed the stability of the solar system. Like a good scientist, he focused on a simplified model situation: the (restricted) three-body problem. His 270 page paper constitutes the first textbook in the qualitative theory of dynamical systems and, as I hope to show, several areas of current research have their origins in it. Poincaré's three-volume treatise [1899] and some of his earlier papers [1880–1890] contain more information and background. In particular, Poincaré describes the role that transverse homoclinic points play in obstructing the existence of "second" integrals of motion and in preventing the convergence of formal asymptotic methods, such as that of Linstedt.

Although we will be more specific later, we remark that a point q in the phase space of a dynamical system is called *homoclinic* to a fixed point p if the orbit is asymptotic to p as $t \to +\infty$ and as $t \to -\infty$. (It is called *heteroclinic* if the orbit is asymptotic to distinct points $p_+ \neq p_-$ as $t \to +\infty$ and $t \to -\infty$, respectively.) The point is transverse if the manifolds of initial conditions asymptotic to p as $t \to +\infty$ and as $t \to -\infty$, which necessarily intersect at q, do so transversely: that is, their tangent spaces span the whole phase space at q. (That these sets of initial conditions form smooth manifolds is the content of the fundamental stable manifold theorem, which I discuss in the second section). Transverse homoclinic points imply that "chaotic motions" exist nearby. One may wonder why it has taken the computer revolution for the majority of applied mathematicians and other scientists to wake up to these facts.

In the second and third sections of this paper, I outline the problem from celestial mechanics which Poincaré addressed and sketch his approach to it. Everything here is contained in Poincaré's paper, although my presentation takes a more "modern" viewpoint, especially in the treatment of perturbation and "Melnikov's method".

In section 4, I describe the central theorem and touch on the interesting history of this idea, which began when Poincaré realized that transverse homoclinic points would lead to complicated behavior, parts of which were subsequently characterized by Birkhoff, before Smale completed the description. In the process, detailed studies of a specific second-order differential equation due to Cartwright, Littlewood and Levinson, played an important role.

I then return to applications in sections 5 and 6, indicating first the sort of rigorous results which follow fairly easily from the Smale–Birkhoff theorem and perturbation methods, and then (some of) the difficulties which arise when one wishes to extend the ideas to describe the behavior of almost all solutions in a dissipative model which appears to possess a "strange attractor". In section 7 I conclude with a general discussion in which I permit myself some modest speculation.

This is far from a complete treatment of dynamical-systems theory. Although I naturally hope that it will generate interest and modestly inform the reader, this article cannot pretend to be a textbook. Alas, there are no short cuts to mastery of all the techniques. For those wishing to embark on a proper study, the books by Lefschetz [1957], Arnold [1973], Andronov et al. [1966], Hirsch and Smale [1974] or Wiggins [1990] provide good introductory material, while those of Arnold [1982], Palis and de Melo [1982], Irwin [1980], and (succumbing to chauvinism) Guckenheimer and Holmes [1983] contain more advanced material. Devaney [1986] has a treatment of iterated mappings (including complex analytic dynamics), starting at an elementary level. Readers with a background in classical mechanics will find that Lichtenberg and Lieberman [1983] and Arnold [1978] provide good routes to some of the recent ideas. Fundamental as well as more advanced mathematical material can be found in ODE texts such as Coddington and Levinson [1955], Hartman [1964] or Hale [1969].

Here I only deal with ordinary differential equations and I concentrate on Hamiltonian systems until section 6. Many of the ideas and results do, however, generalize to partial differential equations, see, e.g., Henry [1981], Temam [1988] and Constantin et al. [1989].

2. Two bodies, three bodies, reduction and Poincaré maps

In courses on classical mechanics (cf. Percival and Richards [1982], Goldstein [1980]) we learn that Newton's famous second law, $F = ma$, is equivalent to the elegant formulation of Hamilton in cases that the total energy is conserved. The equations of motion are then derived directly from the Hamiltonian, a real valued function $H(\boldsymbol{q}, \boldsymbol{p})$ defined on the $2n$-dimensional phase space, (locally) coordinatised by n configuration variables $\boldsymbol{q} = (q_1, q_2, \ldots, q_n)$ and their conjugate momenta $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$,

$$\dot{q}_j = \partial H/\partial p_j\,, \qquad \dot{p}_j = -\partial H/\partial q_j\,. \tag{2.1}$$

If $H(\boldsymbol{q}, \boldsymbol{p})$ does not depend explicitly on time, then a simple calculation using the chain rule confirms that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{j=1}^{n} \left(\frac{\partial H}{\partial q_j} \dot{q}_j + \frac{\partial H}{\partial p_j} \dot{p}_j \right) = \sum_{j=1}^{n} \left(\frac{\partial H}{\partial q_j} \frac{\partial H}{\partial p_j} - \frac{\partial H}{\partial p_j} \frac{\partial H}{\partial q_j} \right) \equiv 0 , \tag{2.2}$$

and thus the Hamiltonian (energy) is conserved.

As Newton showed, in the case of two bodies moving under their mutual gravitational attraction, use of the principles of conservation of linear and angular momentum permits one to reduce the study to a single degree of freedom: thus $n = 1$, and eqs. (2.1) and (2.2) take the simple forms

$$\dot{q} = \partial H / \partial p , \qquad \dot{p} = -\partial H / \partial q , \tag{2.3}$$

$$H(q, p) = h , \quad \text{a constant} . \tag{2.4}$$

The level sets of H are therefore curves in the two-dimensional phase space which are *invariant* under the evolution of (2.3); a solution started in a particular level h remains on that level for all time, positive and negative, (unless it runs off to infinity). Ranging through the values of h we cover the whole phase space, which is said to be *foliated* by a one (h) parameter family of such level curves. This implies that the single degree of freedom system is *completely integrable*, both in the classical sense that (2.4) can be inverted to solve for p in terms of q (and h) and the resulting relation integrated by quadratures [Goldstein 1980], and in the geometric sense implicit in Poincaré [1890] that the foliation of two-dimensional phase space by one-dimensional energy levels gives a complete qualitative description of all the solutions. A key point here is that the one-dimensional solution curves cannot intersect or cross one another in the two-dimensional phase space, otherwise uniqueness of solutions would fail. Thus a great order reigns: for the most part solutions which do not escape to infinity run around and around on closed, periodic orbits, as the following example illustrates.

Rather than considering the two-body problem itself (cf. Marion [1970], Goldstein [1980]), in anticipation of the examples to follow we take the simple pendulum, with Hamiltonian

$$H = p^2/2 + (1 - \cos q) , \tag{2.5}$$

and equations of motion

$$\dot{q} = p , \qquad \dot{p} = -\sin q . \tag{2.6}$$

The geometric structure of the phase space, with three families of periodic solutions separated by the level set $H = 2$, is well known (fig. 1), as is the quadrature

$$p = [2(h - 1 + \cos q)]^{1/2} = \mathrm{d}q/\mathrm{d}t$$

or

$$\int_{q_0}^{q(t)} \frac{\mathrm{d}q}{[2(h - 1 + \cos q)]^{1/2}} = t - t_0 , \tag{2.7}$$

which can be evaluated explicitly in terms of elliptic functions. Note that the separatrices $H = 2$ are composed of homoclinic points; solutions which are foward- and backward-asymptotic to the fixed point

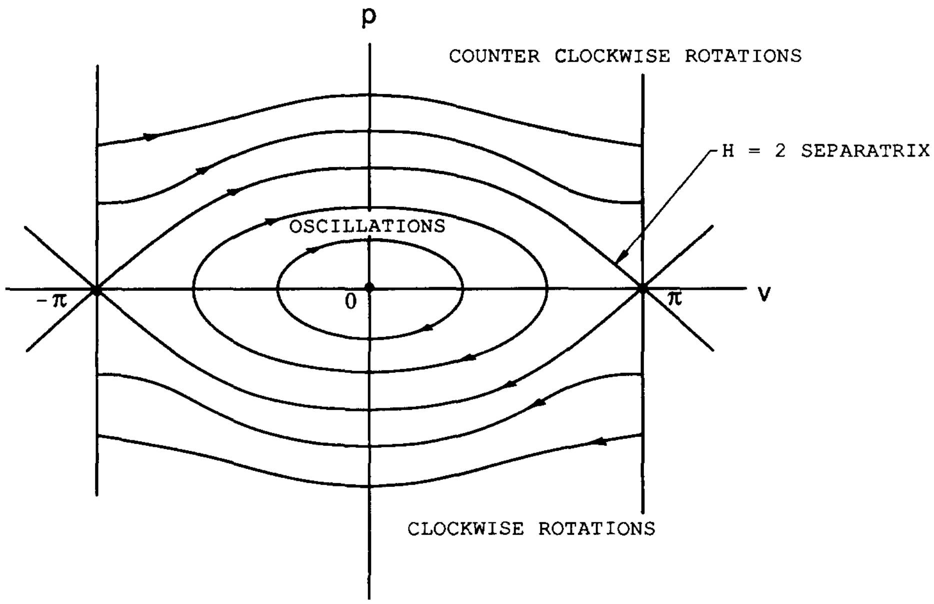

Fig. 1. The phase space of the simple pendulum.

$(q, p) = (\pm\pi, 0)$. The figure shows two distinct equilibria, but since they both correspond to the "upside down" pendulum, we should really identify all points $q = \pm\pi$ and wrap the phase space onto a cylinder. The separatrices will be of great importance later.

The restricted three-body problem comes in various flavors, one of which, the planar case, involves two massive bodies moving in circular Keplerian orbits on a plane with a third, small body moving under the influence of the resulting gravitational potential. If it is sufficiently small, the third mass does not influence the primaries and one may move to a rotating frame in which the two degrees of freedom are described by the position coordinates q_1, q_2 of the third body and their conjugate momenta, fig. 2. In this rotating coordinate system the Hamiltonian is no longer the total mechanical energy, but is rather the Jacobi integral which is time-independent for circular primary orbits.

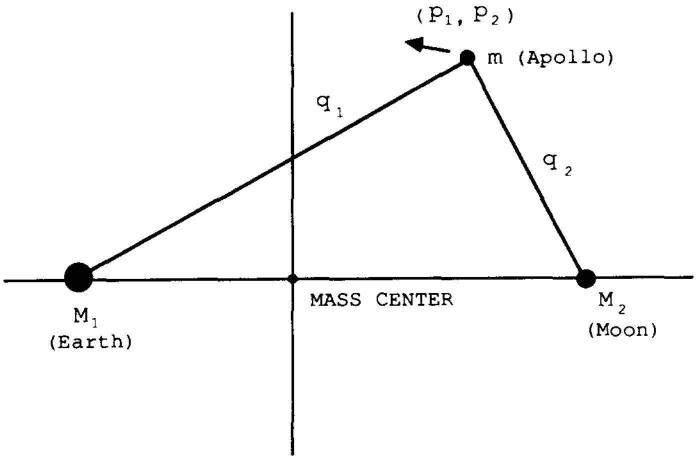

Fig. 2. The restricted planar, circular three-body problem.

The Hamiltonian is now a function of four variables (q_1, q_2, p_1, p_2) so that, while it is still conserved, its level sets are three-dimensional, allowing the solutions much greater freedom; see fig. 3. However, we may still solve for p_2, say, in terms of the remaining variables to obtain

$$p_2 = P_h(q_1, p_1; q_2)\,, \tag{2.8}$$

upon inversion of

$$H(q_1, q_2, p_1, p_2) = h\,. \tag{2.9}$$

Differentiation of (2.9) yields

$$\frac{\mathrm{d}H}{\mathrm{d}q_1} = 0 = \frac{\partial H}{\partial q_1} + \frac{\partial H}{\partial p_2}\frac{\partial P_h}{\partial q_1}\,, \qquad \frac{\mathrm{d}H}{\mathrm{d}p_1} = 0 = \frac{\partial H}{\partial p_1} + \frac{\partial H}{\partial p_2}\frac{\partial P_h}{\partial p_1}\,, \tag{2.10}$$

and if the coordinates are chosen so that the quantity

$$\mathrm{d}q_2/\mathrm{d}t = \partial H/\partial p_2 \neq 0 \tag{2.11}$$

does not vanish on (some subset of) the energy surface $H = h$, then we may eliminate explicit t-dependence in Hamilton's equations and write, using eq. (2.10),

$$\frac{\mathrm{d}q_1}{\mathrm{d}q_2} = \dot{q}_1/\dot{q}_2 = \frac{\partial H/\partial p_1}{\partial H/\partial p_2} = -\frac{\partial P_h}{\partial p_1}\,, \qquad \frac{\mathrm{d}p_1}{\mathrm{d}q_2} = \dot{p}_1/\dot{q}_2 = -\frac{\partial H/\partial q_1}{\partial H/\partial p_2} = +\frac{\partial P_h}{\partial q_1}\,. \tag{2.12}$$

Finally, letting $(\ )'$ denote $\mathrm{d}/\mathrm{d}q_2$, we obtain the *reduced* equations

$$q_1' = -\partial P_h/\partial p_1\,, \qquad p_1' = \partial P_h/\partial q_1\,, \tag{2.13}$$

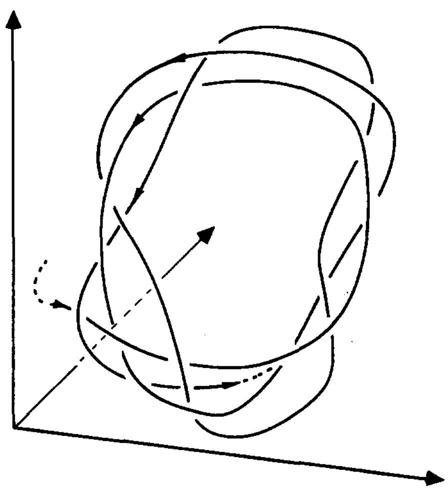

Fig. 3. In the three-dimensional manifold $H = h$, solutions can tie themselves in knots.

which are again Hamiltonian with the one-parameter family of functions $-P_h(q_1, p_1; q_2)$ as the new "energy".

If q_2 (and p_2) are classical angle (and action) variables, then dq_2/dt in (2.11) is generally positive and P_h is, moreover, 2π-periodic in q_2. It is then clear that (2.13) takes the form of a periodically forced single degree of freedom system in which the angle variable q_2 plays the role of a timelike coordinate. In fact one can think of reduction as removing one "oscillator" and replacing it with a "time" (q_2-)periodic external driver. Periodically forced systems like (2.13) are sometimes said to have one and a half degrees of freedom [Chirikov 1979]. A nice description of this procedure may be found in Birkhoff [1927]; see also Whittaker [1959], chapter 12, and, for the generalization to n degrees of freedom, Arnold [1978], section 45. More recently, Marsden and his colleagues have greatly extended and generalized the notion of reduction, see Marsden and Ratiu [1990] for example.

At this point note that the "physical" coordinates q_1, q_2 of the planar, circular problem of fig. 2 do not satisfy (2.11) and specifically that q_2 is not an angle variable. However, a suitable canonical transformation yields the required coordinates. Since this is not our main point, we omit the formulation and details.

Observe that the phase space of the reduced system (2.13), on each level set h, is three-dimensional. It is convenient to write the equations in the form

$$q_1' = -\frac{\partial P_h}{\partial p_1}(q_1, p_1; q_2), \qquad p_1' = \frac{\partial P_h}{\partial q_1}(q_1, p_1; q_2), \qquad q_2' = 1, \tag{2.14}$$

and to consider what is now called the *Poincaré map* induced by the flow of (2.14) on the *cross section* D given by $q_2 = 0$. Picking an initial point (q_1^0, p_1^0) on D, the image $(q_1^1, p_1^1) = P(q_1^0, p_1^0)$ under P is then the point at which the solution next intersects D, in other words, we integrate (2.14) until q_2 reaches 2π; see fig. 4. It is clear that a 2π-periodic orbit of (2.14) corresponds to a fixed point of P and a $2\pi k$-periodic orbit to a cycle of period k. Moreover, if P_h is smooth, P is an orientation preserving diffeomorphism: a smooth mapping with a smooth inverse such that the image of a region retains its original orientation. In addition, since eq. (2.14) is Hamiltonian and the flow preserves volume in (q_1, p_1, q_2)-space (Liouville's theorem), P preserves area. Also, stability types of fixed points and cycles of

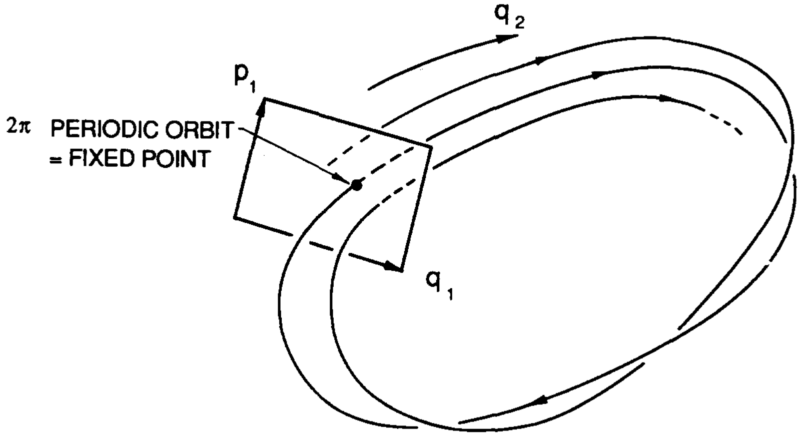

Fig. 4. The Poincaré map for $(q_1, p_1; q_2) \in D \times S^1$.

P and periodic orbits of (2.14) correspond. If γ is a periodic orbit of period 2π for (2.14) then the point (q_1^γ, p_1^γ) at which it intersects D is a fixed point or equilibrium for the map P: if γ has period $2\pi k$ then it intersects D in a k-periodic cycle of points which are mapped one into another by P. The reader may like to sketch two and three periodic cycles in the manner of fig. 4. The two-dimensional map P therefore captures crucial aspects of the solutions of the ordinary differential equation (2.14).

Of course, in general we cannot compute P explicitly (if we could we would have integrated the original equations), but as Poincaré realized, useful *qualitative* information can be drawn from the geometrical picture. In particular, Poincaré concentrated on periodic orbits of (2.14) and the corresponding fixed points of P and on the existence and characteristics of "asymptotic surfaces" belonging to saddle points of P, now called stable and unstable manifolds. This requires a brief review.

Let $P: \mathsf{R}^2 \to \mathsf{R}^2$ be a (smooth) map and $\boldsymbol{p}$ a fixed point $[\boldsymbol{p} = P(\boldsymbol{p})]$. We call the linear system

$$\boldsymbol{x} \mapsto DP(\boldsymbol{p})\boldsymbol{x} \tag{2.15}$$

the *linearization* of P at $\boldsymbol{p}$. $DP(\boldsymbol{p})$ is a 2×2 matrix; denote its eigenvalues λ_1, λ_2. One easily sees that $\boldsymbol{p}$ is stable if both eigenvalues of $DP(\boldsymbol{p})$ lie within the unit circle ($|\lambda_j| < 1, j = 1, 2$). If this is the case, we call $\boldsymbol{p}$ a *sink*. When $|\lambda_1| < 1 < |\lambda_2|$ $\boldsymbol{p}$ is an (unstable) *saddle point* and when $|\lambda_j| > 1, j = 1, 2$ $\boldsymbol{p}$ is a *source*. If $|\lambda_j| \neq 1$ for $j = 1, 2$, we call $\boldsymbol{p}$ *hyperbolic* and the Hartman–Grobman theorem (cf. Devaney [1986], Guckenheimer and Holmes [1983]) guarantees that the dynamical behavior of the linearization (2.1) holds in a neighborhood U of $\boldsymbol{p}$ for the fully nonlinear map P. The names of the fixed points derive from fluid mechanics, in fact one of the key ideas of the modern theory of dynamical systems is to view the phase space geometrically, and to see the totality of solutions of the differential equation as an evolution operator which transports the "phase fluid".

For our example of eq. (2.6), the fixed point(s) $(q, p) = (\pm\pi, 0)$ of the map $P = P_0$ are clearly saddle points. In fact the linearized map can be obtained by integrating the linearized differential equation linearized at $(q, p) = (\pm\pi, 0)$,

$$\dot{\xi}_1 = \xi_2\,, \qquad \dot{\xi}_2 = -\cos(\pm\pi)\,\xi_1 = \xi_1\,. \tag{2.16}$$

Elementary analysis shows that the fundamental solution matrix to this system may be written

$$\begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}, \tag{2.17}$$

and hence that the time T map, which gives DP_0, is

$$DP_0(\pm\pi, 0) \cdot \boldsymbol{\xi} = \begin{bmatrix} \cosh T & \sinh T \\ \sinh T & \cosh T \end{bmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}. \tag{2.18}$$

The matrix DP_0 has eigenvalues

$$\lambda_{1,2} = \cosh T \pm \sinh T = \mathrm{e}^T, \mathrm{e}^{-T}\,, \tag{2.19}$$

and since $\mathrm{e}^{-T} < 1 < \mathrm{e}^T$, the point(s) $(\pm\pi, 0)$ are, as expected, saddle points. As noted earlier, since θ is measured modulo 2π, and both equilibria correspond to the pendulum standing straight up, these points should be identified and the phase space "wrapped up" into a cylinder.

It is reasonable to believe, and possible to prove by a simple application of the implicit function theorem, that, for small perturbations involving (time-dependent) terms of $O(\varepsilon)$, P_0 perturbs to a nearby map $P_\varepsilon = P_0 + O(\varepsilon)$, which has a fixed point $p_\varepsilon = (\pi, 0) + O(\varepsilon)$ with eigenvalues $e^{-T} + O(\varepsilon) < e^{T} + O(\varepsilon)$. We use this fact in perturbation calculations in the next section.

The linear system (2.15) can be put into a convenient form by a suitable similarity transformation. In particular, if the eigenvalues are real and satisfy $|\lambda_1| < 1 < |\lambda_2|$, DP may be diagonalized, so that the linearization uncouples

$$u \mapsto \lambda_1 u\,, \qquad v \mapsto \lambda_2 v\,, \tag{2.20}$$

and the two axes $v = 0$, $u = 0$ are then the invariant *stable* and *unstable* subspaces, E^{s}, E^{u} (fig. 5a). The stable manifold theorem (cf. Guckenheimer and Holmes [1983], Devaney [1986]) asserts that, locally, the structure for the nonlinear system

$$x \mapsto P(x) \tag{2.21}$$

is qualitatively similar. More precisely, in a neighborhood U of p there exist *local* stable and unstable manifolds $W^{s}_{\mathrm{loc}}(p)$, $W^{u}_{\mathrm{loc}}(p)$, tangent to E^{s}, E^{u} at p, and as smooth as the map P. Recall that a (smooth) manifold is a space which locally looks like a piece of Euclidean space of the same dimension. Here we can think of the local manifolds $W^{s}_{\mathrm{loc}}(p)$, $W^{u}_{\mathrm{loc}}(p)$ as graphs – *curves* if both are one-dimensional – modelled on the *flat* stable and unstable subspaces E^{s}, E^{u} (fig. 5). Here the word *local* refers to a neighborhood U of p; a point belongs to $W^{s}_{\mathrm{loc}}(p)$ [or, respectively, to $W^{u}_{\mathrm{loc}}(p)$] if it and its images under p remain in U for all future iterations (or, respectively, backward iterations). That all of these points fit together to form smooth manifolds is the key conclusion of the stable manifold theorem. By taking backward and forward images of arcs contained in these manifolds, one constructs the *global* stable and unstable manifolds,

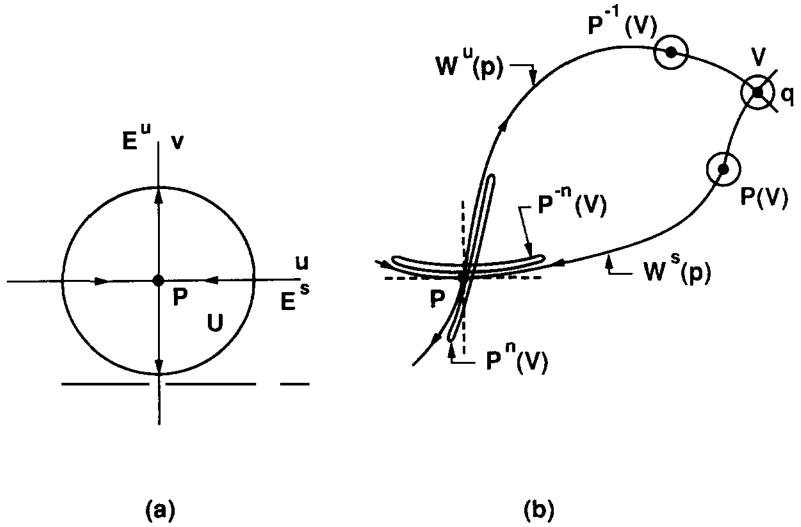

Fig. 5a. Invariant subspaces for the linear map; b. invariant manifolds for the nonlinear map, showing a homoclinic point, q.

$$W^{s}(p)=\bigcup_{n\geq 0} P^{-n}(W^{s}_{\mathrm{loc}}(p)), \qquad W^{u}(p)=\bigcup_{n\geq 0} P^{n}(W^{u}_{\mathrm{loc}}(p)), \tag{2.22}$$

which contain all points $x \in \mathbb{R}^2$ which are forward (or backward) asymptotic to p under iteration of P.

While the local structure is nice, the global structure need not be, and herein lies much of the reason for "chaotic motions", as we shall see. We call a point $q \in W^{u}(p) \cap W^{s}(p)$ a *homoclinic* point, following the terminology of Poincaré [1899]. By definition, the orbit $\{P^{n}(q)\}_{n=-\infty}^{\infty}$ of q is both forward and backward asymptotic to p. At first it may seem odd that one point can belong to both stable and unstable manifolds, but viewing it as an "initial condition" which has specific behavior in the future and the past may help. As I have already remarked, points on the separatrices of fig. 1 provide examples: in fact these separatrices are simultaneously stable and unstable manifolds for the saddle point $(\pm\pi, 0)$. For a two-dimensional differential equation like (2.6), such manifolds are one-dimensional curves and they must therefore either miss altogether, or coincide. For a map, however, they can intersect in other ways. In particular, if the manifolds $W^{s}(p)$, $W^{u}(p)$ intersect transversely at q, then iteration of a small region V containing q causes $P^{n}(V)$ and $P^{-n}(V)$ to "pile up" on $W^{u}(p)$, $W^{s}(p)$ respectively as $n \mapsto \infty$. The map P transports the images $P^{j}(V)$ around "astride" the stable and unstable manifolds for $j>0$ and $j<0$, respectively: once $P^{j}(V)$ is close to the saddle point p the linear contraction and expansion takes control [e.g., eq. (2.20)] and stretches the images as indicated in fig. 5b. (That this occurs in the controlled fashion of C^{1}-convergence of transversals to W^{u}, W^{s} at q is the content of the Lambda lemma [Newhouse 1980; Guckenheimer and Holmes 1983]; one does not need area preservation to prove it.) In such a situation the Smale–Birkhoff homoclinic theorem, described in section 4, shows that V and its images contain a rich and wonderful invariant set.

To anticipate a little, let me give a foretaste of this set in terms of the pendulum example of eq. (2.6) and fig. 1. When a time-periodic perturbation (external forcing, perhaps by a variable torque) is applied, the coincident manifolds forming the separatrix level set typically break up, but some homoclinic points may persist, and with them small neighborhoods of initial conditions which are repeatedly mapped around in the region formerly occupied by the separatrices. As fig. 5b indicates, such regions can now fall "on both sides" of the saddle point, so that two solutions starting nearby may find themselves separated – one corresponding to a rotation and the other to a libration. As they are repeatedly mapped past the saddle point, such solutions must again and again "decide" which route to take. Physically, a gentle tickling of the pendulum has dramatic consequences when it is near its unstable, inverted equilibrium. The global structure of the stable and unstable manifolds in this situation is what Poincaré, in a famous passage, would "not even attempt . . . to draw" [Poincaré 1899, chap. 33]. The remainder of this paper is mainly concerned with Poincaré's own, and subsequent, attempts to analyze this phenomenon.

Returning to our main theme, we ask what the structure of the Poincaré map P of the reduced system (2.14) will be in the event that the original two-degree of freedom Hamiltonian possesses an independent *second integral*, a function $F(q_1, q_2, p_1, p_2)$ which is constant on solutions and independent of H in the sense that $\nabla F \cdot \nabla H \neq 0$ almost everywhere. (The first property is equivalent to vanishing of the Poisson bracket

$$\{F, H\} = \sum_{j=1}^{2} \left(\frac{\partial F}{\partial q_j} \frac{\partial H}{\partial p_j} - \frac{\partial F}{\partial p_j} \frac{\partial H}{\partial q_j} \right), \tag{2.23}$$

and such functions are called integrals in involution [Goldstein 1980].)

Since the level sets of F and H are each three-dimensional, and intersect transversely almost everywhere, the level sets of F foliate the reduced three-dimensional space $H^{-1}(h)$ $((q_1, p_1; q_2)$-space) with a family of two-dimensional surfaces $F = f$. These in turn slice the cross section $D(q_2 = 0)$ in a family of curves, partitioning it in much the same way as the levels $H = h$ partition (q, p)-space in the single degree of freedom example of fig. 1. In this integrable case, the orbits of P, sequences of discrete points $\boldsymbol{x}_j = P(\boldsymbol{x}_{j-1})$, simply march around the "reduced" level curves.

Another way to see this is to recall that, if $H(q_1, q_2, p_1, p_2)$ is completely integrable, then it must (in suitable coordinates) possess a cyclic coordinate. Selecting this to be q_2, H is independent of q_2 and the conjugate momentum, p_2, is the second integral. In this case the reduced Hamiltonian $-P_h(q_1, p_1)$ is likewise independent of q_2 and the reduced system (2.13), (2.14) is an (integrable) single degree of freedom system since $\partial P_h/\partial p_1$ and $\partial P_h/\partial q_1$ are independent of q_2. Orbits of the Poincaré map P march around the level sets $P_h =$ constant of the analogue of fig. 1, for they are obtained by integrating this autonomous equation. Thus, if any homoclinic points exist for such a completely integrable system, they must lie on separatrices formed of smooth, coincident stable and unstable manifolds, much like the level set $H = 2$ of the pendulum.

In his memoir of 1890, Poincaré showed that, after use of perturbation methods and truncation of certain higher-order terms, the Hamiltonian for the restricted three-body problem becomes comletely integrable. Moreover, the reduced system (and hence its Poincaré map) possesses hyperbolic saddle points whose stable and unstable manifolds, being level sets of the second integral, coincide, as they do for the integrable pendulum example of fig. 1. He then asked a question of the type that has become central to the dynamical-systems approach (I paraphrase): "Should I expect this picture to persist if I restore the higher-order terms?" The important notion of *structural stability* refers to the situation in which small perturbations to a system of differential equations, or a map, do not cause qualitative changes in the structure of solutions. It is now known that integrable $n \geq 2$ degree of freedom Hamiltonian systems are not structurally stable in this sense. However, to answer the question in specific cases, such as the three-body problems, we need a little perturbation theory.

3. Perturbation of integrable cases

Here I shall briefly review the method of Melnikov [1963] which permits one to prove the existence of transverse homoclinic points in the Poincaré maps arising from specific examples of periodically perturbed differential equations. I concentrate on the Hamiltonian case, although Hamiltonian structure is not essential to the method. A rather different approach to the same problem can be found in Poincaré's [1890] paper on the three-body problem, and Arnold [1964] applied the idea to Hamiltonian systems around the same time as Melnikov. Thus, as Jerry Marsden has remarked, the method should probably be called the Poincaré–Arnold–Melnikov method.

I outline the simplest version of the method here. See Holmes and Marsden [1981, 1982a, b, 1983] and Wiggins [1988] for extensions to many (even infinitely many) dimensions. Consider a planar ordinary differential equation subject to a small time-periodic perturbation,

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \varepsilon \boldsymbol{g}(\boldsymbol{x}, t), \qquad \boldsymbol{g}(\boldsymbol{x}, t) = \boldsymbol{g}(\boldsymbol{x}, t + T), \quad \boldsymbol{x} \in \mathbb{R}^2. \tag{3.1}$$

Suppose that $\boldsymbol{f}$ and $\boldsymbol{g}$ are sufficiently smooth and bounded on bounded sets and that the unperturbed system is Hamiltonian, so there exists a function $F(\boldsymbol{x})$: $\mathbb{R}^2 \mapsto \mathbb{R}$ such that

$$\dot{x}_1 = f_1(x_1, x_2) = \partial F(x_1, x_2)/\partial x_2\,, \qquad \dot{x}_2 = f_2(x_1, x_2) = -\partial F(x_1, x_2)/\partial x_1\,. \tag{3.2}$$

Furthermore, assume that this unperturbed vector field contains a hyperbolic saddle point $\boldsymbol{p}_0$ lying in a closed set of F; thus there is a (degenerate, nontransversal) loop of homoclinic points, fig. 6a. The orbits on this loop are denoted by $\boldsymbol{x} = \boldsymbol{x}_0(t - t_0)$, where t_0 denotes a shift in the initial condition or base point. For precise technical hypotheses see Guckenheimer and Holmes [1983], section 4.5.

While the unperturbed ($\varepsilon = 0$) equation (3.1) has a two-dimensional phase space and solutions are ordered by the distinct level sets $F =$ constant, as soon as the perturbation is applied ($\varepsilon \neq 0$) this simple picture dissolves. The vector field is now time dependent, solutions can pass the same point $\boldsymbol{x}$ in different directions at different times (or "phases") and so it is better to extend the phase space by including t as a third variable, precisely as was done in eq. (2.14). In fact the attentive reader will notice that the setup under development in this section is perfectly designed for application to the reduced Hamiltonians of section 2.

Now consider the unperturbed and perturbed Poincaré maps P_1, P_ε corresponding to (3.1) with $\varepsilon = 0$ and $\varepsilon \neq 0$. The hyperbolic fixed point $\boldsymbol{p}_0$ of P_0 perturbs to a nearby hyperbolic fixed point $\boldsymbol{p}_\varepsilon = \boldsymbol{p}_0 + \mathrm{O}(\varepsilon)$ for P_ε and its stable and unstable manifolds remain close, as indicated in the sketch of fig. 6b. In fact the power series representations of solutions $\boldsymbol{x}_\varepsilon^{\mathrm{s,u}}$ lying in the perturbed stable and unstable manifolds of the small periodic orbit $\boldsymbol{\gamma}_\varepsilon = \boldsymbol{p}_0 + \mathrm{O}(\varepsilon)$ of eq. (3.1), $\varepsilon \neq 0$, are valid in the following *semi-infinite* time intervals:

$$\begin{aligned} \boldsymbol{x}_\varepsilon^{\mathrm{s}}(t, t_0) &= \boldsymbol{x}_0(t - t_0) + \varepsilon \boldsymbol{x}_1^{\mathrm{s}}(t, t_0) + \mathrm{O}(\varepsilon^2)\,, \quad t \in [t_0, \infty)\,, \\ \boldsymbol{x}_\varepsilon^{\mathrm{u}}(t, t_0) &= \boldsymbol{x}_0(t - t_0) + \varepsilon \boldsymbol{x}_1^{\mathrm{u}}(t, t_0) + \mathrm{O}(\varepsilon^2)\,, \quad t \in (-\infty, t_0]\,. \end{aligned} \tag{3.3}$$

This follows from the usual finite-time Gronwall estimates (e.g., Hartman [1964]) and the fact that these special solutions are "trapped" in the local stable and unstable manifolds and thus have well controlled asymptotic behavior as $t \to \pm\infty$, respectively. One can therefore seek the leading order terms $\boldsymbol{x}_1^{\mathrm{s,u}}(t, t_0)$ as solutions of the first variational equation obtained by substituting (3.3) into (3.1) and expanding in powers of ε,

$$\dot{\boldsymbol{x}}_1^{\mathrm{s,u}} = Df(\boldsymbol{x}_0(t - t_0))\boldsymbol{x}_1^{\mathrm{s,u}} + \boldsymbol{g}(\boldsymbol{x}_0(t - t_0), t)\,. \tag{3.4}$$

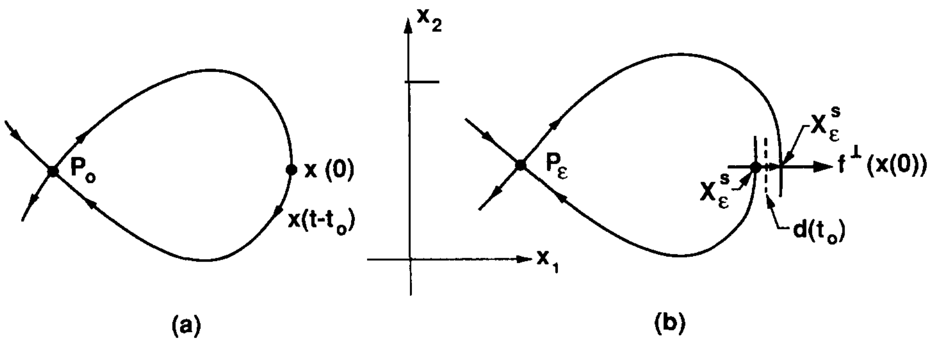

Fig. 6a. The unperturbed loop; b. the perturbed Poincaré map, showing stable and unstable manifolds on a cross section D (x_1, x_2)-space.

Now, while (3.4) is linear, it is usually rather hard to solve, since $Df(\boldsymbol{x}_0(t-t_0))$ is a time-varying 2×2 matrix and is not even periodic. Here Melnikov comes to our rescue. He realized that, to estimate the distance $d(t_0)$ between the perturbed stable and unstable manifolds at a base point t_0 of the unperturbed solution, one need not solve (3.4) explicitly. His method goes as follows.

From (3.3) and fig. 6b, we have

$$d(t_0) = \boldsymbol{x}_\varepsilon^{\mathrm{u}}(t_0, t_0) - \boldsymbol{x}_\varepsilon^{\mathrm{s}}(t_0, t_0) = \frac{\varepsilon(\boldsymbol{x}_1^{\mathrm{u}}(t_0, t_0) - \boldsymbol{x}_1^{\mathrm{s}}(t_0, t_0))\cdot \boldsymbol{f}^{\perp}(\boldsymbol{x}_0(0))}{\|\boldsymbol{f}(\boldsymbol{x}_0(0))\|} + \mathrm{O}(\varepsilon^2)\,, \tag{3.5}$$

where $\boldsymbol{f}^{\perp}(\boldsymbol{x}_0(0))$ denotes the normal to the unperturbed solution vector $\boldsymbol{f}(\boldsymbol{x}_0(0))$. Since $\boldsymbol{a}\cdot\boldsymbol{b}^{\perp} = \boldsymbol{b}\times\boldsymbol{a}$ for vectors in R^2, we can rewrite (3.5) as

$$d(t_0) = \varepsilon\,\frac{\boldsymbol{f}(\boldsymbol{x}_0(0))\times(\boldsymbol{x}_1^{\mathrm{u}}(t_0, t_0) - \boldsymbol{x}_1^{\mathrm{s}}(t_0, t_0))}{\|\boldsymbol{f}(\boldsymbol{x}_0(0))\|} + \mathrm{O}(\varepsilon^2) \stackrel{\mathrm{def}}{=} \varepsilon\,\frac{\Delta^{\mathrm{u}}(t_0, t_0) - \Delta^{\mathrm{s}}(t_0, t_0)}{\|\boldsymbol{f}(\boldsymbol{x}_0(0))\|} + \mathrm{O}(\varepsilon^2)\,. \tag{3.6}$$

If the quantity $\Delta^{\mathrm{u}} - \Delta^{\mathrm{s}}$ has simple zeros as t_0 varies, it follows form the implicit function theorem that, for $\varepsilon \neq 0$ small enough, the distance $d(t_0)$ passes through zero as t_0 varies and consequently that the perturbed manifolds intersect transversely. To compute $\Delta^{\mathrm{u}} - \Delta^{\mathrm{s}}$ we introduce time-varying functions

$$\Delta^{\mathrm{u,s}}(t, t_0) = \boldsymbol{f}(\boldsymbol{x}_0(t-t_0))\times \boldsymbol{x}_1^{\mathrm{u,s}}(t, t_0)$$

and compute

$$\begin{aligned}\dot{\Delta}^{\mathrm{s}} &= Df(\boldsymbol{x}_0)\dot{\boldsymbol{x}}_0 \times \boldsymbol{x}_1^{\mathrm{s}} + \boldsymbol{f}(\boldsymbol{x}_0)\times\dot{\boldsymbol{x}}_1^{\mathrm{s}} = Df(\boldsymbol{x}_0)\boldsymbol{f}(\boldsymbol{x}_0)\times\boldsymbol{x}_1^{\mathrm{s}} + \boldsymbol{f}(\boldsymbol{x}_0)\times[Df(\boldsymbol{x}_0)\boldsymbol{x}_1^{\mathrm{s}} + \boldsymbol{g}(\boldsymbol{x}_0, t)]\\ &= \operatorname{trace} Df(\boldsymbol{x}_0)\boldsymbol{f}(\boldsymbol{x}_0)\times\boldsymbol{x}_1^{\mathrm{s}} + \boldsymbol{f}(\boldsymbol{x}_0)\times\boldsymbol{g}(\boldsymbol{x}_0, t) = \boldsymbol{f}(\boldsymbol{x}_0(t-t_0))\times\boldsymbol{g}(\boldsymbol{x}_0(t-t_0), t)\,.\end{aligned} \tag{3.7}$$

Here we substitute for $\dot{\boldsymbol{x}}_1^{\mathrm{s}}$ from (3.4) and use $\dot{\boldsymbol{x}}_0 = \boldsymbol{f}(\boldsymbol{x}_0)$, a matrix cross product identity, and finally appeal to the fact that

$$\operatorname{trace} Df = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = \frac{\partial^2 F}{\partial x_1\,\partial x_2} - \frac{\partial^2 F}{\partial x_2\,\partial x_1} \equiv 0\,, \tag{3.8}$$

since f is Hamiltonian. Integrating (3.7) we have

$$\Delta^{\mathrm{s}}(t, t_0) - \Delta^{\mathrm{s}}(t_0, t_0) = \int_{t_0}^{t} \boldsymbol{f}(\boldsymbol{x}_0(s-t_0))\times\boldsymbol{g}(\boldsymbol{x}_0(s-t_0), s)\,\mathrm{d}s$$

and, taking the limit $t\to+\infty$ and using the fact that $\boldsymbol{f}(\boldsymbol{x}_0(t))\to\boldsymbol{f}(\boldsymbol{p}_0) = \boldsymbol{0}$ as $t\to\infty$, so that $\Delta^{\mathrm{s}}(t, t_0)\to 0$ we obtain

$$-\Delta^{\mathrm{s}}(t, t_0) = \int_{t_0}^{\infty} (\boldsymbol{f}\times\boldsymbol{g})(\boldsymbol{x}_0(s-t_0), s)\,\mathrm{d}s\,. \tag{3.9}$$

Note that we have used the validity of (3.3) on the semi-infinite interval $[t_0, \infty)$ in this computation. With a similar computation for Δ^{u}, (3.9) yields

$$\Delta^{\mathrm{s}}(t_0, t_0) - \Delta^{\mathrm{s}}(t_0, t_0) \stackrel{\mathrm{def}}{=} M(t_0) = \int_{-\infty}^{\infty} (f \times g)(x_0(s - t_0), s)\, \mathrm{d}s\,,$$

or, translating the variable s,

$$M(t_0) = \int_{-\infty}^{\infty} (f \times g)(x_0(s), s + t_0)\, \mathrm{d}s\,. \tag{3.10}$$

We have completed our sketch of the proof of

Melnikov's theorem. Under the hypotheses stated on (3.1), if $M(t_0)$ has simple zeros, then for $\varepsilon \neq 0$ sufficiently small, the manifolds $W^{\mathrm{s}}(p_\varepsilon)$, $W^{\mathrm{u}}(p_\varepsilon)$ intersect transversely. If $M(t_0)$ is bounded away from zero then $W^{\mathrm{s}}(p_\varepsilon) \cap W^{\mathrm{u}}(p_\varepsilon) = \varnothing$.

Although it is based on a simple perturbation method, this result is of considerable importance. $M(t_0)$ is an *explicitly computable function* which allows us to check in specific examples whether the stable and unstable manifolds intersect transversely or not, essentially by a direct calculation of the approximate distance between these manifolds from the viewpoint of an observer who follows the unperturbed separatrix on a fixed cross section. We will perform a specific computation, related to the pendulum problem, in section 5.

In order to show that stable and unstable manifolds split and intersect transversely, Poincaré [1890, section 19] performed a computation which is equivalent to (3.10), but his derivation involved power series approximations to solutions of the Hamilton–Jacobi equation. A good modern account of this can be found in Arnold et al. [1988, chap. 6.2]. We observe that, if the perturbation $\boldsymbol{g}$ is Hamiltonian with Hamiltonian G [as it will be if (3.1) takes the form (2.13)], then (3.10) can be rewritten in the elegant form

$$M(t_0) = \int_{-\infty}^{\infty} \{F, G\}(x_0(s), s + t_0)\, \mathrm{d}s\,, \tag{3.11}$$

where $\{\cdot, \cdot\}$ is the Poisson bracket, cf. Arnold [1964]. This formulation is useful in higher-dimensional generalizations [Holmes and Marsden 1981, 1982a, b, 1983; Wiggins 1988].

Next we will see why the transverse homoclinic points are so interesting.

4. The Smale–Birkhoff homoclinic theorem

As Poincaré [1890] realized, the presence of homoclinic points can vastly complicate dynamical behavior. However, the very fact that their existence implies recurrent motions makes the situation amenable to at least a partial analysis. Consider the effect of the map P of fig. 5b, containing a transverse homoclinic point $\boldsymbol{q}$ to a hyperbolic saddle $\boldsymbol{p}$, on a "rectangular" strip S containing $\boldsymbol{p}$ and $\boldsymbol{q}$ and segments of the stable and unstable manifolds in its boundary. As n increases, $P^n(S)$ is contracted horizontally and expanded vertically until the image $P^N(S)$ loops around and intersects S and P in a "horseshoe" shape (fig. 7).

To prove that the rates of contraction and expansion are uniformly bounded, one shrinks the width of S until many iterates occur for which $P^j(S)$ lies in a neighborhood U of p and the dynamics is therefore dominated by the linear map $DP(p)$ cf. eq: (2.20)].

A model for the situation was provided by Smale [1963], who introduced the map $F: S \mapsto \mathbb{R}^2$ of the square $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ sketched in fig. 8a. The map is linear on the two horizontal strips H_i whose images are the vertical strips V_i, $i = 0, 1$, the linearizations being

$$DF(x)|_{x \in H_1} = \begin{bmatrix} \lambda & 0 \\ 0 & \gamma \end{bmatrix}, \qquad DF(x)|_{x \in H_2} = \begin{bmatrix} -\lambda & 0 \\ 0 & -\gamma \end{bmatrix}, \tag{4.1}$$

with $0 < \lambda < 1 < \gamma$. Thus $F|_{H_i}$ contracts horizontally and expands vertically in a uniform manner. Smale

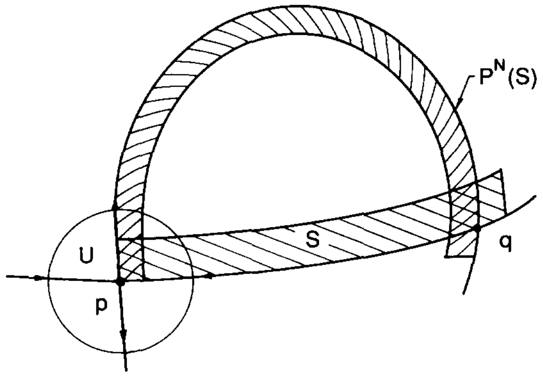

Fig. 7. P^N has a horseshoe.

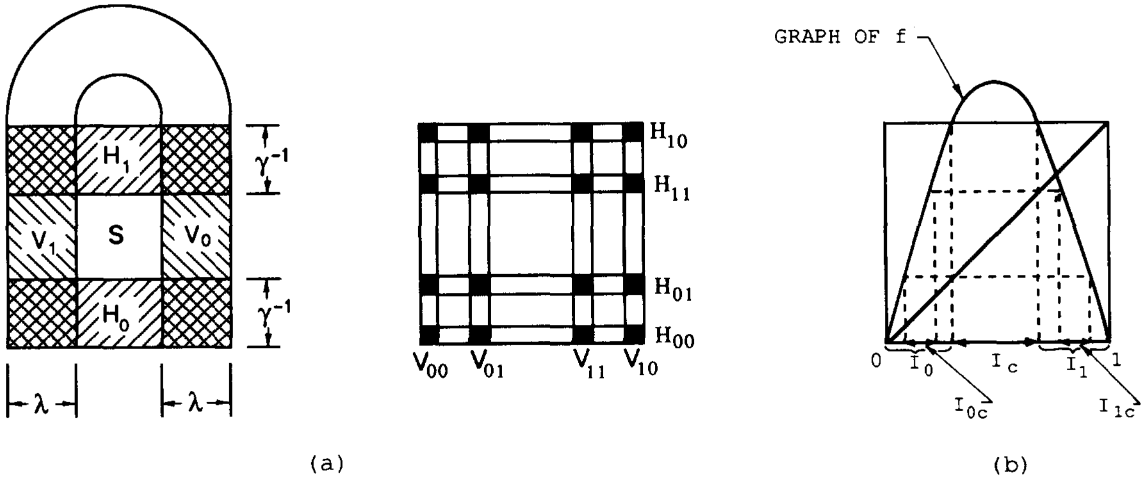

Fig. 8a. The two-dimensional horseshoe and b, its one-dimensional analogue.

studied the structure of the set of points Λ which never leave S under iteration of F. By definition, $\Lambda = \bigcap_{n=-\infty}^{\infty} F^n(S)$; the intersection of all images and preimages of S. Now

$$F^{-1}(S) \cap S = H_1 \cup H_2 , \qquad S \cap F(S) = V_1 \cup V_2$$

so $F^n(S)$ is the union of four rectangles of height γ^{-1} and width λ (fig. 8a). Similarly $\bigcap_{n=-2}^{2} F^n(S)$ is the union of 16 rectangles of height λ^{-2} and width λ^2, $\bigcap_{n=-k}^{k} F^n(S)$ is the union of 2^{2k} rectangles and, passing to the limit, Λ turns out to be a *Cantor set*: an uncountable point set, every member of which is a limit point. (Ian Stewart [1989] points out that the Cantor set was appropriated by Cantor from H. Smith, who constructed the first example some years before Cantor. For that matter "Newton's" equation $F = ma$, of section 2, was first explicitly written down by Euler in 1747 [Truesdell 1968].)

To see this more easily, consider the set of points which never leave $I = [0, 1] \subset \mathbb{R}$ under iteration of the one-dimensional map f of fig. 8b. After one iterate the "middle" interval I_c is lost, after two iterates its preimages I_{0c}, I_{1c} are lost, etc. Removing middle intervals of fixed proportional size (α, say) produces the classic "middle α" Cantor set Λ', the one-dimensional analogue of the construction of Λ. We remark that the map f is qualitatively like the famous quadratic map $x \mapsto ax(1-x)$ or $x \mapsto c - x^2$.

The sets Λ and Λ' can be coded in a way which describes their dynamics. To each $\boldsymbol{x} \in \Lambda$ we assign a bi-infinite sequence of the symbols 0, 1, $\phi(\boldsymbol{x}) = \{\phi_j(\boldsymbol{x})\}_{j=-\infty}^{\infty}$, by the rule $\phi_j(\boldsymbol{x}) = i$ if $F^j(\boldsymbol{x}) \in H_i$ ($i = 0, 1$). Thus $\phi_j(F(\boldsymbol{x})) = \phi_{j+1}(\boldsymbol{x})$ and the action of F on Λ corresponds to the action of the shift σ on the space of symbol sequences Σ. Moreover, every symbol sequence corresponds to an orbit realized by F, since the images V_i lie fully across their preimages H_i. In fact, the map ϕ: $\Lambda \mapsto \Sigma$ is a homeomorphism and the diagram

$$\begin{array}{ccc} \Lambda & \xrightarrow{F} & \Lambda \\ \downarrow\phi & & \downarrow\phi \\ \Sigma & \xrightarrow{\sigma} & \Sigma \end{array}$$

commutes. A homeomorphism is a continuous map with a continuous inverse, and that the diagram commutes means that $\phi(F(\boldsymbol{x})) = \sigma(\phi(\boldsymbol{x}))$. We say that $F|_\Lambda$ is *topologically conjugate to a (full) shift on two symbols*. This means that the dynamics of points $\boldsymbol{x} \in \Lambda$ under F is equivalent, in a strong sense, to the behavior coded in the corresponding sequence $\phi(\boldsymbol{x})$. Moreover, *every one* of the enormous variety of such binary sequences, which are as uncountable as the real numbers themselves, corresponds to a particular orbit of F which weaves its way through Λ.

To prove continuity of ϕ and its inverse, one needs a metric on Σ. The usual choice is

$$d(\boldsymbol{a}, \boldsymbol{b}) = \sum_{j=-\infty}^{\infty} \frac{|a_j - b_j|}{2^{|j|}} , \tag{4.2}$$

thus two sequences $\boldsymbol{a}$, $\boldsymbol{b}$ are close if their entries a_j, b_j agree on a long central block ($a_j = b_j$, $|j| \leq K$, large). One proves continuity of ϕ and ϕ^{-1} by showing that close points in Λ map to close sequences in Σ and vice versa; see Guckenheimer and Holmes [1983], section 5.1.

For $x \in \Lambda'$ one does the same but using only semi-infinite (positive going) sequences since f is non-invertible. The main advantage of this method of *symbolic dynamics* is that one can study the orbits of $F|_\Lambda$ (or $f|_{\Lambda'}$) combinatorially, by examining symbol sequences. For instance, the "constant" sequences $\ldots 000 \ldots \stackrel{\text{def}}{=} (0)'$ and $\ldots 111 \ldots = (1)'$ correspond to fixed points and the periodic sequences (01)′, (011)′, (0001)′, etc. to orbits of periods 2, 3, 4, etc. [here ( )′ denotes periodic extension]. In this way one proves the following:

Proposition. The invariant set Λ of the horseshoe contains: (1) a countable infinity of periodic orbits, including orbits of arbitrarily high period ($\simeq 2^k/k$ orbits of each period k); (2) an uncountable infinity of nonperiodic orbits, including countably many homoclinic and heteroclinic orbits, and (3) a dense orbit.

The dense orbit is obtained by concatenating all possible finite sequences of 0, 1; thus, as one can see from (4.2), by shifting one comes as close as one wishes to *any* other sequence. Since $F|_{H_1 \cup H_2}$ contracts uniformly by λ horizontally and expands by γ vertically, the eigenvalues $\mu_{1,2}$ of DF^k for any k-periodic orbit satisfy $|\mu_1| = \lambda^k < 1 < |\mu_2| = \gamma^k$ and thus all orbits are (unstable) saddles. In fact, all orbits in Λ have associated with them exponentially strong unstable manifolds and thus almost all pairs of points Λ separate exponentially fast under F^n [their symbol sequences differ for some (large?) N: $a_N \neq b_N$]. This *sensitive dependence on initial conditions* leads to what we popularly call "chaos." [Li and Yorke 1975]. Specifically, suppose that we know the initial conditions of some system with accuracy sufficient to determine only the first 50 binary "places". If the initial point lies in Λ then after 50 iterations we do not know whether the image lies in H_0 or H_1. And even more strikingly, since *every* bi-infinite sequence in Σ corresponds to an orbit of $F|_\Lambda$, there are uncountably many orbits which behave in a manner indistinguishable from the outcome of repeated tossing of a (probabilistic) coin; a quintessentially random process.

Perhaps most important is the fact that Λ is a *structurally stable* set; small perturbations $\tilde{F}$ of F possess a topologically conjugate set $\tilde{\Lambda} \sim \Lambda$; both are conjugate to the same shift on two symbols. In fact to prove the existence of such sets one does not need linearity of F or f, as in the example here; it is sufficient to establish uniform bounds on contraction and expansion. See Smale [1963], Moser [1973], and Guckenheimer and Holmes [1983] for more details. Wiggins [1988] does a nice job on the n-dimensional case and generalizations thereof. Infinite-dimensional versions of the theorem are also available.

The constructions we have sketched above and in figs. 7 and 8 lead one to the fundamental

Smale-Birkhoff homoclinic theorem. Let $P: \mathsf{R}^2 \mapsto \mathsf{R}^2$ be a diffeomorphism possessing a transversal homoclinic point $\boldsymbol{q}$ to a hyperbolic saddle point $\boldsymbol{p}$. Then, for some $N < \infty$, P as a hyperbolic invariant set Λ on which the Nth iterate P^N is topologically conjugate to a shift on two symbols.

Birkhoff [1927] had already proved the existence of countably many periodic points in any neighborhood of a homoclinic point, but Smale's construction provided a more complete picture and he extended it to R^n.

It is worth reflecting on how striking this result is. Relatively simple hypotheses, which can be checked in specific examples by explicit calculations such as those of section 3, lead not only to the remarkable conclusion that an infinite collection of "chaotic" and unstable periodic orbits exists, all gloriously mixed together, but a *complete qualitative description* of them is provided by their symbol sequences. Is it not perhaps foolish to call this "chaos"?

In a delightful essay, Smale [1980] describes how he was led to the horseshoe construction when his attention was drawn to a paper of Levinson [1949]. Levinson in turn was trying to simplify the complicated analysis of Cartwright and Littlewood [1945] who had studied Van der Pol equation for "relaxation oscillations",

$$\ddot{y} - k(1 - y^2)\dot{y} + y = b\lambda k \cos(\lambda t + a)\,, \tag{4.3}$$

with k large, in connection with problems arising in the British effort to develop radar during the Second World War. Note that eq. (4.3) is a periodically forced, single degree of freedom oscillator, although it is of course *not* Hamiltonian. Nonetheless, most of the remarks on three-dimensional phase space of sections 2 and 3 hold good. In particular, for certain ranges of (b, λ), Cartwright and Littlewood were able to prove the existence of an infinite set "of (unstable periodic motions) of a great variety of structures", as well as a set "of the power of the continuum, of nonperiodic limiting trajectories" and to show that each point of the former is a limit point for both points of itself and of the latter. This was essentially the horseshoe, the geometric structure of which Smale revealed and generalized. It is interesting to observe their footnote on this "very bizare . . . behavior.":

"Our faith in our results was at one time sustained only by the experimental evidence that stable subharmonics of two different orders did occur [Van der Pol and Van der Mark 1927]. It is this that leads to the startling consequences; the consequences themselves relate to nonstable motions (which the experiments naturally did not reveal)." The last sentence will return to haunt us below.

For a geometrical study, in the spirit of Smale, of the Van der Pol problem, see Levi [1981]. As this problem was successively studied and "simplified" the length of the papers increased in the interesting series 1, 9, 26, 147, . . . (pages).

To close this section we briefly indicate why the existence of a transverse homoclinic point precludes the existence of an analytic integral of motion independent of the total energy H, in the two degree of freedom case. We paraphrase Moser's [1973] argument. Recall that any such function is constant on solutions of the reduced problem and so is constant on orbits of its Poincaré map. Consequently it is constant on orbits contained in the horseshoe Λ, and in particular it is constant on the dense orbit. Such an analytic function must therefore be constant on Λ and hence on a neighborhood of Λ, such as the "square" S of fig. 8a. But then it can only be, trivially, constant on the whole phase space $(q_1, p_1; q_2)$ of the reduced system. The fact that a single dense orbit stitches the whole set Λ together is crucial to this argument. In the integrable case (cf. fig. 1) the complete set of periodic orbits themselves is generally dense, but this does not rule out the existence of further analytic constants of motion. Using different arguments involving periodic solutions and power series approximations, Poincaré [1890, section 22] also showed that no analytic integrals independent of energy existed in the presence of transverse homoclinic points.

5. Chaotic swinging

To illustrate how the Melnikov method and Smale–Birkhoff theorem apply in a particular example, we borrow the two degree of freedom Hamiltonian

$$H(q_1, q_2, p_1, p_2) = -p_2 - p_1^2 + 2\mu \sin^2(q_1/2) + \mu\varepsilon \sin q_1 \cos q_2 \tag{5.1}$$

from Poincaré's memoir [1890, section 19]. His variables have been renamed to conform with the discussion of section two, above. Note that here there are two parameters, μ and ε. We think of μ as fixed and consider the effect of increasing ε on the integrable system corresponding to $\varepsilon = 0$, in which limit q_2 is a cyclic coordinate and p_2 a second integral.

Hamilton's equations corresponding to (5.1) are

$$\dot{q}_1 = -2p_1 , \qquad \dot{q}_2 = -1 ; \qquad \dot{p}_1 = -\mu \sin q_1 - \mu\varepsilon \cos q_1 \cos q_2 , \qquad \dot{p}_2 = \mu\varepsilon \sin q_1 \sin q_2 . \tag{5.2}$$

Thus, $\dot{q}_2 < 0$ and we can invert (5.1) to obtain

$$p_2 = P_h(q_1, p_1; q_2) = h - p_1^2 + 2\mu \sin^2(q_1/2) + \mu\varepsilon \sin q_1 \cos q_2 , \tag{5.3}$$

and, via the reduction procedure of section two,

$$q_1' = -\partial P_h/\partial p_1 = 2p_1 , \qquad p_1' = \partial P_h/\partial q_1 = \mu \sin q_1 + \mu\varepsilon \cos q_1 \cos q_2 . \tag{5.4}$$

Observe that, for $\varepsilon = 0$ and after some rescaling, (5.4) is simply our old friend the pendulum of (2.6). In particular, it possesses a pair of homoclinic orbits connecting the saddle points at $(q_1, p_1) = (0, 0)$, $(2\pi, 0)$. These orbits lie in the level set $P_h = 0$, or

$$p_1 = \pm\sqrt{2\mu} \sin(q_1/2) , \tag{5.5}$$

and thus the region enclosed by them is of width $O(\sqrt{\mu})$. This will become significant later.

Applying the formula (3.10) to the example of (5.4), we find that the Melnikov function is

$$M(q_{20}) = \int_{-\infty}^{\infty} 2\mu p_1(q_2) \cos q_1(q_2) \cos(q_2 + q_{20}) \, dq_2 , \tag{5.6}$$

where $p_1(q_2)$ and $q_1(q_2)$ are the components of the unperturbed homoclinic motion. Using (5.5) we may re-express $\cos q_1 = 1 - p_1^2/\mu$, so that (5.6) becomes

$$M(q_{20}) = \int_{-\infty}^{\infty} 2\mu p_1 \left(1 - \frac{p_1^2}{\mu}\right) \cos(q_2 + q_{20}) \, dq_2 , \tag{5.7}$$

where, as the reader can check, the unperturbed (upper, +) solution is given by

$$p_1(q_2) = \sqrt{2\mu} \,\mathrm{sech}(\sqrt{2\mu} q_2) . \tag{5.8}$$

Since p_1 is even, the integral of (5.7) reduces to

$$M(q_{20}) = 2\mu I \cos(q_{20}) , \qquad I = \sqrt{2\mu} \int_{-\infty}^{\infty} S(1 - 2S^2) \cos q_2 \, dq_2 , \quad S = \mathrm{sech}(\sqrt{2\mu} q_2) , \tag{5.9}$$

and, to verify that transverse homoclinic points exist for $\varepsilon \neq 0$, small, we need only show that the integral $I \neq 0$.

Letting $\sqrt{2\mu} q_2 = \tau$, we have

$$I = \int_{-\infty}^{\infty} S(1 - 2S^2) \cos \omega\tau \, d\tau , \qquad S = \mathrm{sech}\, \tau , \quad \omega = \frac{1}{\sqrt{2\mu}} . \tag{5.10}$$

Integrating twice by parts and using the fact that the boundary terms vanish, since $\operatorname{sech}\tau \to 0$ as $\tau \to \pm\infty$, (5.10) becomes

$$I = -\omega^2 \int_{-\infty}^{\infty} S \cos \omega\tau \, \mathrm{d}\tau ,$$

which is easy to evaluate by the method of residues to yield

$$I = -(\pi/2\mu) \operatorname{sech}(\pi/2\sqrt{2\mu}) . \tag{5.11}$$

Note that this is nonzero for all $\mu \neq 0$, but that as $\mu \to 0$

$$M(q_{20}) = -\pi \operatorname{sech}(\pi/2\sqrt{2\mu}) \cos(q_{20}) \sim \mathrm{e}^{-\pi/2\sqrt{2\mu}} \tag{5.12}$$

approaches zero faster than any algebraic order in μ. Thus, without special care, we cannot treat μ and ε as independent parameters; μ must be fixed $\neq 0$ as ε is taken sufficiently small (depending on μ) for the implicit function arguments of section 3 to work.

It is worth comparing this computation with that of Poincaré, who obtains precisely the same answer for his integral J (up to a multiplicative factor of $\mathrm{i}/2\mu$, p. 223) or would have if he had not mysteriously dropped a factor of 2 in his calculation [1890, p. 222, third equation]. It is also worth remarking that, while explicit computations such as those above are not always possible, it is often easy to estimate Melnikov integrals and show that they are nonzero.

Poincaré also points out (p. 224) that the "splitting distance" (BB′, fig. 9, p. 220) is of exponentially small order in $\sqrt{\mu}$, cf. eq. (5.12). Since the model problem (5.1) represents the leading terms in a perturbation expansion close to a "resonant layer" of width $O(\sqrt{\mu})$ [cf. eq. (5.5)], in the original three-body problem as $\varepsilon \to 0$, $\mu \to 0$ also. The two small parameters are *not* independent and the validity of the simple perturbation technique, which depends upon fixing $\mu \neq 0$ and taking ε sufficiently small, is questionable [terms $O(\varepsilon^2)$ are larger than terms of $O(\varepsilon\, \mathrm{e}^{-c/\sqrt{\mu}})$ for μ small!]. Arnold [1964] and Melnikov [1963] also brushed against this problem but did not really come to grips with it. The present author at first failed to recognize it [Holmes 1980] (cf. Sanders [1982]), but recently Neishtadt [1984] and Holmes et al. [1988] have obtained uniform exponentially small splitting estimates (upper *and* lower bounds in the latter case) for some problems of this type. Kruskal and Segur [1987], Segur and Kruskal [1987], and Amick and McLeod [1990] have also recently studied problems in which manifolds are separated by distances "beyond all orders" in a small parameter. However, in many multiparameter cases, limits may be taken in such a way that the difficulty does not arise. This is a subtle and difficult problem, related to the removal of "tails" in normal form theory [Sternberg 1958] and yet another instance of Poincaré's prescience.

We conclude this section by showing that remarkable physical conclusions follow from the analysis above. Recall that the Poincaré example (5.1) is essentially a simple pendulum weakly coupled to a linear oscillator. In fig. 9 we indicate how a modest generalization of the horseshoe construction rises in the Poincaré map of this perturbed pendulum. The "horizontal" strips H_R, H_L are carried by P^N into "vertical" strips $P^N(H_\mathrm{R}) < P^N(H_\mathrm{L})$ as indicated. Since the saddle points near $(q_1, p_1) = (0, 0)$ and $(2\pi, 0)$ are identified, these images intersect H_R, H_L much as in the canonical Smale example of fig. 7 (cf. fig. 5).

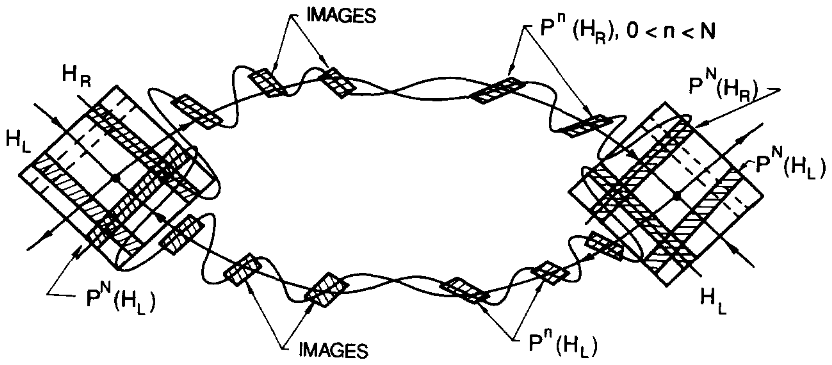

Fig. 9. Poincaré map for the perturbed pendulum.

As in section four, one obtains a homeomorphism between the shift on the two symbols R, L and some iterate P^N of the Poincaré map restricted to a suitable (Cantor) set $\Lambda^N = \bigcap_{n=-\infty}^{\infty} P^{nN}(H_R \cup H_L)$. Note that our construction guarantees that a point lying in H_R will be mapped around near the stable and unstable manifolds with $q_1' = p_1 > 0$ while a point lying in H_L is mapped around with $p_1 < 0$. Thus, an R in the symbol sequence corresponds to a passage of the pendulum bob past $(q_1, p_1) = (0, 0)$ with $p_1 > 0$ and an L to a passage with $p_1 < 0$. Since we have a full shift [$P^N(H_L)$ and $P^N(H_R)$ both lie across $H_L \cup H_R$] we conclude that any "random" sequence of the symbols L, R corresponds to an orbit of the pendulum, rotating "chaotically" to the left and to the right. If the reader finds fig. 9 confusing then (s)he will appreciate Poincaré's reluctance to draw it!

This conclusion is perhaps not too surprising, if we consider the effect of a small periodic perturbation of the undamped, Hamiltonian pendulum of fig. 1 swinging near its seperatrix orbit $H(q, p) = 2$. Each time the pendulum reaches the top of its swing, near the inverted, unstable state, the oscillation supplies a small push either to the left or right depending on the phase (time). Thus the precise time at which the bob arrives near this position is crucial and this, in turn, is determined by the time at which it left the same position after the preceding swing. Here is the physical interpretation of sensitive dependence upon initial conditions.

6. How attractive is chaos?

The analysis we have sketeched above leads to the rigorous proof of an infinite set of "chaotic" motions in a particular equation. The same ideas can be generalized and applied to a wide range of problems (cf. Wiggins [1988]). This seems like a satisfying state of affairs. But at this point honesty compels us to point that all is not rosy in the study of chaotic dynamics. Although this analysis establishes that a specific deterministic differential equation possesses chaotic orbits and provides an estimate for the parameter range(s) in which they exist, it does not necessarily imply that we have a *strange* or *chaotic attractor*. An *attractor* for a flow or map in an *indecomposable*, *closed*, *invariant* set for the flow or map, which attracts all orbits starting at points in some neighborhood. The maximal such neighborhood is the *domain of attraction* or *basin*. We call an attractor *strange* if it contains a transverse homoclinic orbit [Guckenheimer and Holmes 1983, chap. 5]. Here *indecomposable* means that the set cannot be separated into smaller, "basic" pieces: the existence of a dense orbit, as in the horseshoe, implies indecomposability. *Invariant* means that orbits starting in the set remain in it for all forward and backward time. The main physical consequence of indecomposability is that typical orbits attracted to the set continually wander about, exploring its entirety, and not settling down to some "simpler" subset. A strange attractor is therefore one in which almost all orbits exhibit the chaotic dynamics typical of the horseshoe orbits which correspond to the nonperiodic symbol sequences described in section four.

Since their flows preserve volume, the Hamiltonian systems we have considered here cannot possess attractors, but one can easily incorporate (small) damping terms into the Melnikov analysis and conclude that horseshoes exist in the weakly dampled, driven pendulum, for example,

$$\dot{q} = p\,, \qquad \dot{p} = -\sin q + \varepsilon \cos t - \delta p\,. \tag{6.1}$$

It is easy to see that all orbits of (6.1) remain trapped in a band $B = \{(q, p) \mid |p| \le \Gamma\}$ in the (cylindrical) phase space. If we choose $\Gamma > (1+\varepsilon)/\delta$, then the second component of (6.1) admits the bounds

$$\dot{p} \le |-\sin q| + \varepsilon|\cos t| - \delta|p| \le -\delta|p| + 1 + \varepsilon\,, \quad p > 0\,, \tag{6.2a}$$

$$\dot{p} \ge \delta|p| - (1+\varepsilon)\,, \quad p < 0\,, \tag{6.2b}$$

and thus the vector field points into B for all t. B is therefore a forward invariant region or *trapping region* for the Poincaré map P; the forward images of B all lie in the interior of B. At the same time, the reader can modestly extend the computations of section 5 to include the dissipation term and to conclude that, for (6.1),

$$M(t_0) = -8\delta + 2\varepsilon\pi\,\mathrm{sech}(\pi/2)\cos t_0\,, \tag{6.3}$$

and hence that there are transverse homoclinic orbits for ε, δ small provided that

$$\varepsilon\pi > 4\delta\cosh(\pi/2)\,. \tag{6.4}$$

Intuitively, if the force amplitude (ε) sufficiently exceeds the dissipation (δ), then motions which repeatedly pass the unstable, inverted equilibrium can be sustained. These enable the chaotic motions and the Melnikov calculation allows us to compute an approximate "stability boundary" for the existence of such motions.

In this case, the attracting set A can be defined as the intersection of all forward images of B and, since $\det DP = \exp(-2\pi\delta) < 1$ (a nice exercise), P contracts areas by a constant factor and the set

$$A = \bigcap_{n=0}^{\infty} P^n(B) \tag{6.5}$$

has zero area. The attracting set is therefore very "thin", but it need not be a topologically simple curve. In fact, in this case, A certainly contains the homoclinic points and their attendant horseshoes displayed above, and any attractors are certainly contained in A, but A itself need not be indecomposable. To display parameter values for which A *as a whole* and not just Λ, which is only *part* of A, contains a dense orbit seems to be very difficult. In fact, work of Newhouse [1980] on wild hyperbolic sets and the presence of infinitely many stable periodic orbits at certain parameter values for maps like

P shows that there are a lot of values for which A *cannot* be indecomposable. (Any candidate for a dense orbit, once trapped in the basin of a sink, would be out of the running!) Thus a "typical" solution approaching A might eventually settle down to stable periodic behavior, perhaps after a chaotic transient played out near Λ. (Recall that, in the horseshoe construction, almost all points starting in S lie in some preimage of the "middle α" strip and therefore leave S; they do not approach Λ.) In spite of the suggestive nature of numerical simulations, this issue still awaits clarification. It is better to say that P has a *strange attracting set*.

However, the canonical example of the iterated one-domensional map

$$x_{n+1} = \lambda x_n(1 - x_n) \stackrel{\text{def}}{=} f_\lambda(x_n) \tag{6.6}$$

is relevant here. In almost every popular or introductory article on chaos this example is used and it is pointed out that, as λ increases from 3 to 3.58... a "universal" sequence of period doubling bifurcations occurs after which "chaos sets in", punctuated by windows (in λ) of "periodic behavior". However, the question of the relative measure of *periodic* and *chaotic* λ, values in the range 3.58... to 4 was not settled until Jakobsen [1978, 1981] proved that there is a Cantor set Λ_c of strictly positive Legesgue measure of "chaotic" λ values. This means that, picking a parameter value between 3.58... and 4 at random, one has a finite probability of picking a map with a strange attractor, even through every such "chaotic" λ is surrounded by a (tiny) open set of periodic λ, for which f_λ has a simple periodic attractor. More precisely, for each $\lambda \in \Lambda_c$, f_λ possesses an absolutely continuous invariant measure supported on a collection of intervals. Almost all orbits therefore display "statistical" behavior and we have a genuine strange attractor, dense orbit and all. It is *almost* proved that the complement of Λ_c, clearly open, is also dense. But $\Lambda - \Lambda_c$ contains some rather pathological cases as well as those λ values for which f_λ has a stable periodic orbit (cf. Johnson [1987], Guckenheimer and Johnson [1990]). Even for this example the complete story is surprisingly complicated.

It now appears that part of the strange attractor problem for two-dimensional diffeomorphisms such as the Poincaré map of the damped, driven pendulum has also been solved. Benedicks and Carleson [1988] have recently announced a theorem implying that the Hénon map

$$x_{n+1} = y_n\,, \qquad y_{n+1} = -\varepsilon x_n + \mu - y_n^2 \tag{6.7}$$

does possess genuine strange attractors for a set of (μ, ε) values of strictly positive measure, with $\varepsilon \neq 0$ and small. More precisely, they show that (6.7) has a strictly positive Lyapunov exponent for (μ, ε) contained in such a set; the existence of smooth invariant measures is still open. Mora and Viana [1990] have now extended this result. (For $\varepsilon = 0$, (6.7) collapses to (6.6), cf. Holmes and Whitley [1984]). This is an essential step in the study of maps which fold rather than those, like the Lorenz model, which cut. Moreover, Hénon-like maps appear near homoclinic points as the latter are created in (quadratic) tangencies of stable and unstable manifolds; consider, for example, what happens in (6.1) as ε increases for fixed $\delta > 0$ (cf. Guckenheimer and Holmes [1983], sections 6.6, 6.7). When these results are put together, we may be close to a satisfactory resolution of the strange attractor problem.

7. Discussion

In the preceding pages we have seen how a simple perturbation calculation, together with the Smale–Birkhoff theorem, can provide proof of the existence of "chaos" in specific differential equations. We have seen how the resulting "unpredictability", or sensitive dependence on initial conditions, follows naturally from the existence of a topological conjugacy between orbits in phase space and the shift map on symbol sequences. However, we have also learned that this is not sufficient to conclude the existence of a strange attractor: that almost all orbits behave unpredictably. What morals can be drawn from this type of analysis?

The first is that the *paradigms* of abstract dynamical-systems theory, with its stress on structural stability and generic properties (cf. Smale [1967]), provide invaluable guidance in the study of specific problems. This sort of theory tells us what to look for. It is commonplace in physics that theoretical biases and expectations influence our interpretation of the data. Before applied scientists woke up to the possibility of strange attractors and chaos, they tended to ignore a lot of "anomolous numerical results". Of course, now there is a tendency to over-react; I doubt that strange attractors are *quite* as prevalent, or important, as the many announcements in Phys. Rev. Lett., Phys. Lett. A, etc., suggest. However, the general theory does provide us with lists of certain species and hunting licenses. This brings me to my second point.

Methods are now becoming available for the detection of solutions with interesting global properties in specific nonlinear models. That discussed in the present paper is only one example; the reader will find many more in the references. Numerical simulations play an important part here, although I have not considered them in this paper. There is also a very nice interplay between numerical experiment and rigorous results, clearest perhaps in the careful simulations of Jim Yorke and his colleagues and the results on numerical approximation of "true" orbits which they have obtained (e.g., Hammel et al. [1988], Nusse and Yorke [1989a, b]). A further important strand is the increasing quantity and quality of experimental evidence which demonstrates that chaotic motions and strange attractors do appear in a diverse range of physical systems, including apparently simple electrical circuits and mechanical devices (Bergé, et al. 1987, Moon 1987]. (The evidence in turbulence, with respect to which the relevance of strange attractors was suggested almost twenty years ago [Ruelle and Takens 1971], is more problematic, especially in open flows and fully developed turbulence. See Holmes [1990], for example.) The ideas of dynamical-systems theory has also led to a number of new methods of data analysis; dimension computations, Lyapunov exponents, phase-space reconstruction, spectra of dimensions, etc., although much remains to be learned about the applicability and validity of such methods.

The "paradigm" mode has its dangers. Attempts to understand (pieces of) the world often start with metaphor: one studies a simpler system which "looks like" the real object of interest, but one does not insist on fundamental connections or derivations "from first principles". The fact that small sets of differential equations and iterated maps can exhibit complicated behavior has stimulated a lot of such metaphorical studies. Dynamical-systems theory has potentially dangerous side effects in this resepct. Certain persons seem to prefer to abandon hard won, detailed knowledge of problems like turbulence in boundary or shear layers in favor of metaphors, such as coupled map lattices, which have little obvious connection with the underlying physics. In most cases it still remains to relate these to fundamental principles, and so in turn the metaphors into models. However, it is certainly true that the notion of *universality* – topological and metric behavior which is, to some extent, *independent* of the precise system – confers respectability on some such metaphors and reveals underlying similarities of structure across fields of application. Unfortunately, none of the famous "bifurcation to chaos" scenarios – period doubling, quasiperiodic, etc. – are sufficiently general that we can entirely dispense with analyses to determine what actually occurs in specific models. The fact that we have methods as well as paradigms is central to this enterprise. (In the biological sciences, where the problems are more difficult and less clearly defined, the metaphorical approach is of considerable value, e.g., Glass and Mackey [1988]).

The third moral is that mathematics and physics are coming closer again. Dynamical systems theory is just one instance of a general movement. In *The Value of Science*, Poincaré [1921] observes,

> "The mathematician should not be for the physicist a mere purveyor of formula; there should be between them a more intimate collaboration. Mathematical physics and pure analysis are not merely adjacent powers, maintaining good neighborly relations; they mutually interpenetrate and their spirit is the same."

In spite of this, I do not really see a "new science" here, in particular I do not see that "chaos theory" even exists as a coherent object, for example like the quantum and relativity theories. I would certainly hesitate to label it SUPERB, as Penrose [1989] labels these two theories in his recent critique of another new science: artifical intelligence. However, we do have a loose collection of ideas and methods, many of the latter inherited from "classical" applied mathematics, which we can add to the scientist's toolbox. There is a great ferment of excitement and activity. The artificial distinction between pure and applied mathematics is weakening. Mathematicians and scientists from different fields are talking to one another. Some are even listening.

References

Adams, D., 1979, The Hitchiker's Guide to the Galaxy (Pocket Books, New York).

Amick, C.J. and J.B. McLeod, 1990, A singular perturbation problem in needle crystals, Arch. Ration. Mech. Anal. 109, 139–171.

Andronov, A.A., E.A. Vitt, and S. E. Khaiken, 1966, Theory of Oscillators (Pergamon, New York).

Arnold, V.I., 1964, Instability of dynamical systems with several degrees of freedom, Sov. Math. Dokl. 5, 581–585.

Arnold, V.I., 1973, Ordinary Differential Equations (M.I.T. Press, Boston).

Arnold, V.I., 1978, Mathematical Methods of Classical Mechanics (Springer, New York).

Arnold, V.I., 1982, Geometrical methods in the Theory of Ordinary Differential Equations (Springer, New York).

Arnold, V.I., V.V. Kozlov and A.I. Neishtadt, 1988, Mathematical Aspects of Classical and Celestial Mechanics (Springer, New York). (Encyclopedia of Mathematical Sciences, Vol. III, ed. V.I. Arnold (originally published in Moscow, 1985).

Benedicks, M. and L. Carleson, 1988, The dynamics of the Hénon map, submitted for publication.

Bergé, P., Y. Pomeau and C. Vidal, 1987, Order Within Chaos (Wiley, New York) (French original, 1984).

Birkhoff, G.D., 1927, Dynamical Systems (A.M.S. Publications).

Cartwright, M.L. and J.E. Littlewood, 1945, On nonlinear differential equations of the second order, I: the equation $\ddot{y} - k(1 - y^2)\dot{y} + y = b\lambda k \cos(\lambda t + a)$, k large. J. London Math. Soc. 20, 180–189.

Chirikov, B.V., 1979, A universal instability of many dimensional oscillator systems, Phys. Rep. 52, 263–379.

Coddington, E.A. and N. Levinson, 1955, Theory of Ordinary Differential Equations (McGraw-Hill, New York).

Constantin, P., C. Foias, R. Temam and B. Nicolaenko, 1989, Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations (Springer, Berlin).

Devaney, R.L., 1986, An Introduction to Chaotic Dynamical Systems (Benjamin/Cummings, Menlo Park, CA).

Glass, L. and M.C. Mackey, 1988, From Clocks to Chaos: the Rhythms of Life (Princeton Univ. Press, Princeton).

Gleick, J. 1987, Chaos: Making a New Science (Viking, New York).

Goldstein H, 1980, Classical Mechanics (Addison Wesley, Redding, MA).

Guckenheimer, J. and P. Holmes, 1983, Nonlinear Oscillations, Dynamical Systems and Bifurcation of Vectorfields (Springer, New York).

Guckenheimer, J. and S. Johnson, 1990, Distortion of S-unimodal maps, Ann. Math. 132, 71–130.

Hale, J.K., 1969, Ordinary Differential Equations (Wiley, New York).

Hammel, S.M., J.A. Yorke and C. Grebogi, 1988, Numerial orbits of chaotic processes represent true orbits, Bull. A.M.S. (New Series) 19, 2, 465–469.

Hartman, R., 1964, Ordinary Differential Equations (Wiley, New York).

Henry, D., 1981, Geometric Theory of Semilinear Parabolic Equations, Springer Lecture Notes in Mathematics No. 840 (Springer, New York).

Hirsch M.W. and S. Smale, 1974, Differential Equations, Dynamical Systems and Linear Algebra (Academic Press, New York).

Holmes, P., 1980, Averaging and chaotic motions in forced oscillators, SIAM J. Appl. Math. 38, 65–80.

Holmes, P. 1990, Can dynamical systems approach turbulence? in: Whither Turbulence, ed. J.L. Lumley, Springer Lecture Notes in Applied Physics, 357, 195–249, 306–309.

Holmes, P. and J.E. Marsden, 1981, A partial differential equation with infinitely many periodic orbits: chaotic oscillations of a forced beam. Arch. Ration. Mech. Anal. 76, 135–166.

Holmes, P. and J.E. Marsden, 1982a, Horseshoes in perturbations of Hamiltonians with two degrees of freedom. Commun. Math. Phys. 82, 523–544.
Holmes, P. and J.E. Marsden, 1982b, Melnikov's method and Arnold diffusion for perturbations of integrable Hamiltonian systems, J. Math. Phys. 23, 669–675.
Holmes, P. and J.E. Mardsen, 1983, Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups, Indiana Univ. Math. J. 32, 273–310.
Holmes, P. and D. Whitley, 1984, Bifurcation of one and two dimensional maps. Philos. Trans. R. Soc. London A 311, 43–102.
Holmes, P., J.E. Marsden and J. Scheurle, 1988, Exponentially small splittings of separatrices with applications to KAM theory and degenerate bifurcations, in: Hamiltonian Dynamical Systems, eds K.R. Meyer and D.G. Saari, Cont. Math. 81, 213–243.
Irwin, M.C., 1980, Smooth Dynamical Systems (Academic Press, New York).
Jakobsen, M.V., 1978, Topological and metric properties of one dimensional endomorphisms, Sov. Math. Dokl. 19, 1452–1456.
Jakobsen, M.V., 1981, Absolutely continuous invariant measures for one parameter families of one dimensional maps, Commun. Math. Phys. 81, 39–88.
Johnson, S., 1987, Singular measures without restrictive intervals, Commun. Math. Phys. 110, 185–190.
Kruskal, M. and H. Segur, 1987, Asymptotics beyond all orders in a model of dendritic crystals, preprint.
Lefschetz, S., 1957, Ordinary Differential Equations: Geometric Theory (Interscience, New York; reissued by Dover Publ., New York, 1977).
Levi, M., 1981, Qualitative analysis of the periodically forced relaxation oscillations, Memoirs A.M.S. 214, 1–147.
Levinson, N., 1949, A second-order differential equation with singular solutions, Ann. Math. 50, 127–153.
Li. T.Y. and J.A. Yorke 1975, Period three implies chaos. Am. Math. Monthly 82 985–992.
Lichtenberg, A.J. and M.A. Lieberman, 1983, Regular and Stochastic Motion (Springer, New York).
Marsden, J.E. and T. Ratiu, 1990, Mechanics and Symmetry (forthcoming book).
Marion, J.B., 1970, Classical Dynamics of Particles and Systems (Academic Press, New York).
Melnikov, V.K. 1983, On the stability of the center for time periodic perturbations, Trans. Moscow Math. Soc. 12, 1–57.
Moon, F.C., 1987, Chaotic Vibrations (Wiley, New York).
Mora, L. and M. Viana, 1990, Abundance of strange attractors, preprint, IMPA, Rio de Janeiro, Brasil.
Moser, J., 1973, Stable and Random Motions in Dynamical Systems, (Princeton Univ. Press, Princeton).
Neishtadt, A., 1984, The separation of motions in systems with rapidly rotating phase, P.M.M. USSR 48, 133–139.
Newhouse, S.E., 1980, Lectures on dynamical systems, in: Dynamical Systems, CIME Lectures (Bressanone, Italy, 1978) (Birkhauser, Boston) pp. 1–114.
Nusse, H.E. and J.A. Yorke, 1989a, A procedure for finding numerical trajectories on chaotic saddles, preprint.
Nuse, H.E. and J.A. Yorke, 1989b, Analysis of a procedure for finding numerical trajectories on chaotic saddles, preprint.
Palis, J. and W. de Melo, 1982, Geometric Theory of Dynamical Systems: An Introduction, (Springer, New York).
Penrose, R., 1989, The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics (Oxford Univ. Press, Oxford).
Percival, I. and D. Richards, 1982, Introduction to Dynamics, (Cambridge Univ. Press, Cambridge).
Poincaré, H., 1880–1890, Mémoire sur les courbes définies par les équations différentielles I–IV, Oeuvre I (Gauthier Villars, Paris).
Poincaré, H., 1890, Sur les équations de la dynamique et le problème des trois corps. Acta Math. 13, 1–270.
Poincaré, H., 1899, Les Methodes Nouvelles de la Mécanique Celeste, Vols. 1–3 (Gauthier Villars, Paris).
Poincaré, H., 1921, The Foundations of Science (Science and Hypothesis; The Value of Science; Science and Method) transl. G.B. Halstead (Science Press, New York).
Ruelle, D. and F. Takens, 1971, On the nature of turbulence, Commun. Math. Phys. 20, 167–192; 23, 343–344.
Sanders, J.A., 1982, Melnikov's method and averaging, Celestial Mech. 28, 171–181.
Segur, H. and M. Kruskal, 1987, Non-existence of small-amplitude breather solutions in ϕ^4 theory, Phys. Rev. Lett. 58, 747–750.
Shiraiwa, K., 1985, Bibliography for Dynamical Systems, Preprint Series No. 1, Nagoya University Dept. of Mathematics.
Smale, S., 1963, Diffeomorphisms with many periodic points, in: Differential and Combinatorial Topology, ed. S.S. Carins (Princeton Univ. Press, Princeton) pp. 63–80.
Smale, S., 1967 Differentiable dynamical systems. Bull. Am. Math. Soc. 73, 747–817.
Smale, S., 1980, How I got into dynamical systems, in: The Mathematics of Time: Essays on Dynamical Systems, Economic Processes and Related Topics (Springer, New York).
Sternberg, S., 1958, On the structure of local homeomorphisms of Euclidean n-space II. Am. J. Math. 80, 623–631.
Stewart, I.N., 1989, Does God Play Dice? The Mathematics of Chaos (Basil Blackwell, Oxford).
Temam, R., 1988, Infinite-Dimensional Dynamical Systems in Mechanics and Physics, (Springer, New York).
Truesdell, C.A., 1968, Essays in The History of Mechanics (Springer, New York).
Van der Pol, B. and J. Van der Mark, 1927, Frequency demultiplication, Nature 120, 363–364.
Whittaker, E.T., 1959, A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, 4th Ed. (Cambridge Univ. Press, Cambridge).
Wiggins, S., 1988, Global Bifurcations and Chaos: Analytical Methods (Springer, Berlin).
Wiggins, S., 1990, Introduction to Applied Nonlinear Dynamical Systems and Chaos (Springer, Berlin).