

SINGULARITIES IN CLASSICAL MECHANICAL SYSTEMS

Robert L. Devaney *

Singularities in the equations of motion of a classical mechanical system usually play a dominant role in the global phase portrait of the system. By a singularity we mean a point or set of points where the system is undefined, as in the case of a collision between two or more of the particles in the n-body problem. Such singularities often lead to a complicated global orbit structure. Not only do certain solutions tend to run off the phase space, but also nearby solutions tend to behave in an erratic or unpredictable manner. Numerical studies of such systems are often inconclusive because of this erratic behavior. And power series or other analytic techniques often yield only a very local description of solutions near the singularity, one which gives no hint of the global complexity of the system.

Our goal in these notes is to use geometric techniques to gain a complete picture of the local behavior of the solutions near a singularity. Our main tool is a device due to McGehee by which we "blow up" the singular set and replace it with an invariant boundary. The dynamical system extends smoothly (after a scaling of time) over this boundary, and so we get a new flow on an augmented phase space. It turns out that this

* Research partially supported by NSF Grant MCS 79-00430.

new flow restricted to the boundary is extremely simple to understand — usually it is a gradient-like, Morse-Smale flow. So on the boundary of one of the most complicated types of dynamical systems — Hamiltonian systems — we find one of the simplest types of systems. It is this fact that enables us to readily understand the behavior of solutions near the singularity.

It also turns out that this geometric approach to singularities gives us at least a partial grasp on the global orbit structure of systems with singularities. Using symbolic dynamics, one can often associate to each trajectory of the system a doubly infinite sequence of integers in a natural way which adequately describes the qualitative behavior of solutions. This association is often onto the space of all possible such sequences, and so we get a rough idea of the possible complexity of the set of all solution curves in the system.

We will illustrate this idea with two important classical mechanical systems: the anisotropic Kepler problem and the isosceles three body problem in the plane. In fact, examples such as these play an important role throughout these notes. In each of the three major parts, we work mainly with two or three specific examples, leaving the formulation of general theorems to others.

These notes are divided into three parts. In the first we deal with McGehee's technique of blowing up the singularity. We illustrate this technique by describing the flow on the invariant boundary in the case of Newtonian and non-Newtonian

central force problems, as well as in the examples mentioned above. Here we prove that the extended flow is generically Morse-Smale.

In the second part, we switch our emphasis to mappings with singularities, rather than vector fields. We introduce symbolic dynamics via the classical Smale horseshoe mapping, and show how the shift automorphism arises in the well-known Hénon mapping. The baker transformation and another mapping associated to the restricted three body problem provide important examples of mappings with singularities, and we show how symbolic dynamics enters in these cases.

The final part combines the techniques of the previous two to describe at least partially the global orbit structure of the anisotropic Kepler problem and the isosceles three-body problem. The key idea here is the reduction of each differential equation to a Poincaré mapping with singularities on a surface of section. The symbolic dynamics of part two as well as the McGehee method of part one then give an adequate description of this mapping, and thereby the associated phase portrait.

These notes grew out of a series of lectures delivered by the author at the University of Maryland in the spring of 1980. It is a pleasure to thank A. Katok and N. Markley for arranging my visit, and M. Brin, W. Neumann, and M. Paul for many helpful discussions and comments.

Part One: Singularities in Classical Mechanical Systems.

Our goal in this section is to study some of the properties of solutions of Newton's equations

$$M\ddot{q} = -\nabla V(q).$$

Here q is a point in Q , an open subset of \mathbb{R}^n , and the potential energy function $V: Q \rightarrow \mathbb{R}$ is sufficiently smooth. M , the mass matrix, is a diagonal matrix with positive entries m_1, \dots, m_n .

Newton's equations arise in many different areas of mechanics, and we will study in detail several such examples. Most important are:

1. The Kepler or Newtonian central force problem. Here $M = I$ and $V(q) = -1/|q|$.
2. Other central force problems. In this case, $M = I$, but $V(q) = -1/k|q|^k$. All of these systems are integrable systems and their phase portraits are fairly well understood.
3. The anisotropic Kepler problem. Here we keep the Kepler potential $V(q) = -1/|q|$, but make the mass matrix anisotropic, i.e.,

$$M = \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix}$$

with $\mu > 1$.

4. Special cases of the three body problem.
 - A. The collinear (rectilinear) problem with potential

$$V(q_1, q_2, q_3) = -\sum_{i < j \leq 3} m_i m_j / |q_i - q_j|.$$

B. The isosceles problem with mass matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1+\epsilon/\epsilon \end{pmatrix}$$

and potential $V(q_1, q_2) = -1/8q_1 - 2\epsilon/|q|$.

These last two problems are non-integrable and their dynamics are considerably more complicated than the first two. Note that, for each of these examples, V is real analytic and homogeneous of degree $-k$, i.e., $V(\lambda q) = \lambda^{-k}V(q)$. We will in general restrict attention to such potentials, though much that follows can be extended to other classes of potentials.

We also note that, for each of the examples above, the potential suffers a singularity at the origin, and hence the differential equation is undefined there. Our first goal will be to understand how this affects the local behavior of solutions of the system. In particular, for each of these examples, we will examine in this section how the singularity affects orbits which pass close to it. Later, we will describe how the presence of a singularity affects the global phase portrait of the system.

§1.1 Hamiltonian Systems.

Before considering specific examples, we deal with Newton's equations in a general setting. Assume that the potential energy function V is real analytic, homogeneous of degree $-k$, and has an isolated singularity at the origin. Newton's equation

$$M\ddot{q}'' = -\nabla V(q) \quad (1)$$

is more conveniently written as a first order system by introducing the momentum vector $p = Mq'$. Then (1) may be written

$$\begin{aligned} q' &= M^{-1}p \\ p' &= -\nabla V(q). \end{aligned} \tag{2}$$

Let $Q = \mathbb{R}^n - \{0\}$. Q is called the configuration space of the system. (2) is a first order system of differential equations or a vector field on $Q \times \mathbb{R}^n$, the phase space of the system.

This system may be written in Hamiltonian form by introducing the Hamiltonian or total energy function

$$H(q, p) = \frac{1}{2}p^T M^{-1}p + V(q). \tag{3}$$

Then (2) assumes the form

$$\begin{aligned} q' &= \frac{\partial H}{\partial p} \\ p' &= -\frac{\partial H}{\partial q}. \end{aligned} \tag{4}$$

Systems which can be written in this form are called Hamiltonian systems, and they enjoy many special properties. We refer the reader to [AM, Ar] for detailed studies of Hamiltonian systems; in this section we will simply recall a few of their basic properties which we will use over and over again.

We remark that our Hamiltonian function is especially simple; one can generate more complicated Hamiltonian systems by taking an arbitrary function H on \mathbb{R}^{2n} or any other symplectic manifold.

The term $\frac{1}{2}p^T M^{-1}p$ is called the kinetic energy of the system. Since M is positive definite, the kinetic energy is

a positive definite quadratic form in the momentum variables.

It is well known that H is a first integral or constant of the motion for (4). Indeed,

$$\frac{dH}{dt} = \sum_{i=1}^n \frac{\partial H}{\partial q_i} \frac{dp_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dq_i}{dt}$$

which vanishes identically by (4). Hence we can reduce the dimension of the system by one by considering (4) as a vector field on the surface of constant energy (or energy level) $H^{-1}(e)$. When e is a regular value of H , this is a smooth submanifold of $Q \times \mathbb{R}^n$.

Since the kinetic energy is non-negative, it follows from (3) that, on $H^{-1}(e)$, we have $V(q) \leq e$. The corresponding region in Q is called the Hill's region for energy level e . If we take a solution curve $(q(t), p(t))$ of the system on $H^{-1}(e)$, then $q(t)$ must lie for all time in the corresponding Hill's region.

We can now describe $H^{-1}(e)$ topologically as follows. Over each point q^* in the interior of the Hill's region, we have an ellipsoid in the momentum coordinates defined by

$$\frac{1}{2} p^T M^{-1} p = e - V(q^*) > 0. \quad (5)$$

On the boundary $V(q) = e$, the ellipsoid degenerates to a point. So $H^{-1}(e)$ is a "pinched" sphere bundle over the corresponding Hill's region. We visualize $H^{-1}(e)$ as a union of configurations q^* with attached momentum vectors p^* satisfying (5). See Fig. 1.

In many of our examples, the Hill's region is a two dimensional disk minus a point. So in this case, $H^{-1}(e)$ is

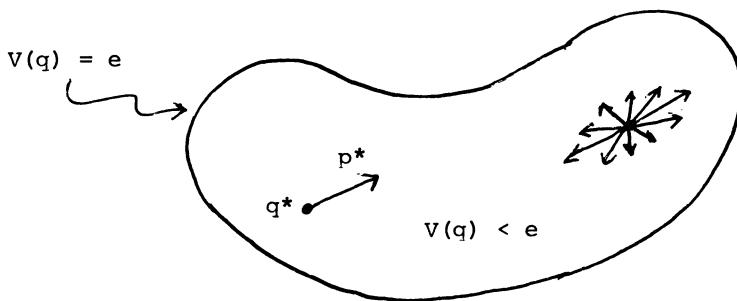


Fig. 1. The Hill's region.

topologically an open solid torus, as one checks easily using the description above.

Note that the projection of a solution curve to the Hill's region is a curve $q(t)$ which may cross itself; at such crossings one has of course different values for the momentum coordinates. Also, solution curves may meet the boundary $V(q) = e$. This boundary is called the zero velocity set, since here the momentum coordinates are all 0. In this case a solution curve $q(t)$ must fall back upon itself and retrace its path in the opposite direction. This follows immediately from the fact that if $(q(t), p(t))$ is a solution of (4), then so too is $(q(-t), -p(-t))$. Systems with this property are called reversible systems.

In particular, we note that if a solution curve has two distinct points of intersection with the zero velocity set, then the corresponding solution curve in phase space is necessarily periodic. These solutions are called symmetric periodic solutions. They are often the easiest types of closed orbits to

find in a reversible system. See Fig. 2.

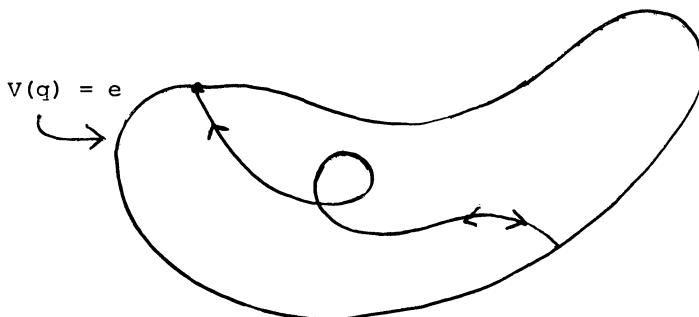


Fig. 2. A symmetric periodic solution in the Hill's region.

§1.2 McGehee Coordinates.

In this section we introduce a remarkable change of coordinates due to R. McGehee [McG 1]. Recall that V is homogeneous of degree $-k$ with an isolated singularity at the origin in configuration space. Since $-\nabla V(q)$ is not defined at $q = 0$, we have a "hole" in phase space at $\{0\} \times \mathbb{R}^n$ where the vector field is undefined. Certain orbits of the system reach this singularity in finite time (the collision orbits), while others begin at the singularity (the ejection orbits). Moreover, the local behavior of the system near $\{0\} \times \mathbb{R}^n$ is usually quite complicated. However, the McGehee coordinates allow us to read off the behavior of these solutions with relative ease.

We introduce these coordinates via a series of three coordinate changes. The first step is to introduce "polar"

coordinates generated by the moment of inertia of the system.

Let

$$\begin{aligned} r &= (q^t M q)^{\frac{1}{2}} \\ s &= r^{-1} q. \end{aligned} \tag{1}$$

The quantity r^2 is called the moment of inertia of the system and s is the configuration. Since $s^t M s = 1$, we think of s as a point in the unit sphere S in \mathbb{R}^n in the metric induced by M , i.e., under the inner product $v_1^t M v_2$. Also define

$$\begin{aligned} y &= s^t p \\ x &= M^{-1} p - y s. \end{aligned} \tag{2}$$

Then y is the radial component of velocity $q' = M^{-1} p$ and x is the component tangent to S . Indeed, we have $s^t M x = 0$, so we think of the pair (s, x) as defining a point in the tangent bundle TS of S .

In these coordinates, the original system becomes

$$\begin{aligned} r' &= y \\ y' &= r^{-1} (x^t M x) + r^{-k-1} k V(s) \\ s' &= r^{-1} x \\ x' &= -r^{-1} (y x + (x^t M x) s) - r^{-k-1} (M^{-1} \nabla V(s) + k V(s) s). \end{aligned} \tag{3}$$

Here we have used the fact that ∇V is homogeneous of degree $-k-1$ so that $\nabla V(rs) = r^{-k-1} V(s)$. Also, Euler's formula gives

$$s^t \nabla V(s) = -k V(s).$$

The system (3) is a real analytic vector field on the manifold $(0, \infty) \times \mathbb{R} \times TS$, and the energy relation gives

$$\frac{1}{2}(x^t M x + y^2) + r^{-k} V(s) = e \quad (4)$$

The system (3) is no longer Hamiltonian, but (4) defines a codimension one invariant set which we continue to call the energy level.

Note that the singularity set for (3) now corresponds to $r = 0$. As $r(t) \rightarrow 0$, we in general have $x^t M x + y^2 \rightarrow \infty$, so the next step is to scale down these components of velocity.

Introduce

$$\begin{aligned} u &= r^{k/2} x \\ v &= r^{k/2} y \end{aligned} \quad (5)$$

The system (3) becomes

$$\begin{aligned} r' &= r^{-k/2} v \\ v' &= r^{-k/2} - \frac{1}{2}(u^t M u + (k/2)v^2 + k V(s)) \\ s' &= r^{-k/2} - \frac{1}{2} u \\ u' &= r^{-k/2} - \frac{1}{2}((k/2 - 1)v u - (u^t M u)s - k V(s)s - \\ &\quad - M^{-1} \nabla V(s)) \end{aligned} \quad (6)$$

This system still has singularities at $r = 0$, but now they can be removed by a change of time scale. Introduce a new time variable τ via

$$\frac{dt}{d\tau} = r^{k/2 + 1} \quad (7)$$

Then (6) becomes

$$\begin{aligned} \dot{r} &= r v \\ \dot{v} &= u^t M u + (k/2)v^2 + k V(s) \\ \dot{s} &= u \end{aligned} \quad (8)$$

$$\dot{u} = (k/2 - 1)vu - (u^t M u)s - kV(s)s - M^{-1}VV(s)$$

where the dot indicates differentiation with respect to τ .

The energy relation goes over to

$$r^k e = \frac{1}{2}(u^t M u + v^2) + V(s). \quad (9)$$

We will study this system in detail in §4, but for now we list some of the most important observations.

1. This is an analytic vector field on $[0, \infty) \times \mathbb{R} \times TS$; that is, the singularities at $r = 0$ have been removed.
2. In their place, the vector field has been extended analytically to the boundary $r = 0$, and this boundary is invariant under the flow, since $\dot{r} = 0$ when $r = 0$.
3. The energy relation also extends to the boundary, giving

$$\frac{1}{2}(u^t M u + v^2) + V(s) = 0 \quad (10)$$

when $r = 0$. In effect, we have glued a boundary onto each energy level, and this boundary is also invariant under the flow. We denote by Λ the boundary of $H^{-1}(e)$ determined by (10); Λ is called the collision manifold of the system.

4. Λ is independent of e ; so each energy surface has the same boundary in $r = 0$.
5. Orbits which previously reached $r = 0$ in finite time are now asymptotic to Λ . And orbits which previously passed close to the singularity now spend

a long time near Λ . Thus the flow on Λ determines the local behavior of solutions near the singularity. In the following sections we will show that in fact the flow on Λ is extremely simple - generically it turns out to be a gradient-like, Morse-Smale flow.

All of the examples we will discuss are systems with two degrees of freedom, i.e., $n = 2$. In this case, the spherical variables (s, u) may be replaced by the more usual polar coordinate θ (in the M-metric) and the component of velocity in the θ direction, which we also call u . The system (8) becomes in this simple case

$$\begin{aligned}\dot{r} &= rv \\ \dot{v} &= u^2 + (k/2)v^2 + kv(\theta) \\ \dot{\theta} &= u \\ \dot{u} &= (k/2 - 1)vu - v'(\theta)\end{aligned}\tag{11}$$

with the energy relation

$$\frac{1}{2}(u^2 + v^2) + v(\theta) = r^k e.\tag{12}$$

We call the variables (r, v, θ, u) McGehee coordinates.

§ 1.3 The Kepler Problem and Other Central Forces.

Before discussing the general properties of the flow on Λ , we pause to describe two simple examples, the Kepler problem and other (non-Newtonian) central force problems. We will concentrate on how the flow on Λ describes the qualitative beha-

avior of orbits which pass close to the singularity.

For the Kepler problem, the potential energy is simply

$$V(q) = -1/|q| \quad (1)$$

and $M = I$. For simplicity, we will restrict to negative energy levels $e < 0$. It follows from (1) that the Hill's regions for negative energy are disks of radius $-1/e$ with the origin removed. Restricted to S , we have $V(0) = -1$, so that the Kepler problem in McGehee coordinates is given by

$$\begin{aligned} \dot{r} &= rv \\ \dot{v} &= \frac{1}{2}v^2 + u^2 - 1 \\ \dot{\theta} &= u \\ \dot{u} &= -\frac{1}{2}vu \end{aligned} \quad (2)$$

with the energy relation

$$\frac{1}{2}(u^2 + v^2) - 1 = re. \quad (3)$$

When $r = 0$, the energy relation shows that Λ is a two dimensional torus in $r = 0$ defined by

$$\begin{aligned} \frac{1}{2}(u^2 + v^2) &= 1 \\ \theta &\text{ arbitrary} \end{aligned} \quad (4)$$

The vector field on Λ is then given by

$$\begin{aligned} \dot{v} &= \frac{1}{2}u^2 \\ \dot{\theta} &= u \\ \dot{u} &= -\frac{1}{2}vu \end{aligned} \quad (5)$$

where we have used the energy relation to simplify \dot{v} . This system is easy to understand. Note first that (5) has rest

points iff $u = 0, v = \pm\sqrt{2}$, θ arbitrary. That is, there are two circles of equilibrium points on Λ . All other solution curves move from the lower circle $v = -\sqrt{2}$ to the upper circle $v = +\sqrt{2}$, since $v > 0$ when $u \neq 0$. In fact, there is a unique unstable manifold associated to the rest point at $\theta = \theta^*$, $v = -\sqrt{2}$, and this unstable manifold has the property that it also forms the stable manifold of the rest point directly "above" it, i.e., at $\theta = \theta^*, v = +\sqrt{2}$. To see this, we introduce the angular variable ψ via

$$u = \sqrt{2} \cos \psi$$

$$v = \sqrt{2} \sin \psi.$$

The system on Λ becomes

$$\dot{\psi} = (1/\sqrt{2}) \cos \psi$$

$$\dot{\theta} = \sqrt{2} \cos \psi$$

so that we have

$$\frac{d\psi}{d\theta} = \frac{1}{2}.$$

The solutions of this vector field are sketched in Fig. 1.

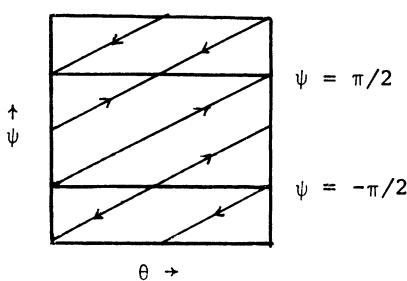


Fig. 1. The flow on Λ for the Kepler problem.

Away from Λ , we also wish to describe the set of orbits which begin and end in collision. For this, we fix $\theta = \theta^*$ and $u = 0$. Then $\dot{\theta} = \dot{u} = 0$, so that the corresponding r, v -plane is invariant under the flow. Restricted to this plane, the system (2) becomes

$$\begin{aligned}\dot{r} &= rv \\ \dot{v} &= \frac{1}{2}v^2 - 1 = re.\end{aligned}$$

This system is easily solved and the phase portrait is sketched in Fig. 2.

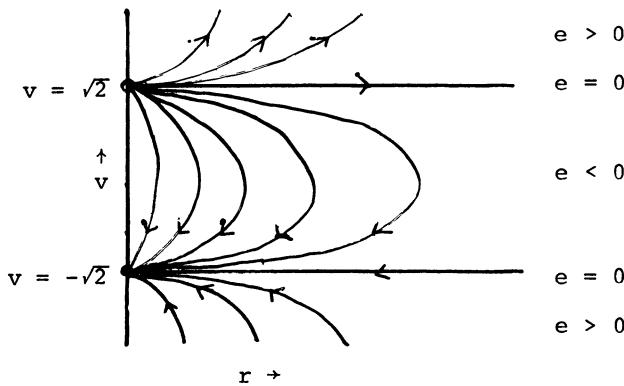


Fig. 2. The flow on an invariant r, v -plane.

For each fixed negative energy level, there is a unique solution which begins and ends at collision with $r = 0$, and which satisfies $\theta = \theta^*$ and $u = 0$ for all time. Such orbits are called homothetic orbits. In configuration space, they simply traverse the ray $\theta = \theta^*$ until meeting the zero velocity curve, whereupon they fall back to collision. See Fig. 3.

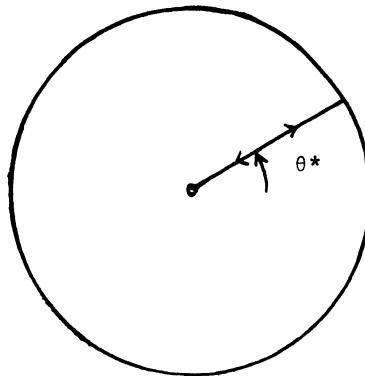


Fig. 3 Homothetic orbits in configuration space.

One may check using the normal hyperbolicity of the circles of equilibria that the homothetic orbits are the only collision and/or ejection orbits for the system. Hence the flow on and near Λ may be pictured as in Fig. 4.

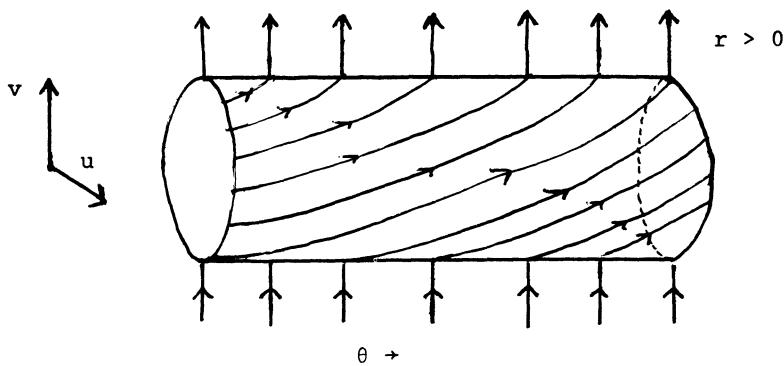


Fig. 4 The flow on and near the collision manifold in the Kepler problem.

We can now read off the local behavior of solutions which pass near the origin. Such an orbit follows a collision orbit until it comes close to Λ . Then it has two choices depending upon which branch of the unstable manifold in Λ it then follows. Each branch, however, forces θ to change by 2π before yielding escape from a neighborhood of collision near an ejection orbit. In one case, θ increases by 2π before leaving a neighborhood of the origin. That is, such solutions circle the origin in a counter-clockwise direction before escaping. In the other case, θ decreases by 2π , forcing solutions to circle the origin clockwise before escaping. See Fig. 5.

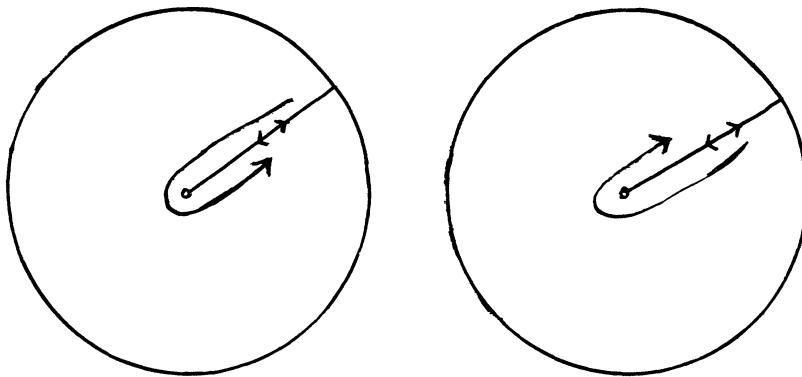


Fig. 5 Local behavior of solutions of the Kepler problem which pass close to collision.

Actually, it is well known that these near-collision orbits lie on ellipses, so the typical near-collision orbit looks like Fig. 6.

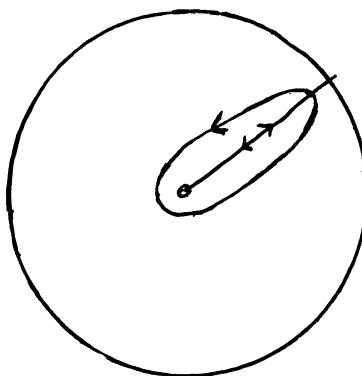


Fig. 6 Near-collision orbits in the Kepler problem.

The same techniques as above also yield a description of collision for other central force potentials, specifically, for those of the form

$$V(q) = -1/k |q|^k \quad k > 0$$

In McGehee coordinates, the system on Λ becomes

$$\begin{aligned}\dot{v} &= (1 - k/2)u^2 \\ \dot{\theta} &= u \\ \dot{u} &= (k/2 - 1)vu\end{aligned}$$

with $\frac{1}{2}(u^2 + v^2) = 1/k$. One checks easily that the circles $u = 0$, $v = \pm\sqrt{2/k}$ are again circles of rest points, and that there is a cylinder of homothetic orbits both coming into and leaving collision. As before, when $e < 0$, these orbits match up exactly, yielding collision/ejection orbits.

The major difference between these systems and the Kepler problem is the non-stationary orbits on Λ . Introducing the

angular variable ψ via

$$u = \sqrt{2/k} \cos \psi$$

$$v = \sqrt{2/k} \sin \psi$$

one finds

$$\frac{d\psi}{d\theta} = 1 - k/2.$$

The resulting phase portrait is different depending on whether $k < 2$, $k = 2$, or $k > 2$. In the first case, $v > 0$, so all orbits still travel from the lower circle of rest points to the upper circle, whereas when $k > 2$, exactly the opposite is true. For $k = 2$, $v = 0$, so all orbits on Λ except the rest points are periodic. Fig. 7 gives a catalogue of these different cases.

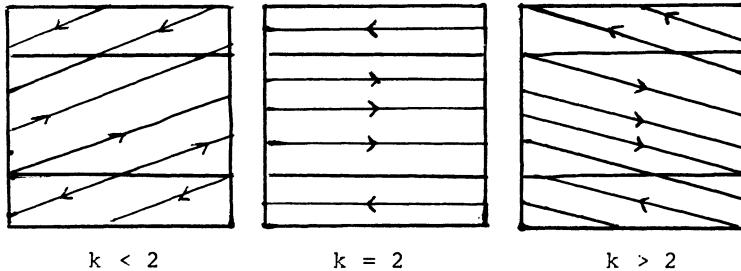


Fig. 7 The flow on the collision manifold for the non-Newtonian central force problems.

When $k > 2$, the lower circle of rest points is now an attractor for the flow off Λ . Hence a large open set of orbits

reach collision in this case. Similarly, a large open set of orbits also begins at collision. We remark that this does not contradict the volume preserving property of the original system, since as the orbits tend to $r = 0$, the velocities become unbounded.

When $k = 2$, we have a transitional case: open sets of orbits still tend to collision.

For $k < 2$, only the homothetic orbits reach $r = 0$, but the local behavior changes dramatically with k . The unstable manifolds at $\theta = \theta^*$, $v = -\sqrt{2/k}$ in Λ do not necessarily join up with the stable manifolds at $\theta = \theta^*$, $v = \sqrt{2/k}$ as was the case in the Kepler problem. Only when we have

$$\frac{d\psi}{d\theta} = 1 - k/2 = 1/2n \quad n = 1, 2, 3\dots$$

do we have this property. That is, if $k = 2 - 1/n$ for a positive integer n , then each branch of the unstable manifold makes n circuits of Λ before rejoining the upper circle of rest points at the same θ -value. When

$$\frac{d\psi}{d\theta} = 1 - k/2 = 1/(2n+1) \quad n = 0, 1, 2\dots$$

or equivalently when $k = 2 - 2/(2n+1)$, the unstable manifolds leaving $\theta = \theta^*$ join up with the stable manifolds at $\theta = \theta^* + \pi$ after making $n+\frac{1}{2}$ circuits. See Fig. 8.

In all other cases, the two branches of the unstable manifolds reach distinct equilibrium points, as in Fig. 9.

The local behavior of solutions may then be described as follows. For $k < 2$, a near-collision orbit makes n revolu-

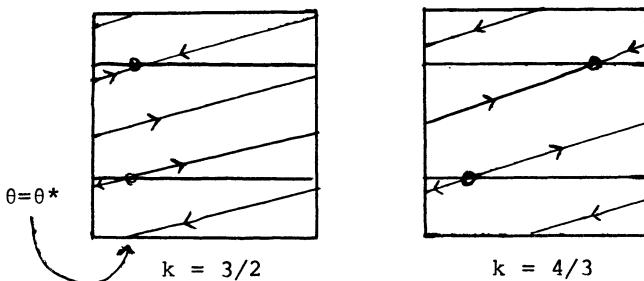


Fig. 8 The flow on Λ in two special cases. The stable and unstable manifolds at $\theta = \theta^*$ match up when $k = 3/2$, while the unstable manifold at θ^* equals the stable manifold at $\theta^* + \pi$ when $k = 4/3$.

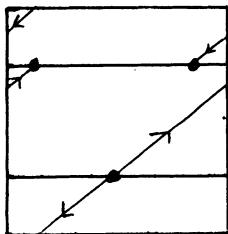


Fig. 9 In this case, the branches of the unstable manifold at θ^* tend to distinct rest points.

tions about $r = 0$ before exiting at an angle which depends on k . In the two special cases depicted in Fig. 8, the orbit either exits in the direction in which it approached collision ($k = 3/2$), or else in exactly the opposite direction ($k = 4/3$). See Fig. 10.

If k is not of one of these two forms, then nearby initial conditions will lead to quite different behavior near collision. This is the basic idea behind Easton's notion of topological regularization [E]. In this case we cannot join orbits coming to collision with orbits leaving collision in a

meaningful way so to make the resulting flow continuous. See Fig. 11.

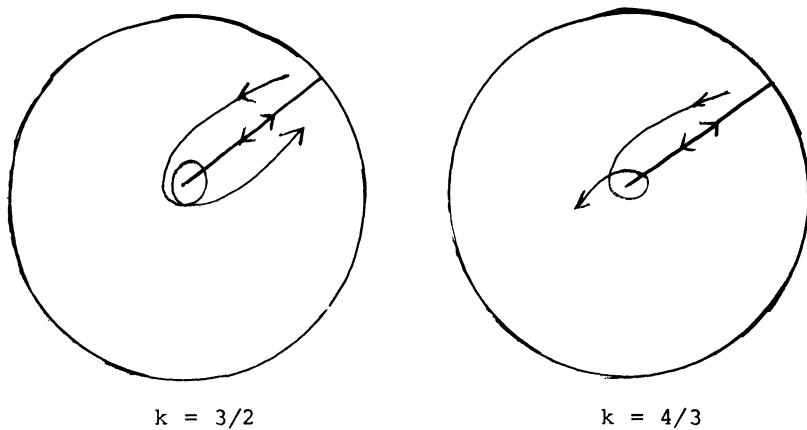


Fig. 10 Behavior of solutions near a homothetic orbit.

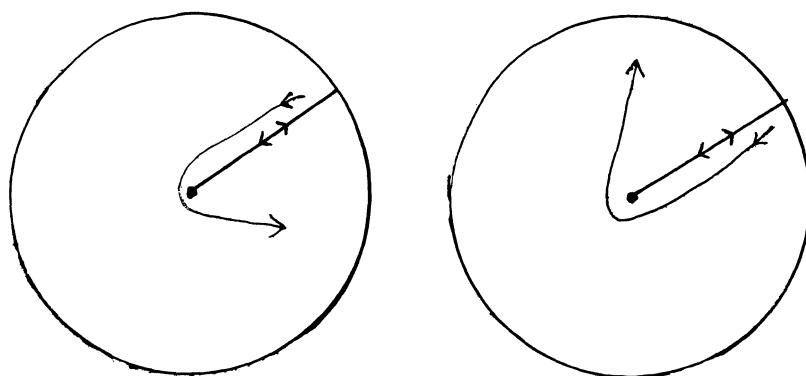


Fig. 11 Behavior of near-collision orbits in the case illustrated in Fig. 9. Note that nearby initial conditions lead to quite different exiting behavior.

§1.4 The Flow on the Collision Manifold

In this section we return to the general setting to describe the behavior of the flow on the collision manifold. Recall that, from §2, in McGehee coordinates, Newton's equations assume the form

$$\begin{aligned}\dot{r} &= rv \\ \dot{v} &= (k/2)v^2 + u^T M u + kV(s) \\ \dot{s} &= u \\ \dot{u} &= (k/2 - 1)vu - (u^T M u)s - kV(s)s - M^{-1}\nabla V(s)\end{aligned}\tag{1}$$

with the energy relation

$$r^k e = \frac{1}{2}(u^T M u + v^2) + V(s).\tag{2}$$

When $r = 0$, the collision manifold Λ is defined by

$$0 = \frac{1}{2}(u^T M u + v^2) + V(s)\tag{3}$$

Recall that V is real analytic, and that $V(s)$ is the restriction of V to the unit sphere S in configuration space.

Proposition 1. If 0 is a regular value of $V(s)$, then Λ is a compact smooth submanifold of the boundary $r = 0$.

Proof. Let $g(v, s, u) = \frac{1}{2}(u^T M u + v^2) + V(s)$ and note that Λ is $g^{-1}(0)$. Λ is clearly closed and bounded. Suppose that 0 is a critical value of g . Then, since $\partial g / \partial u = 0$ and $\partial g / \partial v = 0$, it follows that both $u = 0$ and $v = 0$. Hence $V(s) = 0$ also. But then $dV(s) = 0$, contradicting our assumption.

tion that 0 is a regular value of $V(s)$.

qed

We remark that, in specific examples, one can compute $dV(s)$ explicitly, so that the hypotheses of this proposition are easy to verify in practice. We will call a singularity non-degenerate if $V(s)$ has 0 as a regular value. Clearly, an open and dense set of homogeneous potentials of degree $-k$ have non-degenerate singularities.

We now turn our attention to the flow on Λ in the non-degenerate case. When $r = 0$, (1) reads

$$\begin{aligned}\dot{v} &= (1 - k/2) u^t M u \\ \dot{s} &= u \\ \dot{u} &= (k/2 - 1) v u - (u^t M u) s - k V(s) s - M^{-1} \nabla V(s)\end{aligned}\tag{4}$$

where we have used the energy relation (3).

We remark that the expression $k V(s) s + M^{-1} \nabla V(s)$ is the gradient vector field associated to the restriction of V to S in the metric induced by M . Indeed, in this metric,

$$\langle k V(s) s + M^{-1} \nabla V(s), s \rangle = k V(s) s^t M s + \nabla V(s)^t s = 0$$

by Euler's formula. Furthermore, for any vector w tangent to S at s , we have

$$\langle k V(s) s + M^{-1} \nabla V(s), w \rangle = \nabla V(s)^t w = dV(s)(w)$$

since $s^t M w = 0$. This proves that

$$k V(s) s + M^{-1} \nabla V(s) = \text{grad}_s(V)\tag{5}$$

where $\text{grad}_s(V)$ denotes the gradient vector field of the

restriction of V to S at s .

Now the system (4) has rest points whenever

$$\begin{cases} u = 0 \\ \text{grad}_s(V) = 0 \\ v = \pm\sqrt{-2V(s)} \end{cases} \quad (6)$$

Hence there are two rest points for the flow corresponding to each point s where $\text{grad}_s(V)$ vanishes. Such points are called central configurations. Now $\text{grad}_s(V) = 0$ iff $dV(s) = 0$, so rest points for the flow are in two-to-one correspondence with critical points of the restriction of V to S . We summarize this in a proposition.

Proposition 2. The flow on Λ has a rest point at (v_0, s_0, u_0) iff s_0 is a critical point of the restriction of V to S , $u_0 = 0$, and $v_0 = \pm\sqrt{-2V(s_0)}$.

Recall that a vector field is called gradient-like with respect to a function h if h increases along all non-equilibrium point orbits. For the flow on Λ we have:

Proposition 3. If $k < 2$ (resp. $k > 2$), then the flow on Λ is gradient-like with respect to the v -coordinate (resp. $-v$).

Proof. We have from (4) that

$$\dot{v} = (1 - k/2)u^T M u$$

so that $\dot{v} \neq 0$ when $u \neq 0$. If $u = 0$, we have $\dot{u} = -\text{grad}_s(V)$,

so that $u(\tau)$ is not identically 0 unless $\text{grad}_s(V) = 0$. Consequently, $v(\tau)$ either increases or decreases, depending on whether $k < 2$ or $k > 2$. qed

Gradient-like vector fields on compact manifolds are extremely simple dynamical systems: all non-equilibrium point orbits both begin and end at rest points and there is no non-trivial recurrence.

For the flow on Λ we can in fact say more. Recall that $V(s)$ is called a Morse function if all critical points of $V(s)$ are non-degenerate, i.e., $d^2V(s)$ is a non-degenerate bilinear form at each critical point. On the other hand, a rest point p for the flow on Λ is hyperbolic if all of the eigenvalues of the linearization of the system at p have non-zero real parts. These two notions are linked by the following proposition.

Proposition 4. Suppose $k \neq 2$. If $V(s)$ is a Morse function, then all of the rest points for the flow on Λ are hyperbolic.

Proof. From (1), one computes that the linearization of the flow at a rest point with $r = 0$, $u = 0$, $dV(s_0) = 0$, and $v_0 = \pm\sqrt{-2V(s_0)}$ is given by the matrix

$$A = \begin{pmatrix} v_0 & 0 & 0 & 0 \\ 0 & kv_0 & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & -B & (k/2 - 1)v_0 I \end{pmatrix} \quad (7)$$

Here I is the $(n-1) \times (n-1)$ identity matrix and B represents the linearization of the flow of $\text{grad}_s(v)$ at s_0 . Restricting to an energy surface, it follows easily that the eigenvalues of the linearization of the flow are v_0 together with the eigenvalues of the submatrix

$$A' = \begin{pmatrix} 0 & I \\ -B & (k/2 - 1)v_0 I \end{pmatrix}$$

These eigenvalues are supplied by the following lemma.

Lemma 5. Suppose λ is an eigenvalue of B . Then

$$\zeta^\pm = \frac{1}{2}(\mu \pm \sqrt{\mu^2 - 4\lambda})$$

is an eigenvalue of A' , where $\mu = (k/2 - 1)v_0$.

Proof. Suppose x is an eigenvector of B associated to the eigenvalue λ . Then the vector $(x, \alpha x)$ is an eigenvector for A' iff

$$\alpha^2 - \mu\alpha + \lambda = 0$$

iff $\alpha = \zeta^\pm$. This completes the proof of the lemma.

If $V(s)$ is a Morse function and $k \neq 2$, then all of the eigenvalues of B are real and non-zero, and so all of the eigenvalues of A' have non-zero real part. This follows since, in the non-degenerate case, $v_0 \neq 0$. qed

Hence for each rest point p in Λ one has a stable

and unstable manifold denoted by $w^s(p)$ and $w^u(p)$. One can in fact relate the dimensions of these manifolds to the index of the corresponding critical point and $\text{sgn } v_0$ using Lemma 5. See [De 1] for details. Also, one can prove that, generically, the stable and unstable manifolds of all equilibria meet transversely in Λ . By generically we mean within the set of potentials that are homogeneous of degree $-k$. We again refer to [De 1] for the proof.

Flows on Λ having the above properties are called Morse-Smale flows. More precisely, a flow on a manifold is Morse-Smale if

1. The non-wandering set consists of a finite number of hyperbolic rest points and closed orbits.
2. All stable and unstable manifolds meet transversely.

In our case, the flow on Λ is gradient-like, so that there are no closed orbits on Λ , only rest points. In such a case we call the singularity a Morse-Smale singularity. The central force problems of the previous section do not have Morse-Smale singularities, since the rest points in Λ are not hyperbolic. However, the examples which follow are Morse-Smale and, as we have shown above, are more typical.

To summarize this section, we have shown that

Theorem 6. An open and dense set of potentials which are homogeneous of degree $-k$ and which have isolated singularities at the origin have only non-degenerate, Morse-Smale singularities.

§1.5 Collision Orbits.

We now discuss what happens near the collision manifold in the case of a Morse-Smale singularity. Let C denote the set of solutions which end in collision with the singularity, and let E denote the set of ejection orbits, i.e., those orbits which begin at the singularity.

Proposition 1. Suppose there is a Morse-Smale singularity at the origin. Then both C and E consist of a finite union of submanifolds, one for each central configuration.

Proof. Let s_0 be a central configuration, i.e., a critical point of $V(s)$. So the point $p = (v_0, s_0, 0)$ is a hyperbolic rest point for the flow on Λ . We claim that p is a hyperbolic rest point for the entire flow on an energy surface. Using the results of the previous section and particularly the linearization matrix (7), one checks easily that v_0 is an eigenvalue for the linearization whose associated eigenvector is transverse to Λ . Now $v_0 = \pm\sqrt{-2V(s_0)}$, so that p is a hyperbolic rest point in $H^{-1}(e)$, not just on the boundary Λ , since $V(s_0) \neq 0$ in the non-degenerate case. If $v_0 < 0$, we thus have a stable manifold at p in $H^{-1}(e) - \Lambda$ which is one dimension larger than the dimension of the stable manifold at p in Λ . These are the collision orbits which asymptotically attain the configuration s_0 . When $v_0 > 0$, we similarly find a submanifold of ejection orbits in $H^{-1}(e) - \Lambda$ given by the corresponding unstable manifold. Since the flow on Λ is gradient-like, it follows

easily that in fact all collision and ejection orbits lie in such a stable or unstable manifold. And since there are only a finite number of critical points for $V(s)$, the result follows.

qed

We remark that the above proof shows that corresponding to any central configuration s_0 , there are two invariant manifolds. One consists of collision orbits asymptotic to the rest point $(v_0, s_0, 0)$ with $v_0 < 0$, and the other consists of ejection orbits emanating from $(-v_0, s_0, 0)$.

As in the central force potentials, there are special collision and ejection orbits which are associated to each central configuration. These are the homothetic orbits given as follows. From equation (1) of §1.4, it follows that if s_0 is a central configuration and $u = 0$, then $\dot{s} = 0$ and $\dot{u} = 0$. Hence the r, v -plane defined by $s = s_0$, $u = 0$ is invariant. As in the Kepler problem, the vector field on this plane is given by

$$\begin{aligned}\dot{r} &= rv \\ \dot{v} &= (k/2)v^2 + kV(s_0) = r^k e.\end{aligned}$$

The phase portrait in the r, v -plane is sketched in Fig. 1. In particular, there is a unique homothetic orbit in each negative energy level which begins and ends in collision and which projects to configuration space along the ray $s = s_0$. These homothetic orbits can therefore be interpreted as heteroclinic orbits connecting distinct rest points. As such, they play a key role in the global dynamics of the examples discussed below.

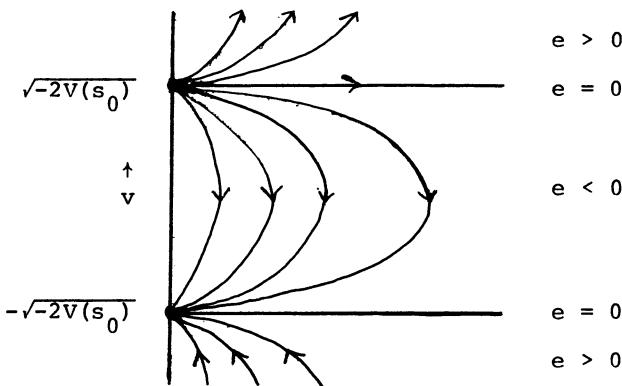


Fig. 1 The flow on an invariant r, v -plane. Each orbit is a homothetic orbit.

§1.6 The Anisotropic Kepler Problem

Our goal in this section is to describe the dynamics of a classical mechanical system which is much more complicated than the previous examples. This is the anisotropic Kepler problem introduced by Gutzwiller in [Gu 1] to model certain quantum mechanical systems. The system is actually a one parameter family of Hamiltonian systems, depending analytically on a parameter $\mu \geq 1$. When $\mu = 1$, we have the usual Kepler problem discussed in §1.3. As μ increases, the system becomes more and more complicated until, for $\mu > 9/8$, numerical evidence indicates that the system is ergodic. We will give partial results in this direction in the third part of these notes.

For $\mu \geq 1$, the anisotropic Kepler problem is described by the equations

$$\begin{aligned}\ddot{q}_1 &= -\mu q_1 / |q|^3 \\ \ddot{q}_2 &= -q_2 / |q|^3\end{aligned}$$

This system differs from the central force problems in that the force vector is directed more toward the q_2 -axis, forcing the solution curves to cross that axis more often than the q_1 -axis. Introducing the mass matrix

$$M^{-1} = \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix}$$

and the potential energy $V(q) = -1/|q|$, one may write this system in Hamiltonian form

$$\begin{aligned}\dot{q} &= M^{-1} p \\ \dot{p} &= -\nabla V(q)\end{aligned}\tag{1}$$

where the momentum vector p is defined by the first equation. The total energy function is

$$H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q).$$

For negative energy, the Hill's region is the disk given by $|q| \leq -1/e$.

We remark that an equivalent system results if one uses the Hamiltonian

$$H'(q, p) = \frac{1}{2} |p|^2 - \frac{1}{\sqrt{\mu q_1^2 + q_2^2}}.$$

Here the potential rather than the kinetic energy is anisotropic.

In McGehee coordinates, (1) goes over to

$$\begin{aligned}
 \dot{r} &= rv \\
 \dot{v} &= u^2 + \frac{1}{2}v^2 + v(\theta) \\
 \dot{\theta} &= u \\
 \dot{u} &= -\frac{1}{2}v u - v'(\theta)
 \end{aligned} \tag{2}$$

where the potential $v(\theta)$ is given by

$$v(\theta) = \frac{-1}{\sqrt{\mu \cos^2 \theta + \sin^2 \theta}} \tag{3}$$

$v(\theta)$ has two non-degenerate maxima at $\theta = 0$ and $\theta = \pi$, and two non-degenerate minima at $\theta = \pi/2$ and $3\pi/2$. The graph of $v(\theta)$ is given in Fig. 1.

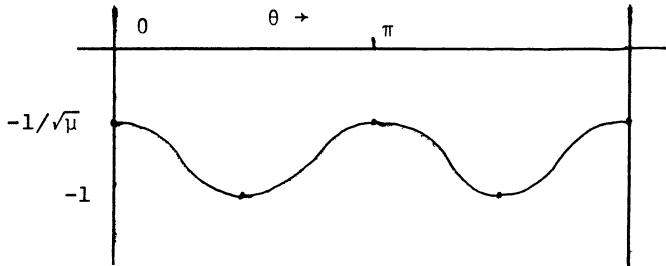


Fig. 1 The graph of $v(\theta)$.

The collision manifold Λ is determined from the energy relation

$$re = \frac{1}{2}(u^2 + v^2) + v(\theta) \tag{4}$$

and is sketched in Fig. 2. When $r = 0$, this equation deter-

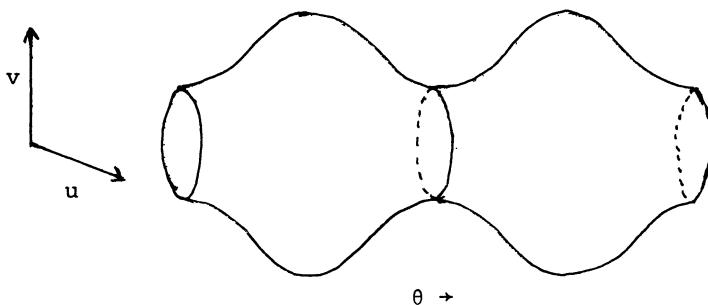


Fig. 2 The collision manifold for the anisotropic Kepler problem.

mimes a torus.

From the results of §4, there are 8 rest points for the flow on Λ : two sinks, two sources, and four saddles. The eigenvalues of the linearization at a critical point of the form $(v_0, \theta_0, 0)$ are given by

$$-\frac{1}{4}v_0 \pm \frac{1}{2}\sqrt{\frac{1}{4}v_0^2 - 4v''(\theta_0)}$$

where $v_0 = \pm\sqrt{-2V(\theta_0)}$. Hence the sinks and sources occur when $V''(\theta_0) > 0$, i.e., at $\theta_0 = \pi/2$ and $3\pi/2$. At these points we have $v(\theta_0) = -1$ and $v''(\theta_0) = (\mu-1)/2$, so that the radical in the above expression reduces to $\sqrt{(9-8\mu)}$. Hence the eigenvalues at the sinks and sources are complex when $\mu > 9/8$. This implies that orbits spiral into and away from the sinks and sources, and, as we show below, this causes nearby solutions to behave quite erratically.

The ultimate behavior of the stable and unstable manifolds of the saddles has been studied by Gutzwiller [Gu 1,2]. Briefly, for all $\mu > 1$, he finds that each branch of one of these invariant manifolds is asymptotic to a distinct sink or source, after

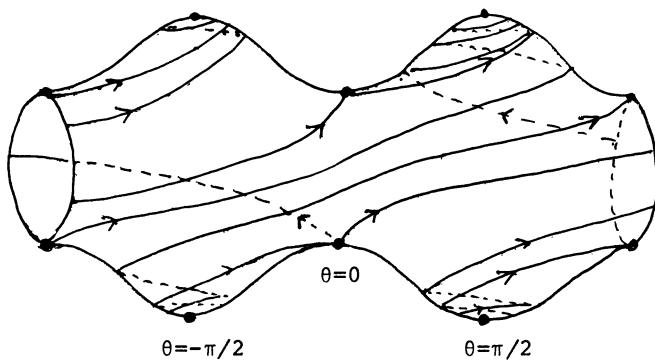


Fig. 3 The stable and unstable manifolds of the saddles. The branches of the invariant manifolds on the backside of Λ can be obtained by the reflection $v \rightarrow -v$, $u \rightarrow -u$.

θ increases or decreases by $3\pi/2$ or $\pi/2$. That is, the two branches of the stable manifold of any saddle emanate from different sources, while the two branches of the unstable manifold die in different sinks.

From §5, we also have homothetic orbits traveling along each ray $\theta = 0, \pi/2, \pi$, and $3\pi/2$. For negative energy, these solutions are bounded and connect distinct rest points in Λ . For $\mu > 9/8$, solutions near these orbits behave quite erratically. Consider an orbit near the collision orbit on $\theta = 0$. The collision orbits along $\theta = \pi$ are handled completely analogously. After passing close to the rest point in Λ , the orbit follows one of the two branches of the unstable manifold emanating from the rest point. At this rest point, $v_0 < 0$, so from Fig. 3, along one of these branches, θ decreases by $3\pi/2$, while along the other, θ increases by $3\pi/2$. Thus we have

two distinct types of qualitative behavior for nearby solutions, depending on which branch the solution follows.

Now after passing close to collision, each solution leaves a neighborhood of the origin spiralling about the corresponding ray $\theta = \pm\pi/2$, at least when $\mu > 9/8$. See Fig. 4.

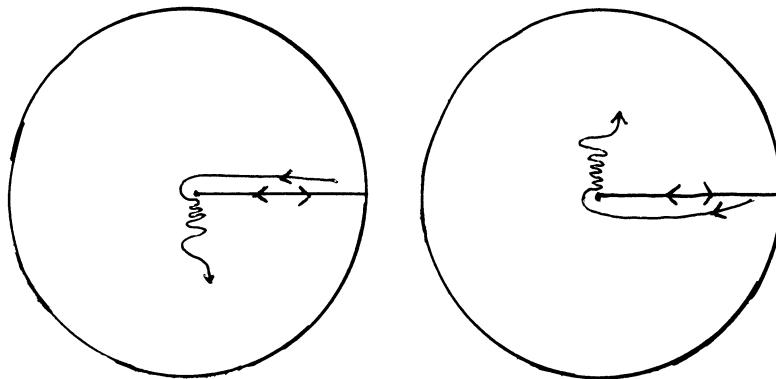


Fig. 4 The qualitative behavior of solutions near the homothetic orbit along $\theta = 0$.

Thus the flow on the collision manifold completely determines the qualitative behavior of orbits close to the collision orbit on $\theta = 0$. One easily makes a similar analysis for the other homothetic orbits.

We also wish to point out how the flow on the collision manifold enables us to find infinitely many distinct collision/ejection and periodic solutions when $\mu > 9/8$. Consider the piece of Λ near the rest points at $\theta_0 = 0$ and $\pi/2$ with $v_0 > 0$. Erect an annular surface of section Λ for the flow defined by $v = v^*$ with $\sqrt{-2v(0)} < v^* < \sqrt{-2v(\pi/2)}$. For r sufficiently small, Λ is transverse to the flow.

Now one branch of the unstable manifold at $\theta = 0$ in Λ crosses A at a point, say p^* . From the results of §1.5, the entire unstable manifold is two dimensional, and it follows that it must meet A transversely (at least near p^*) in a smooth curve γ tending to p^* . Only p^* lies in Λ , however.

Now the homothetic orbit leaving $\theta = \pi/2$, $v_0 > 0$ reaches the zero velocity set Z at a point which we call q^* . Consider a sufficiently small neighborhood N of q^* in Z . Under backwards time, the orbit through each point in N passes close to the rest point and then moves on to meet A . One checks easily that the trace of these points of intersection with A form a spiral winding down to the circular intersection of A with Λ . In particular, this spiral meets γ at infinitely many distinct points, say p_i , $i = 1, 2, \dots$. The orbit through each p_i is an ejection orbit which also meets the oval of zero velocity. By symmetry, such an orbit must fall back upon itself and is therefore also a collision orbit at $\theta = 0$. See Fig. 5.

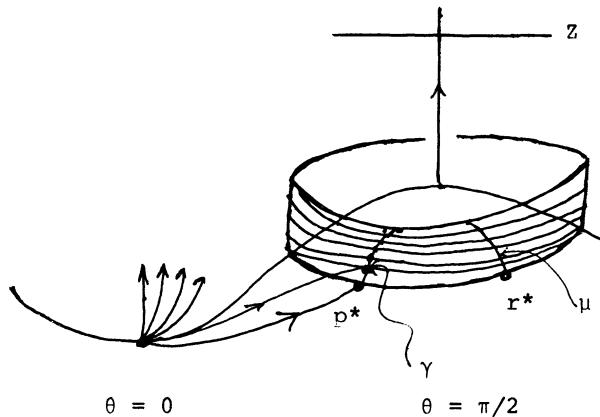


Fig. 5 The proof of Proposition 1.

We have thus proven:

Proposition 1. There exist infinitely many distinct collision/ejection orbits in the anisotropic Kepler problem on each negative energy level when $\mu > 9/8$.

In Fig. 6 we have sketched several of these collision/ejection orbits. Note that they are characterized by a rapid oscillation about the q_2 -axis.

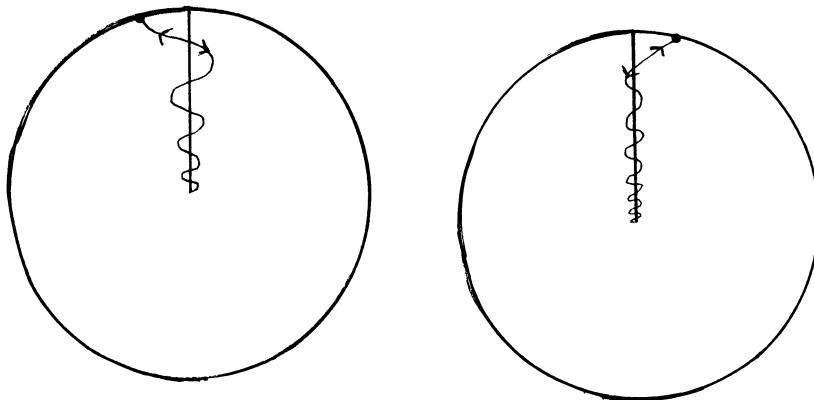


Fig. 6 Collision/ejection orbits in the anisotropic Kepler problem.

We conclude this section by proving the existence of infinitely many symmetric closed orbits for the anisotropic problem, again when $\mu > 9/8$. The homothetic orbit along $\theta = \pi$ meets the zero velocity set at a point s^* ; let P be a small neighborhood of s^* in Z . It is known that Z is transverse to the stable manifold containing s^* . See [De 3].

This time we follow P forward in time under the flow. The orbit of each point in P comes close to the saddle in Λ at $\theta = \pi$ and then follows one of the branches of the unstable manifold. One of these branches meets the annular surface of section at a point r^* , and one may check easily that there is an arc of points μ in Λ limiting at r^* which comes from P . See Fig. 5. Now recall that there is a spiralling curve in Λ which comes from the oval of zero velocity. As before, this spiral intersects μ in infinitely many points. These points lie on orbits which meet the zero velocity set at two distinct points, and by our remarks in §1.1, therefore lie on symmetric closed orbits. We have therefore established:

Proposition 2. There exist infinitely many distinct symmetric closed orbits in each negative energy level when $\mu > 9/8$.

Some of these solutions are sketched in Fig. 7.

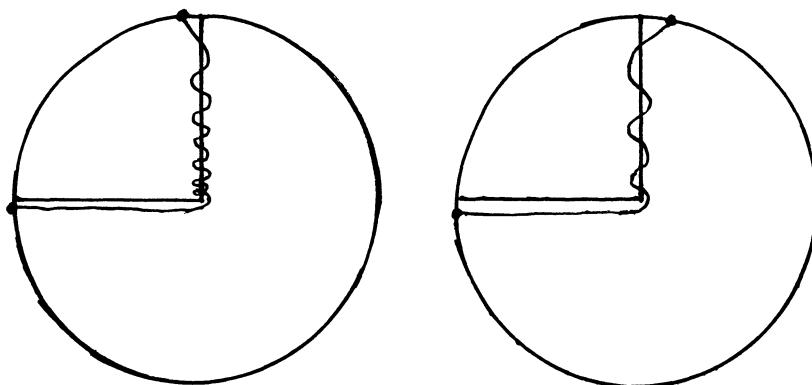


Fig. 7 Symmetric periodic solutions on a negative energy level in the anisotropic Kepler problem.

Again we note that these orbits are characterized by a rapid oscillation about the q_2 -axis, followed by at most one crossing of the q_1 -axis. The number of oscillations about the q_2 -axis may be arbitrarily large. Hence these periodic solutions are quite different from the periodic solutions which occur on negative energy levels for the Kepler problem. Compare Fig. 6 in §1.3. Also, the existence of such orbits suggests the possibility of codifying closed orbits of the anisotropic problem by listing the number of crossings of the q_2 -axis in between successive crossings of the q_1 -axis. This is an idea of Gutzwiller [Gu 2] which motivates the introduction of symbolic dynamics into the problem and which we will treat in much more detail in part 3.

§1.7 The Planar Isosceles Three Body Problem.

In this section, we discuss another example of a non-integrable classical mechanical system, a special case of the three body problem. Our aim here is to show how the techniques of the previous sections allow us to recover classical results of Euler, Lagrange, Siegel, and others relatively easily, as well as to describe the local behavior of near-collision orbits relatively simply.

This problem is remarkably similar to the anisotropic Kepler problem in certain respects, as we will show later. One major difference is that the singularity at the origin is not isolated as in the previous examples. Presumably, our remedy

for this situation extends to other examples with non-isolated singularities.

To state the problem, we take three point masses $m_1 = m_2$ and $m_3 = \epsilon$. Without loss of generality, we may assume that $m_1 = m_2 = 1$. Suppose m_1 and m_2 are positioned symmetrically with respect to the y -axis, and with symmetric initial velocities. And suppose also that the third mass lies on the y -axis with initial velocity parallel to the axis. See Fig. 1.

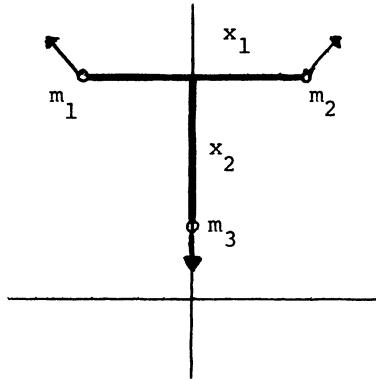


Fig. 1 Jacobi coordinates in the isosceles problem.

If q_i denotes the position of the i^{th} mass, then the differential equation may be written

$$m_i q_i'' = -\nabla_i V$$

where

$$V(q_1, q_2, q_3) = - \sum_{k < j} m_i m_j / |q_i - q_j| .$$

We may make several standard reductions to reduce the number of degrees of freedom in the problem to two. First, by symmetry, if we know $q_1(t)$, then $q_2(t)$ is determined, since the particles always remain in a (possibly degenerate) isosceles triangle configuration. Secondly, we may fix the center of mass at the origin since

$$\sum_i m_i q_i'' = 0$$

and the equations of motion remain unchanged. Hence the positions of all three particles may be determined from Jacobi coordinates x_1, x_2 where $x_1 \geq 0$ denotes the distance between m_1 and m_2 , and x_2 the signed distance between m_3 and the center of mass of m_1 and m_2 . We take $x_2 > 0$ when m_3 lies above the other masses. See Fig. 1.

In these coordinates, the differential equation is given by

$$\begin{aligned}x_1'' &= -2/x_1 - 8\epsilon x_1/(x_1^2 + 4x_2^2)^{3/2} \\x_2'' &= -8(2 + \epsilon)/(x_1^2 + 4x_2^2)^{3/2}.\end{aligned}$$

We refer to Pollard's book [Po] for the derivation of these equations. Note that these equations are singular when $x_1 = 0$, which corresponds to a double collision between m_1 and m_2 , and also when $x_1 = x_2 = 0$, which corresponds to a triple collision or total collapse of the system. Hence this system differs from the previous examples in that the singularity is not isolated.

We first remove the singularity at the origin. Toward that end, we introduce the momenta

$$\begin{aligned} p_1 &= \frac{1}{2}x'_1 \\ p_2 &= (2\epsilon/2+\epsilon)x'_2 \end{aligned}$$

Then the system may be written in Hamiltonian form with the Hamiltonian

$$H = \frac{1}{2}p^T M^{-1} p + V(x)$$

where

$$M^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & (2+\epsilon)/2\epsilon \end{pmatrix}$$

and

$$V(x_1, x_2) = -\frac{1}{x_1} - \frac{4\epsilon}{\sqrt{x_1^2 + 4x_2^2}} \quad (1)$$

The Hill's region for negative energy is sketched in Fig. 2.

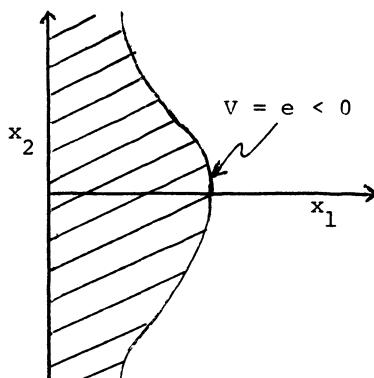


Fig. 2 The Hill's region for negative energy in the isosceles problem.

In McGehee coordinates the equations become

$$\begin{aligned}\dot{r} &= rv \\ \dot{v} &= \frac{1}{2}v^2 + u^2 + v(\theta) \\ \dot{\theta} &= u \\ \dot{u} &= -\frac{1}{2}vu - v'(\theta)\end{aligned}\tag{2}$$

where

$$v(\theta) = -\frac{1}{\sqrt{2\cos\theta}} - \frac{4\varepsilon^{3/2}}{\sqrt{2\varepsilon + 4\sin^2\theta}}\tag{3}$$

is defined for $-\pi/2 < \theta < \pi/2$. The graph of $v(\theta)$ is given in Fig. 3.

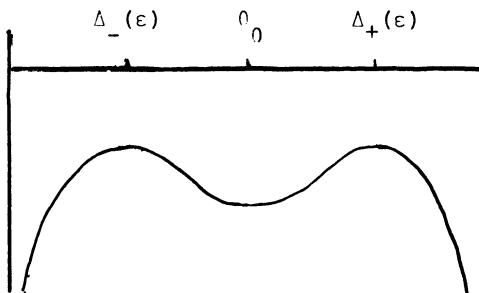


Fig. 3 The graph of $v(\theta)$ in the isosceles problem.

There are three non-degenerate critical points: a local minimum at $\theta = 0$, and two maxima at $\Delta_{\pm} = \arcsin(\sqrt{3\varepsilon}/(2+4\varepsilon))$

We will interpret the flow on the triple collision manifold later. First, however, we will eliminate the singularities in the above equations when $\theta = \pm\pi/2$. This will be achieved by simultaneously regularizing each double collision

via a procedure due to Sundman [Su]. This necessitates a further change of variables. We remark that McGehee carried out a similar procedure for the collinear three body problem [McG 1].

Let $W(0) = -\cos\theta V(0)$. One checks easily that $W(\theta)$ is a positive, real analytic function on $[-\pi/2, \pi/2]$. Now introduce a new tangential component of velocity

$$w = u \cos\theta / \sqrt{W(\theta)}$$

as well as a further change of time scale

$$\frac{dt}{ds} = \frac{\cos\theta}{\sqrt{W(\theta)}}.$$

The differential equation becomes

$$\begin{aligned} \dot{r} &= \frac{\cos\theta}{\sqrt{W(\theta)}} rv \\ \dot{v} &= \sqrt{W(\theta)} [1 - \frac{\cos\theta}{2W(\theta)}(v^2 - 4re)] \\ \dot{\theta} &= w \\ \dot{w} &= \sin\theta [-1 + \frac{\cos\theta}{W(\theta)}(v^2 - 2re)] - \frac{vv\cos\theta}{2\sqrt{W(\theta)}} + \\ &\quad + \frac{W'(\theta)}{W(\theta)} (\cos\theta - w^2/2) \end{aligned} \tag{4}$$

with energy relation

$$w^2/2\cos\theta - 1 = \frac{\cos\theta}{W(\theta)} (re - v^2/2). \tag{5}$$

Here the dots indicate differentiation with respect to s . This gives a real analytic vector field on $[0, \infty) \times \mathbb{R} \times [-\pi/2, \pi/2] \times \mathbb{R}$. Note that the vector field does not vanish when $\theta = \pm\pi/2$. Hence we have extended solution curves through double collision via an "elastic bounce".

The triple collision manifold is given by

$$w^2/2 + v^2 \cos^2 \theta / w(\theta) = \cos \theta . \quad (6)$$

When $\theta = \pm\pi/2$, this gives $w = 0$, v arbitrary, i.e., a straight line. For all other values of θ , this equation determines an ellipse in the corresponding v, w -plane. The triple collision manifold can then be sketched as in Fig. 4.

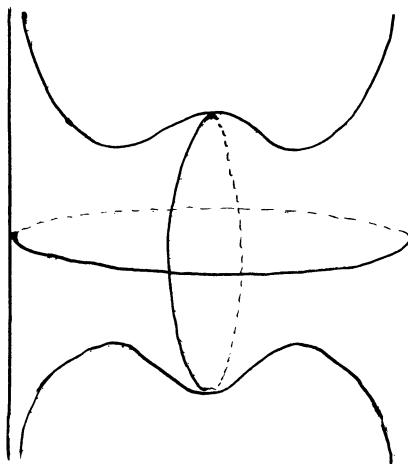


Fig. 4 The triple collision manifold in the isosceles problem after regularization of double collisions.

Note that, in this case, the triple collision manifold is non-compact; topologically, it is a sphere with four points removed.

Using the results of §1.4, one checks easily that there are four saddle points for the flow corresponding to the critical points at $\theta = \Delta_{\pm}$, and a sink and source at $\theta = 0$. Moreover, a simple computation shows that the eigenvalues at the sink and source are complex when $\epsilon < 55/4$. See Fig. 5.

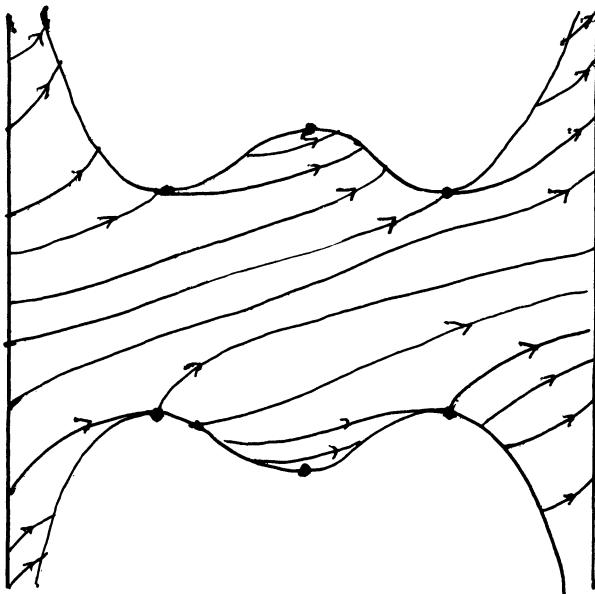


Fig. 5 The flow on the triple collision manifold.

The rest point structure allows us to prove several classical theorems quite easily:

Theorem 1 (Siegel) The set of solutions of the isosceles three body problem which begin or end in triple collision is given by a union of lower dimensional submanifolds.

Proof. There are two dimensional stable manifolds in each energy level asymptotic to the saddle points with $v < 0$, and a one dimensional stable manifold at the source in Λ . All of these orbits are collision orbits. We claim that, in fact, these are the only orbits which end in triple collision. For any orbit which ends in triple collision satisfies $r \rightarrow 0$.

The ω -limit set of such an orbit must then lie in $r = 0$, and one checks easily that it must be a compact set in Λ . Since the flow on Λ is gradient-like, it follows that the ω -limit set must be a rest point.

The proof for the ejection orbits is essentially the same. qed

The relatively simple structure of the flow on the triple collision manifold also allows us to interpret what happens to the three masses as they approach triple collision.

Theorem 2 (Sundman) As they approach triple collision, the three masses tend to either a straight line or else to the vertices of an equilateral triangle.

Proof. First suppose the orbit is asymptotic to the rest point at $\theta = 0$. In terms of the original Jacobi coordinates, x_1, x_2 , it follows that

$$x_2/x_1 = (2+\varepsilon)\sin\theta/4\varepsilon\cos\theta \rightarrow 0$$

so that the particles tend toward a horizontal line. On the other hand, if the orbit is asymptotic to one of the rest points with $\theta = \Delta_{\pm}$, then we find

$$x_2^2/x_1^2 = (2+\varepsilon)\sin^2\theta/4\varepsilon\cos^2\theta \rightarrow 3/4$$

so that the particles tend to form an equilateral triangle in the limit. qed

In the classical literature on the subject, these asymptotic configurations are called central configurations, motivating our more general use of the term.

We remark that, since the corresponding invariant manifolds are one-dimensional, there is a unique orbit tending to and away from the rest point at $\theta = 0$ (the straight line configuration). This must then be the homothetic solution discussed in §1.5. Since $\theta = 0$ along this solution, it follows that x_2 is identically equal to zero along the solution also. Hence the third mass must remain fixed at the center of mass of the binary system. This special solution was first found by Euler [Eu] in the more general setting of the full three body problem.

There are also a pair of homothetic solutions satisfying $\theta = \Delta_{\pm}$. For these special solutions, the particles remain for all time at the vertices of an equilateral triangle. Lagrange [La] was the first to find these solutions. We have sketched the projections of these solutions to configuration space in Fig. 6 for negative energy.

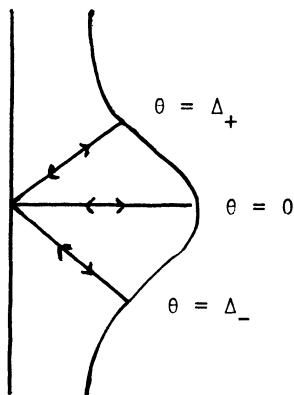


Fig. 6 The projection to the Hill's region of the special homothetic orbits discovered by Euler and Lagrange.

To study the behavior of solutions passing close to triple collision, we need some information about the behavior of the stable and unstable manifolds of the saddle points in Λ . This has been studied by Simo [Sim]. It turns out that there are many possibilities, depending on ϵ . We will discuss only the case of $0 < \epsilon < \epsilon_0$, for some small but positive ϵ_0 determined by Simo. For this case, the invariant manifolds are depicted in Fig. 7.

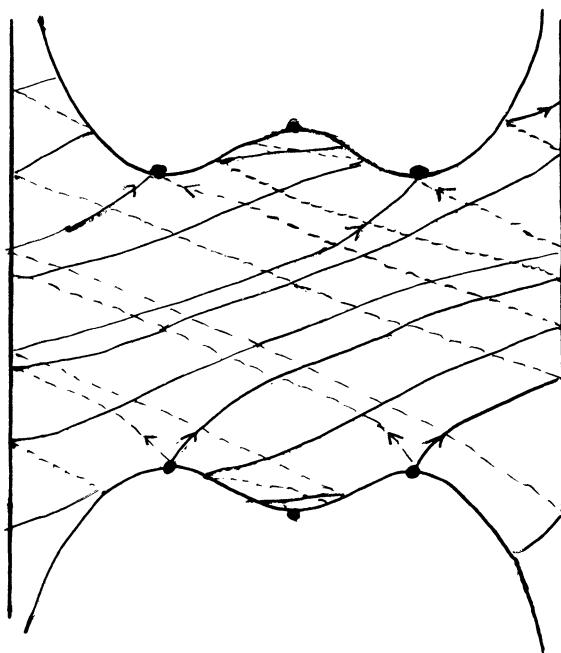


Fig. 7 The behavior of the stable and unstable manifolds of the saddle points in the isosceles problem.

In Fig. 7, note that one branch of each unstable manifold dies in the sink, while the other branch runs up the "arm" of the triple collision manifold. The behavior of the

stable manifolds at the upper saddle points can be obtained via the reflection $v \rightarrow -v$, $\theta \rightarrow \theta$, $w \rightarrow -w$ which maps the unstable manifolds (resp. stable manifolds) of the rest points with $v < 0$ to the stable (resp. unstable) manifolds of the rest points with $v > 0$. Fig. 7 displays most of this information.

When ϵ is sufficiently small, the behavior of a solution curve which passes close to triple collision is easily decided. Consider a solution curve near an orbit which reaches triple collision at $\theta = \Delta_+$. If this orbit is close enough to the triple collision orbit, there are precisely two possibilities for its eventual behavior: either the orbit follows the branch of the unstable manifold which dies in the sink, or else it follows the branch which runs up the arm with $\theta = -\pi/2$. In the first case the orbit leaves a neighborhood of $r = 0$ near the line $\theta = 0$. Note that because of the spiralling near the sink, the orbit oscillates rapidly about $\theta = 0$, i.e., the third mass is left oscillating about the center of mass of the binary system, which in turn separates roughly along the x -axis.

In the second case, one finds dramatically different behavior. One can prove that this branch of the unstable manifold meets the line $\theta = -\pi/2$ infinitely often as it runs up the arm of the collision manifold. One may also check that nearby solutions cross $\theta = -\pi/2$ an arbitrarily large number of times. Each crossing corresponds to a binary collision between m_1 and m_2 .

Finally note that orbits following this branch of the unstable manifold exit from a neighborhood of triple collision

with an arbitrarily large value for v , the radial component of velocity. This means that m_3 is separating from the binary system with arbitrarily large velocity. The third mass travels down the y -axis, while the binary pair moves up the axis in a tight oscillation featuring many double collisions. One may check that this large value of v leads to escape orbits, i.e., with $m_3 \rightarrow -\infty$, $m_1, m_2 \rightarrow +\infty$. The structure of the set of escape orbits is not as yet known.

We remark that the fact that the particles may leave a neighborhood of triple collision with arbitrarily large velocity was first noticed in the collinear three body problem by McGehee [McG 1]. Later, McGehee and Mather exploited this fact to construct a solution of the collinear four body problem which becomes unbounded in finite time [MM]. Their solution, however, admits infinitely many binary collisions, which they regularize as we did above. This then leaves open the question of the existence of a non-collision singularity in the n -body problem. Is there a solution which becomes unbounded in finite time and which does not experience any collisions at all?

Note that the rapid oscillation of solutions leaving a neighborhood of triple collision is reminiscent of the behavior of non-collision orbits in the anisotropic Kepler problem. We will exploit this idea in part three of these notes.

Finally, in Fig. 8 we have displayed the two different types of local behavior of solutions which are close to the homothetic solution along $\theta = \Delta_+$.

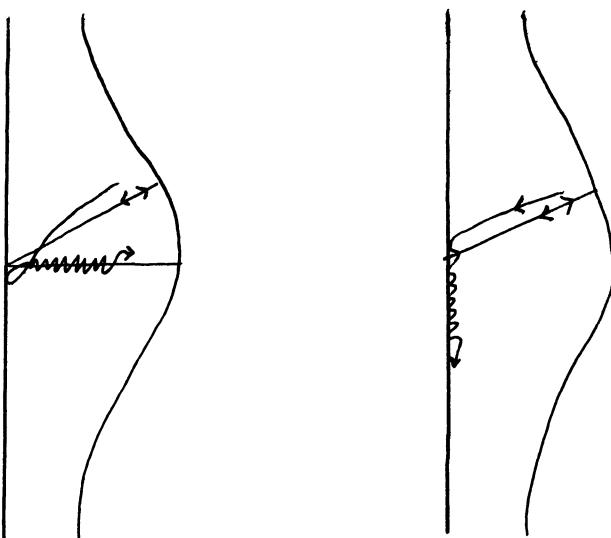


Fig. 8 Local behavior of near-collision orbits in the isosceles problem.

Part Two: Symbolic Dynamics and Mappings with Singularities.

In this section, we move away from the setting of vector fields and classical mechanics to discuss mappings which exhibit random or stochastic behavior. In particular we will describe the shift automorphism and show how it enters into the dynamics associated to mappings. Later, in part three, we will return to the examples of part one and apply the methods of this section to at least partially describe the global dynamics associated with these systems.

We begin with a description of the shift automorphism and the "Smale horseshoe" mapping due to Conley. For proofs, we will refer to Moser's book [Mo]. Conley's axioms are then applied to an interesting example, the Hénon mapping. Later we describe mappings which possess singularities, notably the baker transformation and a nonlinear version of this mapping which is associated with the restricted three body problem.

Mappings which admit singularities are important for our purposes since, in part three, we will show that both the isosceles problem and the anisotropic Kepler problem can be reduced to Poincaré mappings on two dimensional spaces which admit singularities. The singularities will correspond to points on collision or ejection orbits. We will also present a model mapping for the anisotropic problem which is a type of "infinite" baker transformation.

§2.1 The Smale Horseshoe.

In this section we review the construction of the now classical Smale horseshoe mapping. Historically, this was one of the first mappings which had infinitely many periodic points and was also structurally stable. This mapping admits a complicated Cantor set as a closed, invariant set on which the mapping is equivalent to the well known shift automorphism. The presentation below is due mainly to Conley. For proofs we will refer to Moser's book [Mo]. See also Smale's original paper [Sm] and Nitecki's book [Ni].

We first recall the construction of the shift automorphism. Let A be a finite or infinite set of positive integers. The elements of A are called symbols and A itself is called the alphabet. Let $\Sigma = \Sigma_A$ denote the set of all doubly infinite sequences of the form

$$(s) = (\dots s_{-2} s_{-1} s_0; s_1 s_2 \dots)$$

with $s_j \in A$. We define the shift automorphism $\sigma: \Sigma \rightarrow \Sigma$ by $(\sigma(s))_j = s_{j-1}$. That is,

$$\sigma(\dots s_{-1} s_0; s_1 s_2 \dots) = (\dots s_{-2} s_{-1}; s_0 s_1 s_2 \dots).$$

To topologize Σ we take as a neighborhood basis of $(s^*) = (\dots s_{-1}^* s_0^*; s_1^* s_2^* \dots)$ all "cylinder" sets of the form

$$U_j = \{(s) \mid s_i = s_i^* \text{ for } |i| \leq j\}.$$

Note that for any $\alpha_1, \alpha_2, \dots, \alpha_k \in A$ and any set of distinct indices i_1, \dots, i_k , the requirement that $s_{i_j} = \alpha_j$ for all $j \leq k$

defines an open subset of Σ . Also, in this topology, two sequences (s) and (t) are "close" if $s_i = t_i$ for $|i| \leq k$.

For finite alphabets, it is easy to see that Σ is a Cantor set. When A consists of infinitely many symbols, Σ turns out to be non-compact. Later we will describe and use several different compactifications of Σ .

In the topology above, the shift automorphism is clearly a homeomorphism. Moreover, periodic points for σ are given by repeating sequences. Thus periodic points for σ are dense in Σ . Also, σ admits a dense orbit. For example, if A consists of the two symbols 1 and 2, then the following sequence represents a dense orbit in Σ :

$$(s) = (\dots s_{-1} s_0; 1 2 \underbrace{11 12 21 22}_{\text{all 2 blocks}} \underbrace{111 112 \dots}_{\text{all 3 blocks}} \dots)$$

We simply enumerate all possible n blocks of symbols in some order, then all $n+1$ blocks, etc.

We now describe how the shift automorphism arises as an invariant subsystem of a smooth mapping of the plane. Let $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be given by

$$\begin{aligned} x_1 &= F_1(x_0, y_0) \\ y_1 &= F_2(x_0, y_0). \end{aligned}$$

For simplicity, we will concentrate on the restriction of F to the square S in the plane given by $0 \leq |x|, |y| \leq 1$. We define a horizontal strip in S to be the region bounded by the non-intersecting graphs of two smooth functions

$h_1, h_2: [-1,1] \rightarrow [-1,1]$ which satisfy $|h_j'(x)| < 1$ for $j = 1, 2$. Vertical strips are defined analogously. See Fig. 1.

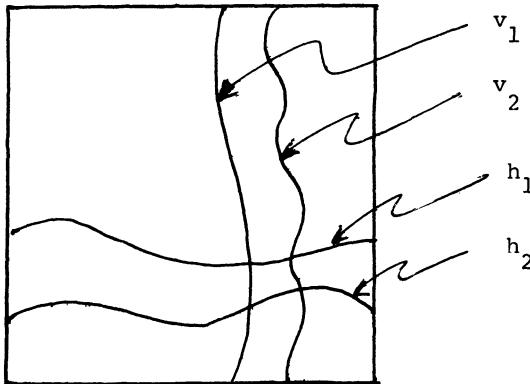


Fig. 1 Vertical and horizontal strips in S .

We now list two axioms which will guarantee that F possesses an invariant set in S on which F is equivalent to σ .

Axiom 1: There exist a finite or infinite number of horizontal and vertical strips H_i and V_i with $i \in A$. The mapping F takes V_i diffeomorphically onto H_i , with horizontal boundaries mapped to horizontal boundaries and vertical boundaries mapped to vertical boundaries.

We call this axiom the "strip" condition. Note that we need only verify this condition for F , not for its powers. Nevertheless, we will eventually be able to conclude something about F and all of its powers.

The next condition is the hyperbolicity condition. Let (ξ_0, η_0) be a tangent vector to S , and denote its image under dF by (ξ_1, η_1) and under dF^{-1} by (ξ_{-1}, η_{-1}) . Consider the sector bundles in the tangent bundle defined by

$$S_p^+ = \{(\xi_0, \eta_0) \in T_p \mathbb{R}^2 \mid |\eta_0| \leq |\xi_0|\}$$

$$S_p^- = \{(\xi_0, \eta_0) \in T_p \mathbb{R}^2 \mid |\eta_0| \geq |\xi_0|\}.$$

Then the second axiom for F is:

Axiom 2: For some $\lambda > 1$, and for any $p \in \cup V_i$, we have

i. $dF(S_p^+) \subset S_{F(p)}^+$

ii. For $(\xi_0, \eta_0) \in S_p^+, |\xi_1| \geq \lambda |\eta_0|$.

Also, for $p \in \cup H_i$, we have

iii. $dF^{-1}(S_p^-) \subset S_{F^{-1}(p)}^-$

iv. For $(\xi_0, \eta_0) \in S_p^-, |\eta_{-1}| \geq \lambda |\xi_0|$.

S^+ is called the bundle of unstable vectors and S^- is the bundle of stable vectors. According to Axiom 2, dF stretches vectors in S^+ while dF^{-1} stretches vectors in S^- .

If F satisfies Axioms 1 and 2, then we have the following theorem due to Smale:

Theorem 1. Suppose F is a smooth mapping $F: \cup V_i \rightarrow \cup H_i$ satisfying Axioms 1 and 2. Then there exists a subset Λ homeomorphic to $\Sigma = \Sigma_A$ on which F is topologically conjugate to the shift automorphism on Σ . More precisely, there is

a homeomorphism $s: \Lambda \rightarrow \Sigma$ such that the following diagram commutes

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

For the proof we refer to Theorem 3.1 in Moser's book [Mo, pg. 72].

We remark that this theorem says that the restriction of F to Λ behaves dynamically exactly as the shift automorphism does on Σ . That is, F -periodic points are dense in Λ , F has a dense orbit in Λ , etc.

In the case of a finite alphabet or in the case of one of the compactifications discussed in later sections for the infinite alphabet case, one can also assert that Λ is a hyperbolic set. We will discuss this in more detail below. We simply remark here that the existence of hyperbolic sets like Λ for a diffeomorphism is the cornerstone for Smale's definition of Axiom A diffeomorphisms. This important class of mappings have been studied exhaustively in recent years. See [Bo] for more details.

One can generalize the notion of the full shift automorphism somewhat. This leads to the notion of subshifts of finite type. These are restrictions of the full shift to certain invariant subsets of Σ_A . For the definition, let B be a $k \times k$ matrix of 0's and 1's where $k = \text{card } A$. Define the subset $\Sigma_B \subset \Sigma_A$ by $(s) \in \Sigma_B$ iff $B_{s_i s_{i+1}} = 1$ for all i .

That is, Σ_B consists of all sequences which are allowable in the following sense: β can follow α in such a sequence iff the (α, β) -entry of B is 1. So the matrix with all entries 1 corresponds to the full shift, while any other matrix yields a σ -invariant subset Σ_B of Σ_A , and the induced shift automorphism is a subshift of finite type. See [Fr] for more details.

We conclude this section by briefly describing the conjugacy s in Theorem 1 above. We will use this construction over and over again in the sequel.

Define inductively

$$\begin{aligned} v_{s_0 s_{-1} \dots s_{-n}} &= \text{closure}(v_{s_0} \cap F^{-1}(v_{s_{-1}}) \cap \dots \cap F^{-n}(v_{s_{-n}})) \\ &= \text{closure}(v_{s_0} \cap F^{-1}(v_{s_{-1} \dots s_{-n}})) \end{aligned}$$

It is easy to check that the $v_{s_0 \dots s_{-n}}$ form a nested sequence of vertical substrips in v_{s_0} as $n \rightarrow \infty$.

Similarly define

$$\begin{aligned} H_{s_1 \dots s_n} &= \text{closure}(F(v_{s_1}) \cap \dots \cap F^n(v_{s_n})) \\ &= \text{closure}(F(v_{s_1}) \cap F(v_{s_2 \dots s_n})) \\ &= \text{closure}(H_{s_1} \cap F(H_{s_2}) \cap \dots \cap F^{n-1}(H_{s_n})). \end{aligned}$$

Again the $H_{s_1 \dots s_n}$ form a nested sequence of horizontal strips in H_{s_1} . The main part of the proof of the above Theorem consists of showing that

$$\bigcap_{n=1}^{\infty} v_{s_0 \dots s_{-n}} \cap H_{s_1 \dots s_n}$$

consists of a unique point in S . Then the homeomorphism s assigns the sequence $(\dots s_{-1} s_0; s_1 s_2 \dots)$ to this point.

§2.2 The Hénon Mapping.

Axioms 1 and 2 in the previous section are often easy to verify in practice. In this section we give a rather simple example which illustrates this: we find a shift automorphism embedded in the dynamics of a quadratic diffeomorphism of the plane. The example is a two parameter family of mappings recently introduced by Hénon [H1] as a simple example of a mapping which seems to possess a "strange attractor". We will not discuss this particular phenomenon; rather, we will consider parameter values far away from those considered by Hénon. For our parameter values, the mapping of Hénon will be a concrete example of the Smale horseshoe.

The original Smale horseshoe mapping may be defined geometrically as follows. The mapping G is pictured in Fig. 1. $G|V_i$ is linear and hyperbolic and maps V_i onto H_i for $i = 1, 2$. Axioms 1 and 2 are easy to verify, so there exists a subset Λ of the square on which G is conjugate to the two-shift.

The mapping of Hénon is geometrically similar and is given analytically by $(x_1, y_1) = F(x_0, y_0)$ where

$$x_1 = A + B y - x^2$$

$$y_1 = x$$

where $A \in \mathbb{R}$ and $|B| < 1$ are two parameters. Actually, Hénon considers a slightly different form of this mapping given by

$$x_1 = 1 + y - Ax^2$$

$$y_1 = Bx$$

but these two forms are related for $A \neq 0$ by the linear conjugacy

$$x = X/A \quad y = BY/A .$$

F has constant Jacobian determinant $-B$, and its global inverse is given by

$$x_{-1} = y$$

$$y_{-1} = (x - A + y^2)/B .$$

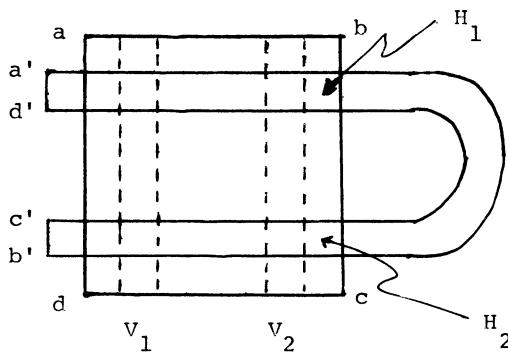


Fig. 1 The geometric description of the Smale horseshoe mapping.

For simplicity we will consider only the orientation reversing case with $0 < B \leq 1$. The orientation preserving case is handled in a similar fashion.

We assume throughout that

$$A > 3(1 + B)^2 \quad (1)$$

Actually, slightly lower A-values work as well, but this estimate is sufficient for our purposes. Let

$$R = R(A) = \frac{1}{2}(1 + B + \sqrt{(1 + B)^2 + 4A}) \quad (2)$$

So R is positive for $A > 0$ and is the larger root of

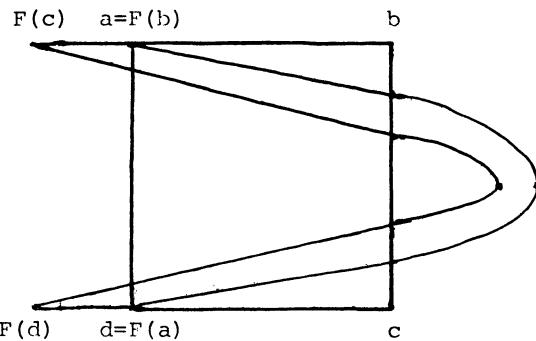
$$R^2 - (1+B)R - A = 0. \quad (3)$$

Note that

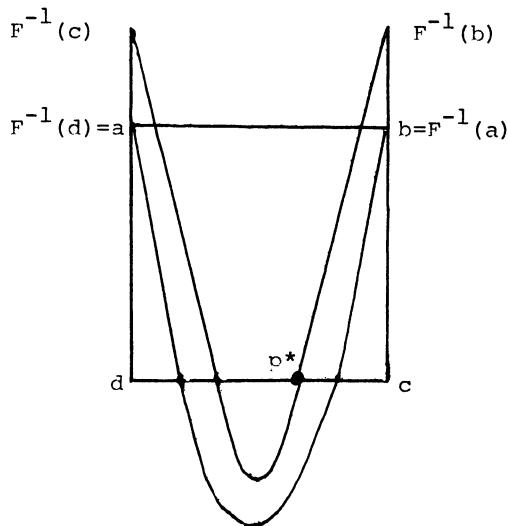
$$A - BR > R \quad (4)$$

for $A > 2(1+B)^2$.

Consider now the square S centered at the origin and having sides of length $2R$. We claim that F maps S into \mathbb{R}^2 as in Fig. 2. Indeed, the verticals $x = \pm R$ are mapped by F to the horizontals $y = \pm R$, while the fact that $F(a) = d$ and $F(b) = a$ follows immediately from (3). The fact that $F(c)$ and $F(d)$ lie to the left of S follows from the fact that $A - BR - R^2 < A + BR - R^2$. Finally, note that F maps $y = \pm R$ to parabolas opening to the left, with vertices on the x -axis. Since $F(0, R) = (A + BR, 0)$ and $F(0, -R) = (A - BR, 0)$, it follows from (4) that both parabolas cut completely across S as in Fig. 2.

Fig. 2 Geometric description of F .

Similarly, F^{-1} maps S across itself as in Fig. 3.

Fig. 3 Geometric description of F^{-1} .

Let p^* be the point on the right-hand intersection of the inner parabolic boundary of $F^{-1}(S)$ and the line $y = -R$ as in Fig. 3. One checks easily that the slope of the tangent line to the parabola at this point is larger than 1 if $A > 3(1+B)^2$. It then follows that the slopes of all tangent lines to vertical boundaries of $F^{-1}(S) \cap S$ are greater than one in absolute value. Hence these curves are vertical curves in the language of the previous section. Similarly, if $A > 3(1+B)^2$, it is easily checked that all horizontal boundaries of $F(S) \cap S$ in Fig. 2 are actually horizontal curves. This verifies Axiom 1 for Hénon's mapping.

To verify Axiom 2, first suppose that $|x| > \frac{1}{2}(\lambda+B)$ where $\lambda > 1$. Then we have, if $|\eta_0| < |\xi_0|$:

$$\begin{aligned} |\xi_1| &= |-2x\xi_0 + B\eta_0| > |2x||\xi_0| - |B\xi_0| \\ &> |2x||\xi_0| - B|\xi_0| \\ &> \lambda|\xi_0| \\ &= \lambda|\eta_1| \\ &> |\eta_1|. \end{aligned}$$

In particular, $|\xi_1| > |\eta_1|$ implies that dF preserves S^+ , while $|\xi_1| > \lambda|\xi_0|$ implies that dF expands vectors in S^+ for these values of x . So Axioms 2i and 2ii are satisfied provided $|x| > (\lambda+B)/2$.

Now let x^* be the x -coordinate of p^* in Fig. 3. We claim that, provided A satisfies (1), we have $x^* > (1+B)/2$.

This implies that $|x| > (\lambda+B)/2$ for some $\lambda > 1$ and for all points in $F^{-1}(S) \cap S$.

To prove this, we note that x^* is determined by the equation

$$(x^*)^2 = A - (1+B)R.$$

If we write $\Lambda = k(1+B)^2/2$, then we find

$$R = (1 + \sqrt{1+2k})(1+B)/2$$

so that

$$(x^*)^2 = (2k - 2 - 2\sqrt{1+2k})(1+B)^2/4.$$

For $k > 0$, this quantity increases with k and the left hand side equals 1 when $2k = 5 + 2\sqrt{5}$. Thus for A satisfying (1) we certainly have $x^* > (1+B)/2$. This proves Axioms 2i and 2ii.

Axioms 2iii and 2iv are verified similarly. We leave these proofs to the reader.

We remark that the invariant Cantor set for the Hénon mapping is a subset of the plane of measure zero. Yet this subset contains all of the interesting dynamics of the mapping. One can prove that all points outside of this set wander to infinity under either forward or backward iteration of F . See [DN]. Thus our set forms the entire non-wandering set for F and, as such, contains all of the non-trivial dynamics.

§2.3 The Baker Transformation and a Mapping Associated to the Restricted Three Body Problem.

In the last section we described a nonlinear mapping of the plane which possessed an invariant subset topologically conjugate to a complicated yet well understood model mapping, the shift automorphism. This conjugacy was valid on only a small subset of the plane, in fact, on a set of measure zero. In this section we describe another nonlinear mapping of the plane. Curiously, this mapping was also first introduced by Hénon [H2], this time in connection with his studies of the restricted three body problem. Unlike the previous mapping, chaotic behavior occurs here on the entire plane. Again we can model the dynamic behavior by a complicated yet well understood mapping, the baker transformation. For more details, we refer the reader to [AA].

We will consider the baker transformation of the open square $0 \leq |u|, |v| < 1$. The mapping is given analytically by

$$\begin{aligned} u_1 &= \frac{1}{2}(1 + u + 2[v]) \\ v_1 &= 2(v - [v]) - 1 \end{aligned}$$

where $[v]$ denotes the greatest integer less than or equal to v . Note that $v = 0$ is mapped to $v = -1$, and hence out of the open square. So the domain of the mapping does not include the u -axis. These points are singular points for the mapping. Similarly, one checks easily that no point maps to the v -axis under this mapping.

Geometrically, the baker transformation is described

in Fig. 1.

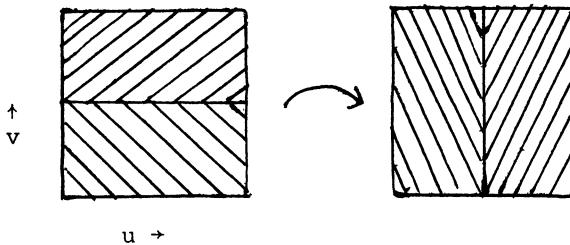


Fig. 1 Geometric description of the baker transformation.

This mapping contracts the upper and lower rectangles in the u -direction and expands them in the v -direction with a cut along the u -axis. As such, the mapping resembles the process of "rolling the dough", hence the curious name.

The u - and v -axes play the role of singularity sets for the baker transformation. Any point which maps to $v = 0$ under forward iteration has an orbit which terminates. Similarly, any orbit which meets $u = 0$ in backward time also terminates. The behavior of collision and ejection orbits in the classical mechanical systems of part one will be mirrored in such singular points.

The baker transformation is intimately related to the shift automorphism introduced earlier. Let $(s) = (\dots s_{-1} s_0; s_1 s_2 \dots)$ be a sequence of +1's and -1's, and define the mapping

$$u = \sum_{i=1}^{\infty} s_i / 2^i \quad v = \sum_{i=0}^{-\infty} s_i / 2^{1-i} \quad (1)$$

taking such sequences into points in the closed square. One may check easily that, under this mapping, the shift automorphism on sequences goes over to precisely the baker transformation, and the topology on the set of sequences induces the usual topology on the square.

Note that any sequence which is constant to the right or left of the semi-colon is mapped to the boundary of the square. Any sequence which is eventually constant (in either direction) may be shifted to such a point; hence such points eventually leave the square under iteration of the baker transformation.

Note that such eventually constant sequences are dense in the square. Indeed, they correspond to points in the square one of whose coordinates is a dyadic rational. Note also that sequences of the form $(\dots s_k +1 -1 -1 -1 \dots)$ and $(\dots s_k -1 +1 +1 +1 \dots)$ are to be identified under the projection above. We will encounter similar identifications later.

Finally note that the baker transformation preserves Lebesgue measure in the square. It can be shown that this mapping is ergodic and even a K-system on the square [AA]. As with the shift automorphism, there is also a dense orbit as well as dense periodic points.

We turn now to a nonlinear analogue of the baker transformation. Consider the mapping $(x_1, y_1) = F(x_0, y_0)$ given by

$$\begin{aligned} x_1 &= x_0 + 1/y_0 \\ y_1 &= y_0 - x_0 - 1/y_0 . \end{aligned}$$

In [H2], Hénon introduces this mapping as an asymptotic form of the equations of motion of a special case of the restricted

three body problem. One checks easily that F preserves Lebesgue measure in the plane, and its inverse is given by

$$x_{-1} = x_0 - 1/(x_0 + y_0)$$

$$y_{-1} = x_0 + y_0 .$$

F is singular along the x -axis while F^{-1} is undefined along the line $y = -x$.

The interesting fact about F is that it is topologically conjugate to the baker transformation of the open square. More precisely, we have:

Theorem 1. There is a homeomorphism $h: \mathbb{R}^2 \rightarrow \{(u,v) | 0 \leq |u|, |v| < 1\}$ which gives a topological conjugacy between F and the baker transformation on their respective domains.

We remark that this result is purely topological; despite the fact that both F and the baker transformation preserve Lebesgue measure, this theorem implies nothing about the ergodicity of F . In fact, it is an interesting open problem whether or not F is ergodic.

The remainder of this section is devoted to sketching a few of the details of the proof of Theorem 1. We refer to [De 6] for complete details.

Define the bundles of sectors in the tangent bundle

$$S^+ = \{(\xi, n) \mid \xi n \leq 0\}$$

$$S^- = \{(\xi, n) \mid \xi n \geq 0\} .$$

So S^- consists of all tangent vectors to \mathbb{R}^2 which lie in

the first or third quadrant. One checks easily that

$$dF(s^+) \subset s^+ \text{ and } dF^{-1}(s^-) \subset s^-.$$

Moreover, if we let

$\|(\xi, \eta)\| = |\xi| + |\eta|$, we also have for $(\xi_0, \eta_0) \in s^+$, $\|(\xi_1, \eta_1)\| \geq \min(2, 1 + 2/y^2) \|(\xi_0, \eta_0)\|$. Also, if $(\xi_0, \eta_0) \in s^-$, then $\|(\xi_{-1}, \eta_{-1})\| \geq (1 + 1/(x+y)^2) \|(\xi_0, \eta_0)\|$ where $(\xi_{-1}, \eta_{-1}) = dF^{-1}(\xi_0, \eta_0)$. These facts together imply that F is "non-uniformly" hyperbolic; in the sum norm, dF expands vectors in s^+ while dF^{-1} expands vectors in s^- . Note that the expansion coefficient tends to 1 as $|y| \rightarrow \infty$, hence the non-uniformity of the hyperbolicity.

We define an unstable curve to be a smooth curve in the plane all of whose tangents lie in the interior of the unstable sector bundle s^+ . Stable curves have tangents lying in s^- . It is clear that F maps unstable curves to unstable curves, and, in so doing, increases the length of these curves.

We now describe the conjugacy between F and the baker transformation. Let $(x_n, y_n) = F^n(x_0, y_0) = F^n(p)$ and suppose that (x_n, y_n) is defined for all integers n . Let $s(p) = (\dots s_{-1} s_0; s_1 s_2 \dots)$ where

$$s_j = s_j(p) = \begin{cases} +1 & \text{if } y_{-j} > 0 \\ -1 & \text{if } y_{-j} < 0. \end{cases}$$

The reversal of signs is confusing, but necessary. So the coding specifies which half-plane the j^{th} iterate of p lies in. We may extend this coding to points whose orbits terminate at one of the singular sets as follows. Suppose $y_{k+1} = 0$ for some $k \geq -1$. Then we let $s(p)$ be the one-sided terminating sequence $[0 s_{-k} \dots s_0; s_1 s_2 \dots]$. If $y_{-k}(p) = -x_{-k}(p)$ for some $k \geq 1$,

then we define $s(p) = (\dots s_{-1} s_0; s_1 \dots s_k 0)$. We also allow finite sequences of the form $[0 s_{-k} \dots s_j 0]$. So to each point $p \in \mathbb{R}^2$, there is assigned a sequence $s(p)$ of the form $[0 s_{-k} \dots s_0; s_1 \dots s_j 0]$ where k and/or j may be infinite. We assign the sequence $[0; 0]$ to the origin. Let Σ denote the set of all such sequences excluding those which begin and/or end with an infinite constant sequence, i.e., excluding sequences of the form $(\dots \alpha \alpha \alpha s_k s_{k+1} \dots)$ and $(\dots s_{k-1} s_k \alpha \alpha \alpha \dots)$. The topological conjugacy is then given as follows. To each $p \in \mathbb{R}^2$, we let $h(p)$ denote the point in the square which corresponds to the sequence $s(p)$ under the map (1). Thus one should think of the excluded sequences of the form $(\dots s_{k-1} s_k 1 -1 -1 -1 \dots)$ and $(\dots s_k -1 +1 +1 +1 \dots)$ as being identified with the allowed sequence $(\dots s_{k-1} s_k 0)$.

The following proposition shows why sequences which begin or end with a constant sequence are disallowed.

Proposition 2. Let $p = (x_0, y_0)$. Suppose $y_0 > 0$. Then there are integers $k > 0$ and $j < 0$ such that $y_k \leq 0$ and $y_j \leq 0$. Similarly, if $y_0 < 0$, there are integers $k > 0$ and $j < 0$ such that $y_j \geq 0$, $y_k \geq 0$.

Proof. We prove the first statement; the others follow similarly.

Observe first that if $y_n > 0$ for all n and $x_0 \geq 0$, then we have

$$x_n = x_{n-1} + 1/y_{n-1} > x_{n-1} > \dots > x_0 > 0.$$

Also

$$y_1 = y_0 - x_0 - \frac{1}{y_0} < y_0 - x_0$$

and by induction

$$y_n < y_0 - nx_0 .$$

Therefore, if $x_0 > 0$, we must eventually have $y_n \leq 0$.

In case $x_0 < 0$ and $y_0 > 0$ we claim that there exists $k > 0$ such that either $x_k \geq 0$ or else $y_k \leq 0$. Indeed, as before, we have $x_n > x_0$ so that

$$x_2 = x_1 + \frac{1}{y_1} > x_0 + \frac{1}{y_0} + \frac{1}{(y_0 - x_0)} .$$

Arguing inductively,

$$x_{n+1} > x_0 + \frac{1}{y_0} + \frac{1}{(y_0 - x_0)} + \dots + \frac{1}{(y_0 - nx_0)} .$$

Since the series $\sum_{n=0}^{\infty} \frac{1}{(y_0 - nx_0)}$ diverges, it follows that $x_n \rightarrow \infty$, which proves the result. qed

Again we refer the reader to [De 6] for the remainder of the proof.

Part Three: Singularities and Global Dynamics.

The object of this final part of these notes is to combine the techniques of the previous two to show how singularities in the equations of motion of a Hamiltonian system lead to a complicated but nevertheless somewhat understandable orbit structure. Although we cannot completely describe the orbit structure, we do find that a knowledge of the structure of the singular orbits provides in some sense a good overview of the global dynamics.

We deal here with two specific examples: the anisotropic Kepler problem and the isosceles three body problem. We will show that each of these systems admits a "singular" cross-section or surface of section. By this we mean a transverse cross-section for the flow which is crossed by all orbits except a few "special" collision and ejection orbits. Thus the associated Poincaré mapping has singularities, much like the baker transformation and mapping of Hénon of §2.3.

The notion of a singular cross-section has not been discussed very often in the literature, yet this type of cross-section seems to occur much more often in specific Hamiltonian systems than the usual cross-sections.

As with the Hénon mapping, we will relate our two examples to the shift automorphism on certain natural symbol spaces. We prove only that the given system projects onto these symbol spaces. It is an outstanding conjecture whether or not these projections are 1-1 (modulo symmetries). Hence our results give a complete classification of all possible types of motion

(classified symbolically) in the given systems. We fail only to show that there is a unique orbit corresponding to each sequence.

§3.1 Singular Cross-Sections.

Let X be a smooth vector field on a manifold M with (not necessarily complete) flow ϕ_t . A singular cross-section S for X is a smooth submanifold of codimension one in M satisfying:

- i. X is transverse to S .
- ii. There exist smooth submanifolds C_i $i = 0, \dots, n$ and E_j $j = 0, \dots, k$ having codimension at least one in S and such that the orbit through any point in each C_i (resp. E_j) does not return to S in forward (resp. backward) time.
- iii. The orbit through any point in $S - (\cup C_i)$ (resp. $S - (\cup E_j)$) returns to S in forward (resp. backward) time.
- iv. Each orbit of X meets S at least once.

Thus S is almost a traditional cross-section for the flow; only a few points in S fail to return in one direction or the other.

Given a singular cross-section S , there is given as usual a Poincaré mapping F defined on S . Actually F maps $S - (\cup C_i)$ diffeomorphically onto $S - (\cup E_j)$, so we think of $F: S \rightarrow S$ as a mapping with singularities. The C_i 's play

the role of immediate collision orbits, while the E_j 's consist of immediate ejection orbits. Of course, points may map onto a C_i after some iterates of F , so many orbits of F eventually reach collision.

Both of the major examples of part one admit singular cross-sections, as we now show. Consider first the anisotropic Kepler problem. From §1.6, the equations of motion are

$$\dot{q} = M^{-1} p$$

$$\dot{p} = -\nabla V(q)$$

where $M^{-1} = \text{diag}(\mu, 1)$ and $V(q) = -1/|q|$. We restrict to a negative energy level $H^{-1}(e)$ given by

$$H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q) = e < 0.$$

Let

$$S^+ = \{(q, p) \in H^{-1}(e) \mid p_2 = 0, q_2 > 0\}$$

$$S^- = \{(q, p) \in H^{-1}(e) \mid p_2 = 0, q_2 < 0\}$$

and $S = S^+ \cup S^-$. We claim that S is a singular cross-section for the flow on $H^{-1}(e)$.

First observe that both S^+ and S^- consist of initial conditions for which the momentum vector is parallel to the q_1 -axis or else zero. Thus, over each point in the interior of the Hill's region with $q_2 \neq 0$, there are two initial conditions in S , while over the zero velocity set there is but one (the zero momentum vector).

It is clear that both S^+ and S^- are two dimensional planes in $H^{-1}(e)$. In fact, we may parametrize S^+ by θ and

p_1 . Here $0 < \theta < \pi$ and $-\infty < p_1 < \infty$. This second inequality follows from the energy relation restricted to S^+ :

$$\frac{1}{2} p_1^2 = e^{-V(\theta)}/r.$$

S^- may be parametrized similarly.

To see that the vector field is transverse to S , we observe that, if $q_2 \neq 0$ and $p_2 = 0$, then $p_2' = -q_2/|q|^3 \neq 0$. So orbits reach a relative maximum q_2 -value whenever they cross S^+ and a relative minimum q_2 -value when crossing S^- .

Now we claim that a solution beginning at S^+ behaves in one of two possible ways. Either the solution immediately reaches collision with q_2 decreasing to zero, or else q_2 decreases until the solution curve meets S^- . Indeed, $q_2' = p_2$ which is negative just after passing S^+ , so q_2 decreases while the orbit remains in the upper half-plane. The solution curve cannot have a limiting q_2 -value with $q_2 > 0$, since $q_2'' < 0$ at such points. Hence the solution must meet the origin or else crosses the line $q_2 = 0$. The solution cannot be tangent to $q_2 = 0$; only the homothetic orbits have this property. Therefore any orbit which does not undergo collision after leaving S^+ must pass into the lower half-plane. Then similar arguments as above give that the solution must eventually pierce S^- , proving the claim. Of course, solutions leaving S^- have a similar property.

Now the orbits which leave S and immediately reach collision must lie on the stable manifolds of the rest points on the collision manifold. In fact, for any orbit in these stable manifolds (excluding those along $q_2 = 0$), there is a

unique point on the orbit which lies in S and which is the last such point before the orbit reaches collision. These are the immediate collision orbits. They form a finite union of submanifolds since they are given by the transverse intersection of the various stable manifolds with S . Note that they are transverse within the energy surface since the vector field is tangent to the stable manifolds but transverse to S . We will henceforth denote the set of points in S leading to immediate collision by C . Similar remarks hold for points in S leading to immediate ejection orbits. We denote these points by E .

We remark that the homothetic orbits along $q_2 = 0$ do not meet S ; but these are the only orbits in each energy level which do not. We will not worry about this slight violation of the definition of a singular cross-section. All other orbits must cross S at least once, since, by the previous arguments, any other orbit must have a maximum or minimum q_2 -value, thereby meeting S^+ or S^- . Hence we have proven that S forms a singular cross-section for the flow on a negative energy level in the anisotropic Kepler problem.

Now we discuss a similar singular cross-section for the isosceles problem. Some of the constructions here are similar to those for the anisotropic problem, so these details will be omitted. Actually, there are two different singular cross-sections for the system, each of which will be useful in the sequel. For the first cross-section, we recall the original equations of motion from §1.7:

$$\begin{aligned}x_1' &= 2p_1 \\x_2' &= (2+\epsilon/2\epsilon)p_2 \\p_i' &= -\partial V/\partial x_i\end{aligned}$$

where

$$V(x_1, x_2) = -\frac{1}{x_1} - \frac{4\epsilon}{\sqrt{x_1^2 + 4x_2^2}}$$

and $x_1 > 0$. Throughout we will restrict to a fixed negative energy level. Define S_1 to be the set of initial conditions satisfying $p_1 = 0$, $x_1 > 0$. Since $p_1 = 0$ precisely when the primaries reach maximum separation, it is intuitively clear that solutions pass through S_1 . More precisely, however, we have

$$p_1' = -\partial V/\partial x_1 < 0$$

along S_1 , so solutions do in fact cross S_1 transversely. We postpone the remainder of the proof for a moment.

To define the second singular cross-section S_2 , we recall McGehee coordinates (r, v, θ, w) from §1.7. In these coordinates the system is:

$$\begin{aligned}\dot{r} &= \frac{\cos \theta}{\sqrt{W(\theta)}} rv \\ \dot{v} &= \sqrt{W(\theta)} [1 - \frac{\cos \theta}{2W(\theta)}(v^2 - 4re)] \\ \dot{\theta} &= w \\ \dot{w} &= \sin \theta [-1 + \frac{\cos \theta}{W(\theta)}(v^2 - 2re)] - \frac{vw\cos \theta}{2\sqrt{W(\theta)}} + \\ &\quad + \frac{W'(\theta)}{W(\theta)}(\cos \theta - w^2/2)\end{aligned}$$

and the energy relation is

$$\frac{1}{2}w^2/\cos \theta - 1 = \frac{\cos \theta}{W(\theta)}(re - \frac{1}{2}v^2)$$

Let S_2 consist of all points in these coordinates with $\theta = \pm\pi/2$. From the energy relation, it follows that $w = 0$ on S_2 , while r and v are arbitrary. So S_2 consists of two planes in each energy level. We denote them by S_2^+ and S_2^- depending on the sign of θ .

Now $\dot{w} = -\sin(\pm\pi/2)$ along S_2 , so we have that solutions cross S_2 transversely. Again intuitively speaking, S_2 consists of the points of (regularized) double collision, so one expects solutions to pass through S_2 .

Now to show that both S_1 and S_2 form singular cross-sections, we first observe:

Proposition 1. Every solution curve $(x_1(t), x_2(t))$ in Jacobi coordinates satisfies: there exist times $t_0 < t_1 < t_2$ such that

- i. $x_1'(t_1) = 0$
- ii. $x_1(t) \rightarrow 0$ as $t \rightarrow t_0$ and t_2 .

The proof of this proposition is entirely analogous to the corresponding statement in the anisotropic problem, and hence is omitted.

Now $x_1 = 0$ goes over to $\theta = \pm\pi/2$ in McGehee coordinates, so solutions oscillate back and forth between S_1 and S_2 . There are only two possibilities for an orbit leaving S_1 : either it eventually meets S_2 (and passes through it), or else

it reaches triple collision. In the former case, the orbit must return again to S_1 by Proposition 1, while in the latter case, the orbit is an immediate collision orbit lying on one of the stable manifolds of a rest point in Λ . It follows that the set of immediate collision points in S_2 form a finite union of submanifolds as before. This proves that S_1 is a singular cross-section for the flow. The proof in the case of S_2 is equally easy.

§3.2 Symbolic dynamics and a Theorem of Gutzwiller.

The goal of this section is to set the stage for the proof of a theorem of Gutzwiller which classifies all possible types of orbits in the anisotropic Kepler problem. We begin with some symbolic dynamics and end with their relation to the differential equation.

Let Λ' denote the set of all doubly infinite sequences of positive integers, that is

$$\Lambda' = \{(s) = (\dots s_{-1} s_0; s_1 s_2 \dots) \mid s_j \in \mathbb{Z}^+\} .$$

We will augment Λ' by allowing certain terminating sequences as in the case of the Hénon mapping of §2.3. Namely, we will allow one-sided terminating sequences of the forms

$$\begin{aligned} & [\infty s_{-k} \dots s_0; s_1 s_2 \dots) \\ & (\dots s_0; s_1 \dots s_j \infty] \\ & [\infty s_{-k} \dots s_0; s_1 \dots s_j \infty] \end{aligned}$$

with $j, k \geq 1$. So $s_0 \neq \infty$. Let Λ denote Λ' augmented by these terminating sequences.

We topologize Λ as follows. If $(s^*) = (\dots s_{-1}^* s_0^*; s_1^* s_2^* \dots)$ is a non-terminating sequence, then we take as a neighborhood basis for (s) the cylinder sets of sequences (s) satisfying

$$\begin{aligned}s_j &= s_j^* \text{ if } |j| \leq k \\ s_j &\text{ arbitrary if } |j| > k\end{aligned}$$

On the other hand, if $(s) = (\dots s_{-1}^* s_0^*; s_1^* \dots s_j^* \infty]$, then we take as a neighborhood basis the sets consisting of sequences of either of the following types:

$$\begin{aligned}s_i &= s_i^* \quad -K \leq i \leq j \\ s_{j+1} &\geq K\end{aligned}$$

and all other s_i arbitrary, or else

$$\begin{aligned}s_i &= s_i^* \quad -K \leq i < j \\ s_j &= s_j^* - 1 \\ s_{j+1} &= 1 \\ s_{j+2} &\geq K\end{aligned}$$

and all other s_i arbitrary. The rationale for this strange topology will be given below. We define similar neighborhood bases for the other terminating sequences.

The shift automorphism is defined in this case by

$$\sigma(\dots s_{-1} s_0; s_1 s_2 \dots) = (\dots s_{-1} s_0 s_1; s_2 \dots),$$

i.e., σ shifts the semi-colon to the right. We find it more

natural to shift in this direction for this problem. Consequently, points of the form $(\dots s_{-1} s_0; \infty]$ are not in the domain of σ , whereas points of the form $[\infty s_0; s_1 s_2 \dots)$ are not in the range. These points will represent the immediate collision and ejection orbits respectively. Hence we denote them by \hat{C} and \hat{E} .

We now return to the anisotropic Kepler problem. Our goal is to associate a sequence in Λ with every trajectory in the system in some meaningful way. This is accomplished as follows.

Let p be an initial condition in S and suppose that $F^j(p)$ is defined. We define the j^{th} passage of p to be the segment of orbit containing $F^j(p)$ beginning at the first prior crossing of $q_2 = 0$ and ending at the next crossing of the q_1 -axis. We include the endpoints of the orbit segment, even if one or both is the origin, i.e., even if this segment of orbit begins and/or ends in collision.

Now let $s_j = s_j(p)$ denote the number of times the orbit through p crosses the q_2 -axis during the j^{th} passage. We count collision and ejection as a crossing.

Proposition 1. Suppose p does not lie on the homothetic orbit along the q_2 -axis. If $F^j(p)$ is defined, then $1 \leq s_j < \infty$.

Proof. The proof that $s_j < \infty$ is trivial. If $s_j(p) = 0$, then there is a point during the j^{th} passage at which $\theta = 0$. For definiteness, let us assume that $0 < \theta < \pi/2$ at this point. Now $\ddot{\theta} = -V'(\theta) > 0$ at any such point, so $\theta(\tau)$ has

a minimum there. It follows, then, that the orbit must cross $\theta = \pi/2$ before reaching $\theta = 0$. The other cases are handled similarly.

qed

So to each point p in S we have associated a sequence of positive integers $s(p) = (\dots s_{-1} s_0; s_1 s_2 \dots)$. The sequence terminates to the right at j if $F^j(p) \in C$ and on the left at $-k$ if $F^{-k}(p) \in E$. We associate to such points terminating sequences in Λ of the form $(\dots s_{-1} s_0; s_1 \dots s_j \infty)$ or $[\infty s_{-k} \dots s_0; s_1 \dots)$. Therefore, to each point p in S there is associated a sequence $s(p)$ in Λ of the form $[\infty s_{-k} \dots s_0; s_1 \dots s_j \infty]$ where k and/or j may be infinite. See Figs. 1 and 2.

So s provides a mapping $s: S \rightarrow \Lambda$ and we obviously have the commutative diagram

$$\begin{array}{ccc} S - C & \xrightarrow{F} & S - E \\ s \downarrow & & s \downarrow \\ \Lambda - \hat{C} & \xrightarrow{\sigma} & \Lambda - \hat{E} \end{array}$$

where \hat{C} denotes the set of "collision" sequences of the form $(\dots s_{-1} s_0; \infty)$ while \hat{E} contains sequences of the form $(\infty; s_1 s_2 \dots)$. Clearly, the sequence $s(p)$ gives a good qualitative picture of the behavior of the solution through p . Gutzwiller's theorem [Gu 2] mainly asserts that the mapping s is surjective. Thus every conceivable type of motion is possible in the anisotropic Kepler problem. Given a possible sequence of numbers of crossings, there is at least one orbit of the differential equation which has the prescribed behavior.

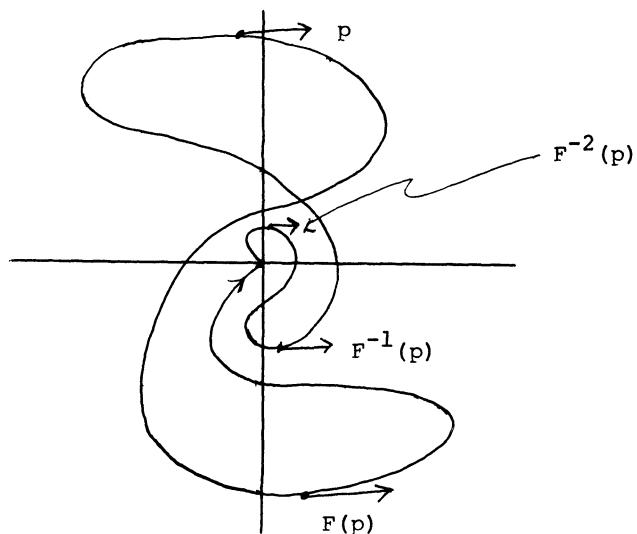


Fig. 1 The orbit through p is a collision/ejection orbit with $s(p) = [\infty \ 2 \ 2 \ 3; \ 3 \ \infty]$.

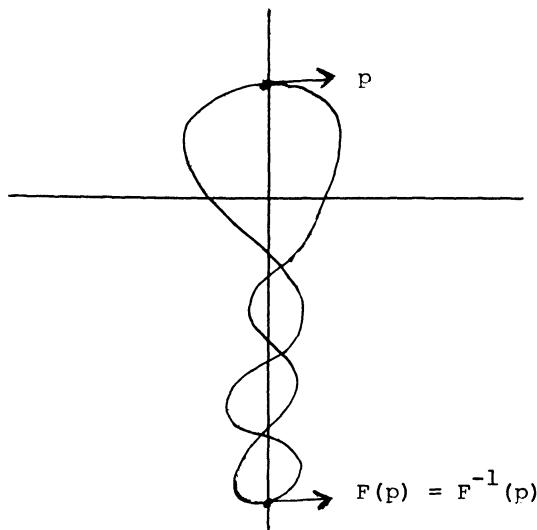


Fig. 2 The orbit through p is periodic with $s(p) = (\dots \ 7 \ 1; \ 7 \ 1 \ 7 \ \dots)$.

One can in fact say more than just surjectivity of s :

Theorem 2. (Gutzwiller [Gu 2]) The mapping $s: S \rightarrow \Lambda$ is a continuous surjection. Moreover, corresponding to any periodic sequence, there is at least one actual periodic solution of the system.

The mapping s is definitely not 1-1 as we have defined it, since the various symmetries of the problem give rise to distinct orbits with the same sequences. However, Gutzwiller has conjectured that, up to these symmetries, there is a unique solution corresponding to each sequence in Λ . If true, this provides a topological conjugacy between the anisotropic Kepler problem and the shift automorphism, a rather unexpected isomorphism!

We might add that Gutzwiller has ample numerical evidence to support his conjecture [Gu 3]. On the other hand, a single generic elliptic closed orbit in the system, however long, would, by the Moser Twist Theorem, serve to disprove the conjecture. No such orbit has been found, however. Thus one might try to prove that the Poincaré mapping F is hyperbolic in some sense. Gutzwiller has shown numerically that the short periodic solutions are hyperbolic [Gu 4], while the author has shown that some of the longest periodic solutions are also hyperbolic [De 1], thus lending some additional credence to the conjecture. Later, we will present a model mapping similar in many respects to F for which one can verify the conjecture. Whether this mapping is conjugate to F , however, is speculation at best at this time.

§3.3 Proof of Gutzwiller's Theorem.

In this section we prove that the mapping $s: S \rightarrow \Lambda$ is both continuous and surjective. Our proof is basically geometric and differs completely from Gutzwiller's proof. It also has the advantage of extending to other mechanical systems such as the isosceles three body problem.

We first prove continuity. Let $p \in S$ and suppose first that $s(p)$ is a non-terminating sequence. We show that points close enough to p in S have associated sequences which agree with $s(p)$ arbitrarily far to the right and left of the semi-colon. Indeed, let $\gamma_j(p)$ denote the j^{th} passage of the orbit through p , i.e., the orbit segment connecting $F^j(p)$ to the q_1 -axis. $\gamma_j(p)$ is compact and is never tangent to the q_2 -axis. Hence $\gamma_j(p)$ intersects the q_2 -axis transversely at a finite number of isolated points which are disjoint from the origin. This follows since, if $\gamma_j(p)$ were tangent to the q_2 -axis, then uniqueness of solutions would force it to lie for all time along this axis, thereby agreeing with the homothetic orbit already lying on this axis. By continuity of solutions with respect to initial conditions, nearby solutions meet the q_2 -axis transversely at nearby points, proving that their associated sequences agree with $s_j(p)$. Hence, given any integer k , for q close enough to p , we have $s_j(q) = s_j(p)$ for all j such that $|j| \leq k$. So s is continuous at such points.

For collision and/or ejection orbits, we argue as follows. Suppose $s(p) = (\dots s_0^*; s_1^* \dots s_j^* \dots)$; the other cases are

handled similarly. We show that nearby solutions have associated sequences which take one of two forms, agreeing with the topology imposed on Λ . For definiteness, suppose that the orbit through p reaches collision at $\theta = 0$, say from above. See Fig. 1. Now nearby solutions either reach collision,

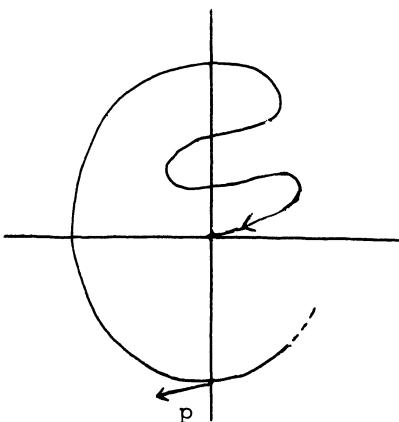


Fig. 1 The initial condition p leads to a typical collision orbit.

or else follow one of the two branches of the unstable manifold emanating from the appropriate rest point in the collision manifold. This was discussed in §1.7. See Fig. 4 in that section. The nearby behavior here is illustrated in Fig. 2 below. For q , we have the associated sequence $(\dots s_0^*; s_1^* \dots s_{j-1}^* 1 k \dots)$ where k is large. For q' , we find $s(q') = (\dots s_1^* \dots s_j^* k \dots)$ again with k large. Both of these sequences are close to $s(p)$ in the topology we have chosen on Λ . Hence s is continuous at these points, also.

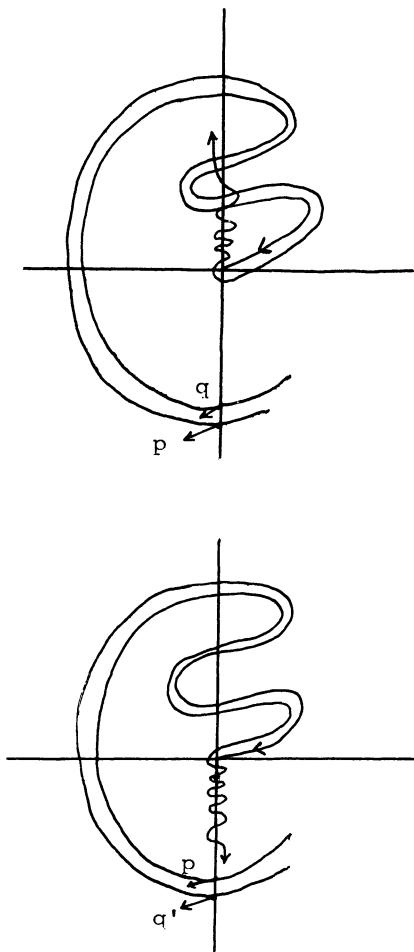


Fig. 2 The two types of nearby behavior for solutions beginning close to p .

We now prove surjectivity. Using θ and p_1 as coordinates, the singular cross-section may be represented as the strip $0 < \theta < \pi$, p_1 arbitrary. The zero velocity set in S^+ is given by $p_1 = 0$. We denote the special central point $p_1 = 0$, $\theta = \pi/2$ by q^+ ; this is the point where the homothete-

tic orbit along $\theta = \pi/2$ meets S^+ . One checks easily that the points of intersection of the other immediate collision and ejection orbits with S^+ form a pair of topological spirals converging to q^+ . By this we mean that $C \cap S^+$ consists of two smooth curves which spiral infinitely often about q^+ as they connect $\theta = 0$ or $\theta = \pi$ to q^+ . These curves are only continuous at q^+ . See Fig. 3. This can be verified

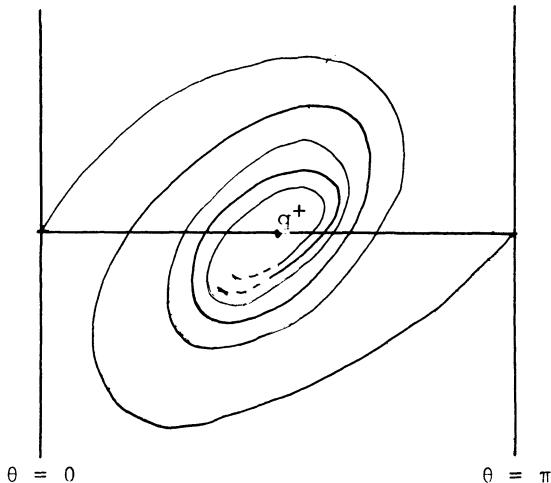


Fig. 3 The set of immediate collision or ejection orbits form a topological spiral in S^+ .

using local analysis about the sink at $\theta = \pi/2$ in the collision manifold, together with the fact that one branch of the unstable manifolds at the saddles falls into this sink.

Note that one branch of C can be obtained from the other via the reflection $\theta \rightarrow \pi - \theta$, $p_1 \rightarrow -p_1$. Also, E can be obtained from C via the reversing reflection $p_1 \rightarrow -p_1$.

The full picture of the immediate collision and ejection orbits in S^+ is given in Fig. 4.

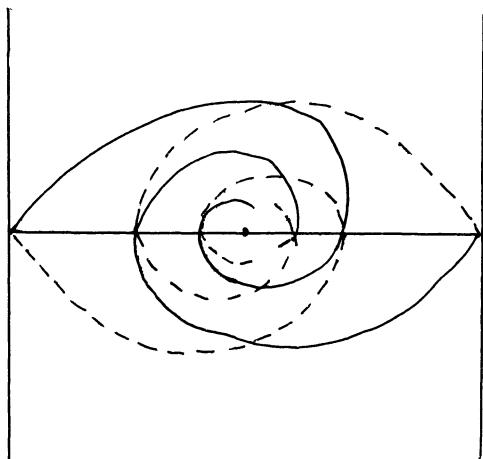


Fig. 4 The immediate collision orbits in S^+ are represented by solid lines, while the immediate ejection orbits are give by broken lines.

One has a similar picture in S^- . In fact, the reflection $q_2 \rightarrow -q_2, p_2 \rightarrow -p_2$ sends the picture in S^+ to that in S^- .

We remark that C and E need not be logarithmic spirals; they may in fact wander quite erratically in S^+ , although numerical evidence indicates that this does not happen. All that is necessary for our purposes is that C wind about q^+ infinitely often and separate both S^+ and S^- into two components. Since each spiral in $C \cap S^+$ lies in the unstable

manifold of a distinct rest point, it follows of course that these two curves cannot intersect.

Note that E and C intersect infinitely often along the zero velocity set and elsewhere, a fact we have already noted in §1.7.

We henceforth denote $C \cap S^+$ by c^+ and $C \cap S^-$ by c^- . Similarly, e^+ and e^- represent $E \cap S^+$ and $E \cap S^-$ respectively. The subset of c^+ asymptotic to the equilibrium point at $\theta = 0$ will be denoted by c_0^+ ; c_π^+ , e_0^+ , e_π^+ , etc. are defined similarly.

Now recall that, for any $p \in S$, $s_0(p)$ denotes the number of times the orbit through p crosses the q_2 -axis between the crossings of the q_1 -axis just prior to and just after p along the orbit. Note that $s_0 = 1$ or 2 near $\theta = 0$ and $\theta = \pi$. This follows from the local description of solutions near the homothetic orbits along $\theta = 0, \pi$ given in §1.7. See Fig. 5. Also, $s_0(p) \rightarrow \infty$ as $p \rightarrow q^+$.

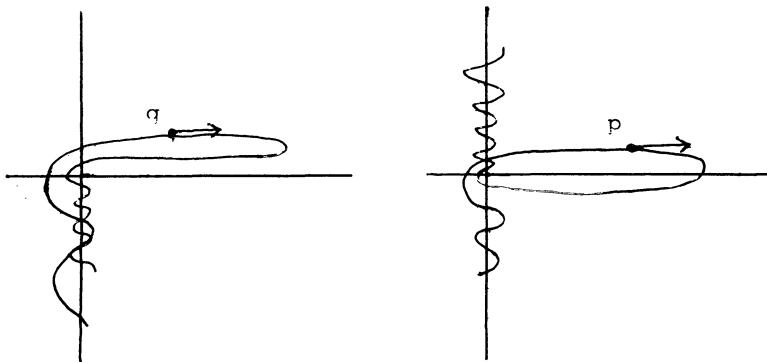


Fig. 5 In these diagrams, $s_0(q) = 2$ and $s_0(p) = 1$.

Now s_0 is continuous on S , except at points in C or E . At such points, s_0 increases or decreases by 1, again by the results of §1.7.

We wish to subdivide S^+ and S^- into a collection of "allowable strips" on which s_0 is constant. Once this is accomplished, we will show that these strips are mapped across each other just as the horizontal and vertical strips were mapped giving the Smale horseshoe. This implies that any positive integer can follow any other in a given sequence. We first need several preliminary lemmas.

Lemma 1. Suppose $p \neq q^+$. Then

- i. If $p \in c_0^+ \cap e_0^+$, then $s_0(p)$ is even.
- ii. If $p \in c_0^+ \cap e_\pi^+$, then $s_0(p)$ is odd.
- iii. If $p \in c_\pi^+ \cap e_\pi^+$, then $s_0(p)$ is even.
- iv. If $p \in c_\pi^+ \cap e_0^+$, then $s_0(p)$ is odd.

Proof. Each of these statements follows from a simple counting argument, since the orbit through p can never be tangent to the q_2 -axis.

Lemma 2. Suppose $p \in c_0 \cap e_\pi$ and $s_0(p) = 2k + 1$. Then there exist closed intervals in both c_0 and e_π ending at p on which $s_0 = 2k+1$. Also, there exist open intervals abutting p in c_0 and e_π on which $s_0 = 2k$.

Proof. By the local behavior of solutions near an ejection orbit, a solution near p in e_π assumes one of two forms.

Either we have $s_0 = 2k$ as in Fig. 6, or else $s_0 = 2k+1$, as in Fig. 7. Similar considerations hold for collision orbits, thus proving the result. qed

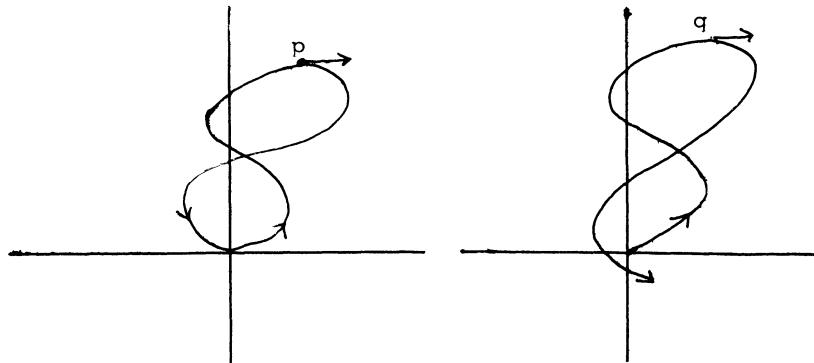


Fig. 6 For the collision/ejection orbit through p , we have $s_0(p) = 2k+1$, while $s_0(q) = 2k$.

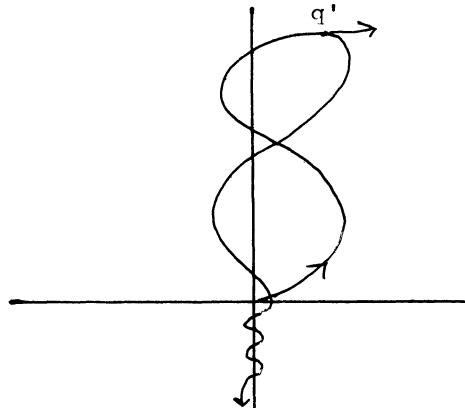


Fig. 7 Here we find $s_0(q') = 2k+1$.

Remark. This lemma shows that, in S , near a point in $c_0 \cap e_\pi$, we have the local picture illustrated in Fig. 8:

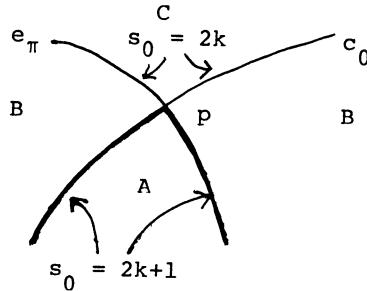


Fig. 8 $s_0(p) = 2k+1$, while $s_0 = 2k$ or $2k+1$ on abutting arcs in C and E . In A , we have $s_0 = 2k+1$, while $s_0|B = 2k$ and $s_0|C = 2k-1$.

The typical solution in the region A is depicted in Fig. 9.

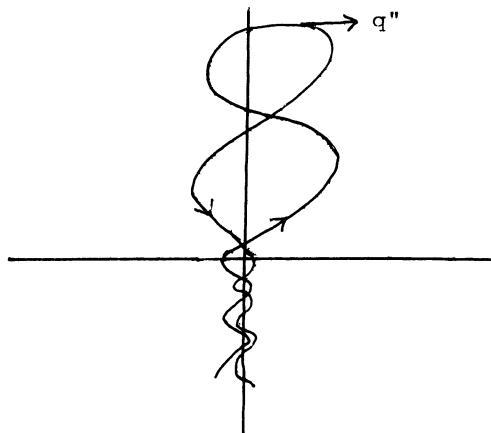


Fig. 9 $s_0(q'') = 2k+1$, and the initial condition q'' is close to p in Figs. 6 and 8.

Remark. The lemma and following remark are obviously true if $p \in c_\pi \cap e_0$.

Lemma 3. Let $p \in c_0 \cap e_0$ and suppose $s_0(p) = 2k$. Then there exist closed intervals in c_0 and e_0 ending at p on which $s_0 = 2k$. Also, there exist open intervals abutting p in c_0 and e_0 on which $s_0 = 2k-1$.

Proof. The proof is similar to that of Lemma 2 and hence is omitted. qed

Again we remark that a similar result holds for $c_\pi \cap e_\pi$, and that the appropriate local picture is as given in Fig. 10.

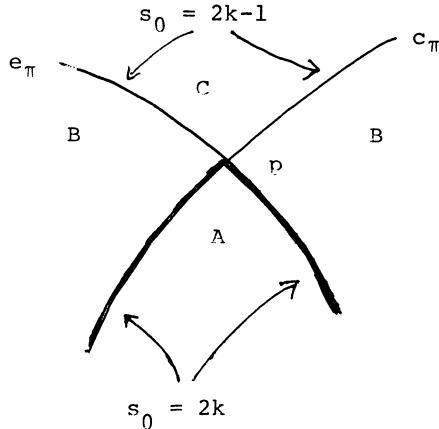


Fig. 10. $s_0(p) = 2k$, while $s_0|_A = 2k$, $s_0|_B = 2k-1$, $s_0|_C = 2k-2$.

Now consider one of the components of $S^+ - E$. This is an infinite spiralling strip winding down to q^+ in S^+ as shown in Fig. 11.

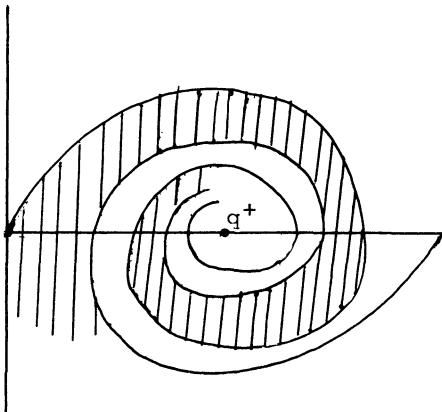


Fig. 11 The components of $S^+ - E$ form infinite spiralling strips converging to q^+ .

The function s_0 changes its values in this strip only at points in c_0 or c_π , and there s_0 either increases or decreases by 1. Now C meets the boundary of this strip in one of two ways: either C cuts completely across the strip or else C meets one of the two boundaries twice in succession. See Fig. 12. In the former case, we call the arc in C cutting across the strip an admissible boundary. In the latter case, C bounds a two dimensional disk in the strip which we call an extraneous disk. Note that two or more of these disks may be nested. Now consider the strip with all extraneous disks removed. The remaining admissible boundaries

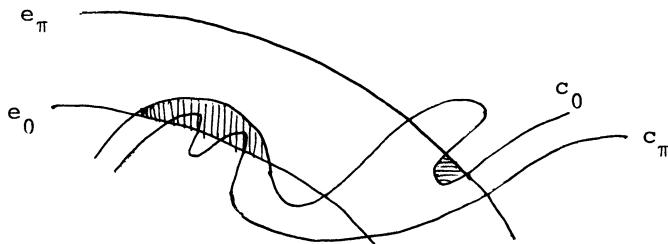


Fig. 12 The shaded regions represent extraneous disks.

cut the strip into a series of substrips on which s_0 is necessarily constant. If $s_0 = k$ on such a strip, we call it a k-strip. Since s_0 is onto the positive integers and changes by 1 upon crossing an admissible boundary, it follows that, for each integer k , there exists at least one k -strip in each component of $S - E$. A typical k -strip is depicted in Fig. 13.

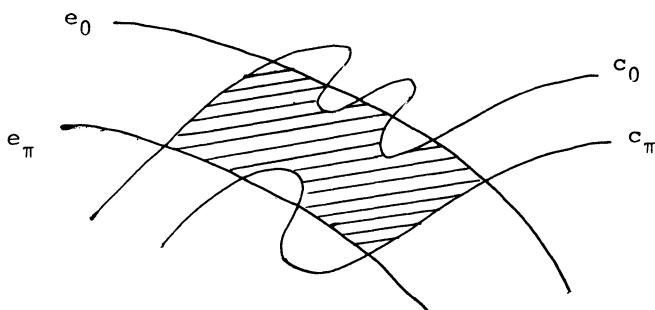


Fig. 13 The shaded region represents a typical k -strip.

So a k -strip is bounded by two admissible boundaries in C as well as several arcs in C (composing boundaries of extraneous disks) and several in E . Each pair of consecutive admissible boundaries in C cuts off a pair of arcs in E , and together these four arcs bound a rectangular region in S which consists of exactly one k -strip together with finitely many extraneous disks.

Lemma 4. Suppose the admissible boundaries of a k -strip both lie in c_0 (or in c_π). Then s_0 assumes the same value on each abutting strip.

Proof. The proof involves repeated applications of lemmas 2 and 3 to take care of successive extraneous disks. Hence we merely present a sketch which gives the relevant details.

Suppose we have a $2n$ -strip whose boundaries lie in c_0 and which is abutted by a $(2n+1)$ -strip. See Fig. 14. We claim that

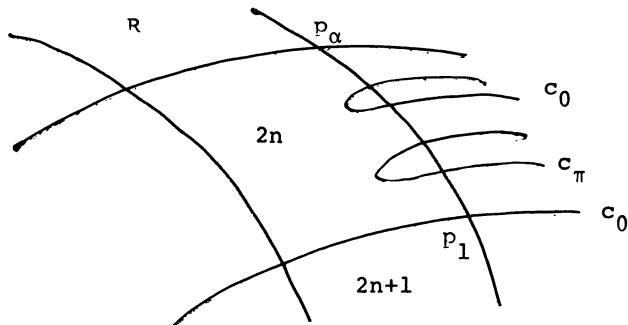


Fig.14.

$s_0(p_\alpha) = 2n+2$, so that, by Fig. 10, $s_0|_R = 2n+1$ also. To see this, observe that $s_0(p_1)$ must equal $2n+2$ by lemma 2.

Lemmas 1, 2, and 3 can be used repeatedly to fill in the other s_0 values as in Fig. 15:

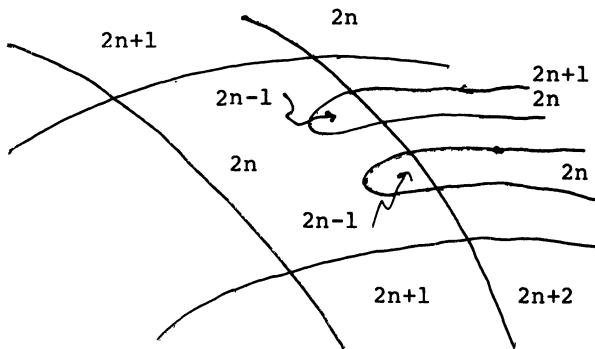


Fig.15.

No matter how many disks meet the $2n$ -strip, we always have $s_0|_R = 2n+1$. All other possibilities may be handled similarly. qed

As a consequence of the preceding lemma, we can prove:

Proposition 5. In each component of $S - E$ and for each integer $k \geq 3$, there exists at least one k -strip R_k one of whose admissible boundaries lies in c_0 and the other in c_π .

Proof. For simplicity, we consider only S^+ . For $k \geq 3$, the region in S^+ with $s_0 \geq k$ is compact. In any of the two components of $S^+ - E$, the number of k -strips is finite, and

so, given k , we can find the last k -strip in this component, i.e., the "closest" k -strip in this component to q^+ . Hence closer n -strips in this component satisfy $n \geq k+1$. In particular, one of the abutting strips must be a $k+1$ strip. By the lemma, if both admissible boundaries lie in c_0 (or in c_π), the other abutting strip must also be a $(k+1)$ -strip. Continuing in this fashion, we must eventually reach a k -strip satisfying the hypotheses of the proposition, or else reach a contradiction since s_0 is onto the positive integers. qed

Remark 1. The proposition does not necessarily extend to the case where $k = 1$ or $k = 2$ since these regions may abut $\theta = 0$ or $\theta = \pi$. In this case, the associated strips must exist (see Fig. 5), but they may be unbounded.

Remark 2. The numerical work of Gutzwiller indicates that, in fact, the k -strips are all bona fide "rectangles" and all extraneous disks are absent. Indeed, in our model mapping, this is the case.

We call k -strips whose admissible boundaries consist of one arc in c_0 and one in c_π admissible k -strips. We now investigate how the Poincaré mapping F affects admissible k -strips. Let R_k be any k -strip in S^+ . F maps R_k into one of the components of $S^- - E$, as immediate ejection orbits are not in the range of F . In fact, we claim that $F(R_k)$ intersects every admissible j -strip in this component. To see this, we first prove:

Lemma 6. Let f_0 and f_π denote the boundary components in e_0 and e_π respectively of an admissible k-strip. Then $s_0 = k$ on one of f_0 or f_π , and $s_0 = k+1$ on the other.

Proof. Suppose $s_0 = k$ on the interior of the admissible strip. For each point p in the interior of this strip, the 0^{th} passage begins on the same side of the origin on the q_1 -axis. Similarly, the terminal point of the 0^{th} passage for any point p always lies on the same side of the origin. For definiteness, let us assume that the 0^{th} passage of all of these orbits begins to the left of the origin. Then, by deforming orbits to the boundaries of the strip, one checks easily that $s_0 = k$ along f_0 while $s_0 = k+1$ along f_π . See Fig. 16. Similar methods also give the following:

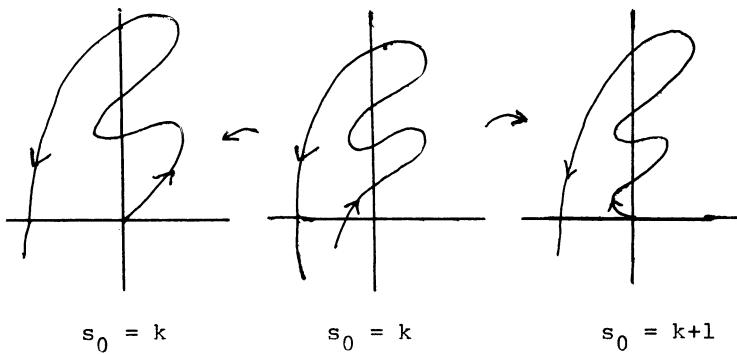


Fig. 16 The center orbit is a typical orbit in a k-strip, while the left and right orbits are deformations to the boundaries f_0 and f_π .

Lemma 7. Let g_0 and g_π be the admissible boundaries of an admissible k-strip in S . Then $s_0 = k$ on one of g_0 or g_π , while $s_0 = k+1$ on the other.

We can now finish the proof of Gutzwiller's Theorem:

Proof of surjectivity: Suppose R'_j is an admissible strip in the component of $S^- - E$ intersected by $F(R_k)$. We claim that $F(R_k)$ meets both admissible boundaries of R'_j . Indeed, by Lemma 7, for p on one admissible boundary of R'_j , we have $s_0(p) = k$, while $s_0 = k+1$ on the other. By the local behavior of near-ejection orbits, we have that $s_1(p) = 1$ for p near the boundary with $s_0 = k+1$. Moreover, as p approaches this boundary, $F(p)$ approaches the collision manifold. On the other hand, $s_1 \rightarrow \infty$ as $p \rightarrow$ the other boundary, and, in fact, $F(p) \rightarrow q^-$. Therefore, $F(R_k)$ meets every admissible j -strip in its component in S^- . Note that, in fact, $F(R_k)$ cannot cross the ejection boundaries of any R'_j . Hence $F(R_k)$ must cross each admissible boundary of R'_j . Moreover, F is undefined on C , so that $F(R_k)$ must be bounded entirely by the image of the ejection orbits. In particular, it follows that $F(R_k) \cap R'_j$ contains a substrip bounded by ejection orbits and connecting the admissible boundaries of R'_j . See Fig. 17.

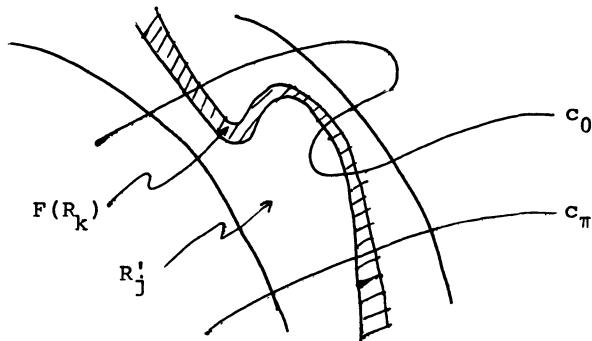


Fig. 17

 $F(R_k) \cap R'_j$

Of course, this substrip may intersect extraneous disks in R_j' .

Now consider inverse images. Arguing just as before, one finds that $F^{-1}(R_k)$ meets every admissible j -strip in its component of $S^- - C$. Moreover, for any such strip R_j' , we have that $F^{-1}(R_k)$ meets R_j' in a substrip which does not cross any of the collision boundaries of R_j' and which connects both ejection boundaries of R_j' as in Fig. 18. Standard arguments as in the case of the Smale horseshoe in §2.1 then give surjectivity. Note that we do not have hyperbolicity and that prevents us from concluding uniqueness.

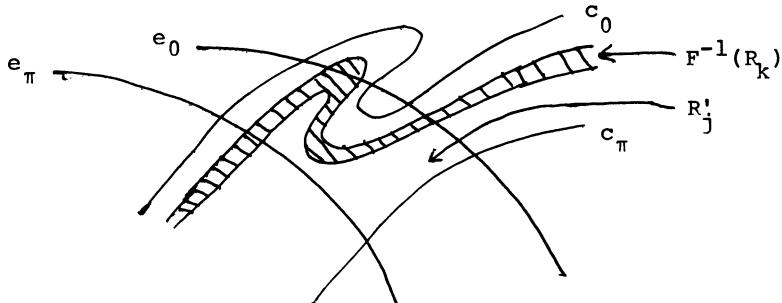


Fig.18 $F^{-1}(R_k) \cap R_j'$

§3.4 A Model Mapping for the Anisotropic Kepler Problem.

In this section we present a mapping which resembles the Poincaré mapping for the anisotropic problem described in the previous section, and for which we can verify topological conjugacy with the shift mapping. Whether this mapping is

actually conjugate to F or not, we do not know. We remark that this mapping is similar to Hénon's mapping associated with the restricted three body problem in the sense that the singular points are handled in the same way.

The mapping G is defined on the open annulus $I \times S^1$ parametrized by (x, θ) , with $0 < x < 1$ and θ defined mod 1. Via polar coordinates we also think of this mapping as being defined on the punctured two-disk. After removing all symmetries in the anisotropic problem, the Poincaré mapping F in this system reduces to a mapping on such a space (the "hole" in the middle corresponds to the homothetic solution along the q_2 -axis). We will use both viewpoints in this section. From an ergodic theorist's point of view, however, the first viewpoint might be preferable since our mapping preserves Lebesgue measure on $I \times S^1$.

We define G in two steps. First let $G_1(x, \theta) = (x, \theta - 1/x)$. G_1 is an infinite twist of the annulus. We also define $G_2(x, \theta) = (-\theta, x)$. G_2 interchanges the x and θ coordinates (with a sign change), and is clearly discontinuous along the line $\theta = 0$. In fact, $\theta = 0$ is mapped out of the annulus, and therefore plays the role of the singular set for G_2 . Now finally define $G = G_1 \circ G_2 \circ G_1$. So the mapping G may be visualized as a twist followed by a flip followed by another twist. As both G_1 and G_2 preserve Lebesgue measure on $I \times S^1$, it follows that G does also. G is given by

$$x_1 = -\theta + 1/x$$

$$\theta_1 = x + 1/(\theta - 1/x)$$

and its inverse by

$$x_{-1} = \theta + 1/x$$

$$\theta_{-1} = -x + 1/(\theta + 1/x) .$$

So G is singular along the spiral $\theta = 1/x$, while G^{-1} is undefined along $\theta = -1/x$. This first spiral should be thought of as representing the immediate collision orbits in the singular cross-section, while the second represents the immediate ejection orbits. See Fig. 1.

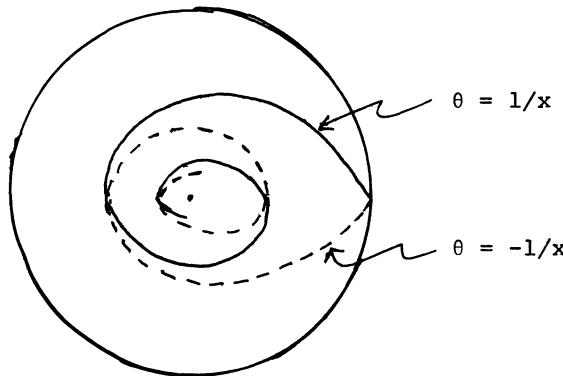


Fig. 1 The solid line represents immediate collision orbits, while the broken line represents immediate ejection orbits.

As in the case of the Hénon mapping, one of the most important features of this mapping is its non-uniform hyperbolicity. Define

$$S^u = \{(\xi, \eta) \mid \xi_n \geq 0\}$$

$$S^s = \{(\xi, \eta) \mid \xi_n \leq 0\} .$$

Letting $(\xi_1, \eta_1) = dG(\xi, \eta)$, we have

$$\begin{aligned}\xi_1 \eta_1 &= \left(\frac{1}{x^2} + \frac{1}{x^2(\theta - 1/x)^2} - 1 \right) \left(1 + \frac{1}{(\theta - 1/x)^2} \right) \xi \eta \\ &\geq \left(\frac{1}{x^2(\theta - 1/x)^2} \right) \left(1 + \frac{1}{(\theta - 1/x)^2} \right) \xi \eta \\ &> 0.\end{aligned}$$

provided $\xi \eta \geq 0$. Hence dG preserves S^u . Moreover, if $\xi \eta \geq 0$, we have

$$|\xi_1| + |\eta_1| \geq \frac{1}{(\theta - 1/x)^2} (|\xi| + |\eta|)$$

so that, in the sum norm, dG expands vectors in S^u .

In a similar fashion, one verifies easily that dG^{-1} preserves and expands S^s . Note that the rate of expansion is $1/(\theta - 1/x)^2$, which tends to 1 as $\theta - 1/x$ tends to 1 from below. Hence G is non-uniformly hyperbolic.

Such non-uniformly hyperbolic mappings have been studied recently by Pesin [Pe]. The extension to mappings with singularities has been made by Katok and Strelcyn [KS]. We will not invoke their work, however, since our results are purely topological. Presumably, the work of Pesin and Katok/Strelcyn applies to show that G is ergodic.

Note that the spirals $\theta = 1/x$ and $\theta = -1/x$ partition the annulus into a grid of "rectangles" as in Fig. 2. The rightmost rectangle is actually a "triangle". We enumerate these rectangles from right to left by R_1, R_2, \dots as in Fig. 2.

For technical reasons, we include the two right-hand

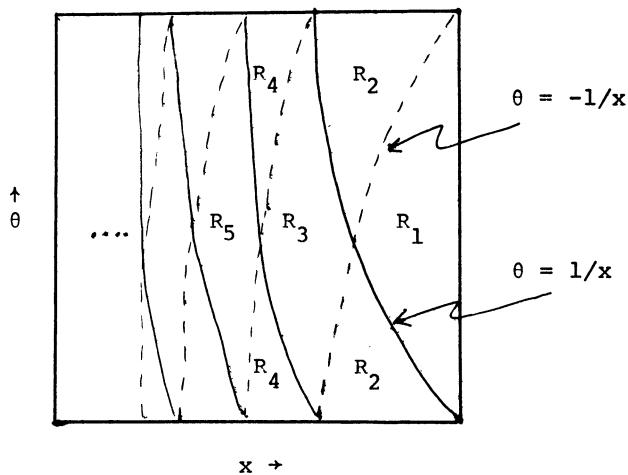


Fig. 2 Construction of the rectangles in $I \times S^1$.

boundaries in each R_j , so that the left hand boundaries of R_{j-1} are part of R_j . One can also view the R_j in the punctured disk as in Fig. 3.

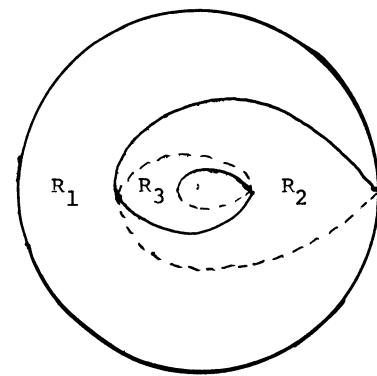


Fig. 3 Construction of the rectangles in the punctured disk.

Each rectangle R_i except R_1 is bounded by a pair of arcs in $\theta = 1/x$ and a pair in $\theta = -1/x$. We call the boundaries in $\theta = 1/x$ stable boundaries, while the boundaries in $\theta = -1/x$ are called unstable. G is undefined on stable boundaries, and G^{-1} is undefined on unstable boundaries.

Lemma 1. Let γ be a smooth curve in R_j connecting both stable boundaries and having tangents in the unstable sector bundle S^u . Then $G(\gamma)$ is a smooth spiral in $I \times S^1$ beginning at $x = 1$, spiralling down to $x = 0$, and meeting each rectangle R_j .

Proof. First observe that $G_1(\gamma)$ is a smooth curve connecting $\theta = 0$ to $\theta = 1$. Hence $G_2(G_1(\gamma))$ is a smooth curve connecting $x = 0$ to $x = 1$. Since G_1 is an infinite twist, it follows that $G(\gamma)$ is a smooth spiral connecting $x = 1$ to $x = 0$. Now the spiral $\theta = -1/x$ is not in the range of G . Hence $G(\gamma)$ cannot meet this curve. It therefore follows that $G(\gamma)$ must meet every rectangle. qed

Remark. Since the tangents to $G(\gamma)$ must also lie in the unstable sector S^u , it also follows that $G(\gamma)$ meets each R_i in an unstable curve which is "parallel" to the unstable boundaries.

Corollary 2. $G(R_j)$ is a spiral band in $I \times S^1$ which meets every rectangle. Moreover, $G(R_j) \cap R_k$ is a sub-rectangle of R_k with two boundaries contained in the stable boundaries of

R_k and the other two boundaries parallel to the unstable boundaries of R_k .

Proof. The proof is an immediate consequence of Lemma 1. See Fig. 4. qed

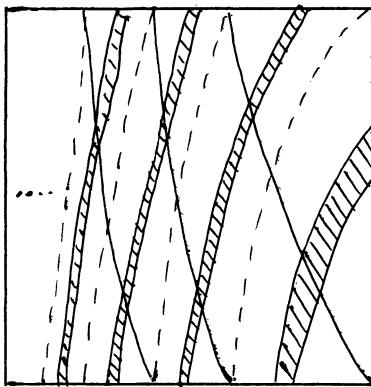


Fig. 4 $G(R_j)$ is a spiral band in $I \times S^1$. Note that $G(R_j)$ meets each R_k in a sub-rectangle.

For G^{-1} we have a similar situation. $G^{-1}(R_j)$ is a spiral band in $I \times S^1$ with boundary parallel to $\theta = 1/x$. See Fig. 5. Hence $G^{-1}(R_j)$ meets each R_k in a substrip with boundaries parallel to the stable boundaries of R_k . This of course gives the familiar Smale horseshoe construction. One verifies immediately that Axioms 1 and 2 of 2.1 hold on the interiors of the R_j , and so there is a subset Ω of $I \times S^1$ on which G is topologically conjugate to the shift automorphism on infinitely many symbols. Any point in Ω has the property that its orbit never meets either of the spirals

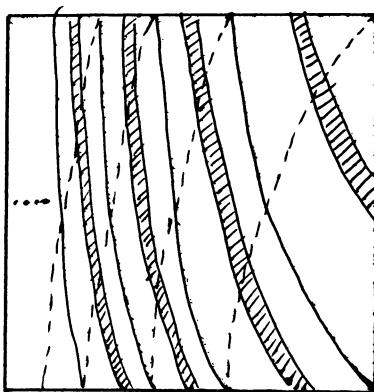


Fig. 5 $G^{-1}(R_j)$ is also a spiral band.

$$\theta = 1/x \text{ or } \theta = -1/x.$$

We can in fact extend this conjugacy to the singular orbits. We claim that G is topologically conjugate to the shift automorphism on the augmented symbol space Λ introduced in §3.2. Recall that Λ consists of all sequences of positive integers of the form $[\infty s_{-k} \dots s_0; s_1 \dots s_j \infty]$ where $s_0 \neq \infty$ and where j and/or k may be infinite. When j and k are infinite, we have a non-terminating sequence which corresponds to a point in Ω . All other sequences terminate on one or both ends and correspond to singular orbits as follows.

Let $p \in I \times S^1$ and suppose $F^j(p)$ is defined. The sequence $s(p) = (\dots s_0(p); s_1(p) s_2(p) \dots)$ is defined by $s_j(p) = \alpha$ iff $F^j(p) \in R_\alpha$. So $s_0(p)$ specifies which rectangle p originally lies in. We set $s_{j+1}(p) = \infty$ for $j \geq 0$ if $F^j(p)$ belongs to the spiral $\theta = 1/x$. Similarly, we set $s_{-k-1}(p) = \infty$ for $k \geq 0$ if $F^{-k}(p)$ belongs to the spiral

$\theta = -1/x$. So to each point in $I \times S^1$ we have associated a sequence in Λ .

Now recall that the shift automorphism σ on Λ is defined by $\sigma(\dots s_0; s_1 \dots) = (\dots s_0 s_1; s_2 \dots)$. So the domain of σ is all of Λ except "collision" sequences with $s_1 = \infty$, and the range of σ excludes "ejection" sequences with $s_{-1} = \infty$. Therefore, s maps points on $\theta = 1/x$ to the collision sequences and points on $\theta = -1/x$ to the ejection sequences. Moreover, we clearly have a conjugacy $\sigma \circ s = s \circ G$ on the respective domains.

We finally claim that s is a homeomorphism. That s is 1-1 and onto follows immediately from hyperbolicity. Hence we confine our attention to continuity of s . For this we need the topology on Λ . If (s^*) is a non-terminating sequence then we take as a neighborhood basis of (s^*) the usual cylinder sets. That is, (t) is close to (s^*) in this topology provided $t_i = s_i^*$ for $|i| \leq K$ with K large. Neighborhoods of terminating sequences of the form $(s^*) = (\dots s_0^*; s_1^* \dots s_j^* \infty]$ are somewhat different; they consist of two types of sequences. Either

$$\begin{aligned}s_i &= s_i^* & -K \leq i \leq j \\ s_{j+1} &\geq K\end{aligned}$$

with all other s_i arbitrary, or else

$$\begin{aligned}s_i &= s_i^* & -K \leq i \leq j-1 \\ s_j &= s_j^* - 1 \\ s_{j+1} &= 1\end{aligned}$$

and all other s_i are arbitrary. Note that this topology on Λ is slightly different from that introduced in §3.2.

Now s is clearly continuous at non-terminating sequences, so we will restrict our attention to sequences of the form $(s^*) = (\dots s_0^*; s_1^* \dots s_k^* \infty)$. Suppose $s(p) = (s^*)$. Therefore $F^k(p)$ lies on the stable boundary separating R_k and R_{k-1} . Let N be a neighborhood of p , and let $N_k = N \cap R_k$ and $N_{k-1} = N \cap R_{k-1}$. Note that $G_1(N_{k-1})$ abuts the lower boundary $\theta = 0$, while $G_1(N_k)$ abuts the boundary $\theta = 1$. Hence $G_2(G_1(N_{k-1}))$ abuts $x = 1$, while $G_2(G_1(N_k))$ abuts $x = 0$. Consequently, $G(N_k)$ is a spiral band converging to $x = 0$, and, in particular, meeting every R_ℓ for ℓ large. Similarly, $G(N_{k-1})$ is contained in R_1 .

It follows that, if $q \in N_k$, we have $s_{k+1}(q) \geq \ell$, whereas, if $q' \in N_{k-1}$, then $s_{k+1}(q') = 1$. This proves that the conjugacy is continuous.

§3.5 Classification of Motion in the Isosceles Three Body Problem.

The symbolic dynamics approach used in Gutzwiller's Theorem in §3.3 to classify solutions in the anisotropic Kepler problem can also be used in a natural way to classify motion in the isosceles three body problem. The most important ingredient in the proof of Gutzwiller's Theorem was a good understanding of the behavior of solutions which pass close to the singularity. In the isosceles problem, McGehee's coordinates

also give a good description of near-collision orbits, so the analogous theorem is easy to prove.

For the remainder of this section, we will assume that ϵ , the mass of the third particle, is smaller than the numerical value ϵ_0 computed by Simo [Sim 2]. In particular, this implies that the structure of the stable and unstable manifolds on the triple collision manifold is as in Fig. 7 of §1.7, and that the local behavior of solutions near triple collision is as given in Figs. 8 and 9 of that section.

We now mimic the statement of Gutzwiller's Theorem. Recall that the singular cross-section S_2 consists of all points in a fixed negative energy surface with $\theta = \pm\pi/2$.

S_2 is a union of two planes, S_2^+ and S_2^- . Heuristically, S_2 consists of all (regularized) initial conditions for which the primaries are exactly at collision. Let F be the Poincaré mapping defined on S_2 .

Let $p \in S_2$ and suppose $F^j(p)$ is defined. We define the j^{th} passage through p to be the orbit segment beginning at $F^j(p)$ and terminating at the next intersection with S_2 . So the 0^{th} passage through p is the orbit segment connecting p to the next crossing of S_2 along the orbit.

We can now introduce symbolic dynamics into the problem. If $r > 0$, any point in S_2 leads to an orbit of the system which is not a homothetic solution. Such an orbit must cross the q_1 -axis transversely (or not at all), for otherwise the orbit would agree with the homothetic orbit along $\theta = 0$. Since the Hill's region meets this axis in a compact interval, it follows that the number of intersections of the j^{th} passage

through any point with this axis is finite. Let $s_j = s_j(p)$ denote the number of such intersections.

As above, one checks easily that $s_j < \infty$. Unlike the anisotropic problem, however, s_j may equal 0. We therefore let the symbol space Σ consist of all doubly infinite sequences of non-negative integers of any of the following types:

$$\begin{aligned} & [\infty s_{-k} \dots s_0; s_1 \dots] \\ & (\dots s_0; s_1 \dots s_j \infty] \\ & [\infty s_{-k} \dots s_0; s_1 \dots s_j \infty] \end{aligned}$$

with $1 \leq j, k \leq \infty$. Again, $s_0 \neq \infty$. Then to each point p in S_2 , we assign the corresponding sequence $(s) = (s_j(p))$. If the orbit ends in collision after the j^{th} passage with $j \geq 0$, we let $s_{j+1} = \infty$ and terminate the sequence. If the orbit begins in ejection before the $-k^{\text{th}}$ passage with $k \geq 1$, we let $s_{-k-1} = \infty$ and again terminate the sequence. So sequences of the form $(\dots s_0; \infty]$ are immediate collision sequences while sequences of the form $[\infty s_0; s_1 \dots]$ are immediate ejection sequences. See Figs. 1 and 2.

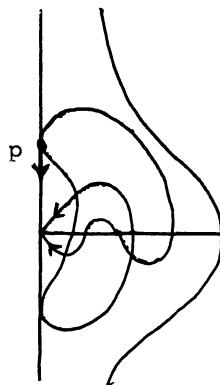


Fig. 1 The associated sequence $s(p) = [\infty 4 1; 2 \infty]$

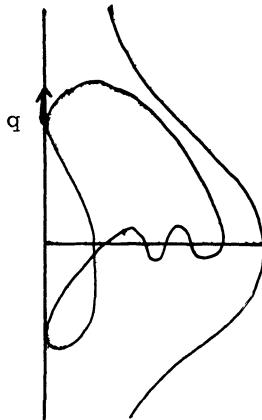


Fig. 2 The associated sequence $s(q) = (\dots 5 \ 1 \ 5; 1 \ 5 \ 1 \dots)$, corresponding to the closed orbit through q .

We topologize Σ as follows. For non-terminating sequences, we take the usual cylinder sets as neighborhood bases. If $(s^*) = (\dots s_0^*; s_1^* \dots s_j^* \infty]$ is a terminating sequence, then we take as neighborhoods $N_{K,L}$ of (s^*) all sequences (s) of one of the following two types. Either

$$\begin{aligned} s_i &= s_i^* & -K \leq i \leq j \\ s_{j+1} &> L \end{aligned}$$

or

$$\begin{aligned} s_i &= s_i^* & -K \leq i \leq j-1 \\ s_j &= s_j^* - 1 \\ s_{j+N} &= 0 & 1 \leq N \leq L. \end{aligned}$$

Note that these neighborhoods are quite different from the corresponding neighborhoods in the anisotropic problem.

The rationale for this topology rests on our discussion

of the local behavior of solutions near a collision orbit.

Recall from §1.7 that near-collision orbits behave in two dramatically different ways depending upon which branch of the unstable manifold in Λ the solution follows. These are illustrated in Figs. 3 and 4.

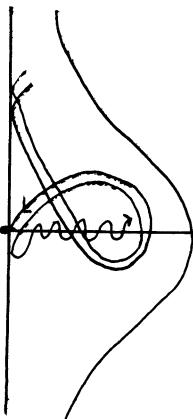


Fig. 3 The nearby orbit has sequence $(\dots s_0; \dots s_j K \dots)$ with K large.

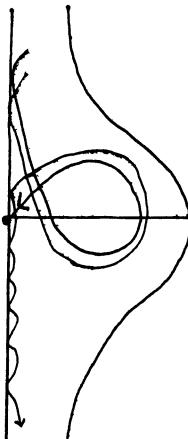


Fig. 4 The nearby orbit has sequence $(\dots s_0; \dots s_{j-1} 1 0 0 0 \dots)$

Using geometric methods similar to those in §3.3, one can prove the following: let $s: S_2 \rightarrow \Sigma$ be the mapping which associates to points in S_2 the corresponding sequence in Σ . On Σ we have the usual shift automorphism $(\sigma(s))_j = s_{j+1}$. The shift is again a homeomorphism defined on $\Sigma - C$, where C represents the immediate collision sequences. Our main result is then

Theorem 1. The correspondence s is a continuous surjection and we have the following commutative diagram:

$$\begin{array}{ccc} S_2 - C & \xrightarrow{F} & S_2 - E \\ s \downarrow & & \downarrow s \\ \Sigma - \hat{C} & \xrightarrow{\sigma} & \Sigma - \hat{E} \end{array}$$

Remark 1. Unlike the anisotropic Kepler problem, the Hill's region for the isosceles problem is non-compact. This means that open sets of orbits of the system can escape (or be captured). At present, the complete structure of these sets is not well understood. However, one can prove that any sequence which terminates to the right (resp. left) with an infinite string of zeroes is an escape (resp. capture) orbit. By the theorem above, these orbits exist.

Remark 2. Again unlike the anisotropic problem, there is no conjecture that the projection s is injective (even up to symmetries). One in fact expects that open sets of orbits escape.

Remark 3. Our work on the isosceles problem is very similar to

the well-known work of Sitnikov and Alekseev [Al]. They study the three dimensional restricted three body problem in which the primaries revolve on Keplerian ellipses and the third mass travels along the axis of symmetry. They classify solutions of this system which oscillate far above and below the center of mass of the primaries. Their methods are similar to ours in the sense that they study a mapping on a cross-section which turns out to be a Smale horseshoe. A nice treatment of their work is contained in Moser's book [Mo].

References

- [AA] Arnold, V. and Avez, A. Ergodic Problems in Classical Mechanics. New York: Benjamin, 1968.
- [Al] Alekseev, V. Quasirandom dynamical systems, I,II,III. Math. USSR Sbornik 6 (1969), 489-498.
- [AM] Abraham, R. and Marsden, J. Foundations of Mechanics. Reading, Mass.: Benjamin/Cummings, 1978.
- [Ar] Arnold, V. Mathematical Methods of Classical Mechanics. New York: Springer-Verlag, 1978.
- [Bo] Bowen, R. On Axiom A Diffeomorphisms. CBMS Regional Conference Series 38 (1978).
- [Br] Broucke, R. On the isosceles triangle configuration in the planar general three body problem. Astron. Astrophys. 73 (1979), 303-313.
- [De 1] Devaney, R. Collision orbits in the anisotropic Kepler problem. Inventiones Math. 45 (1978), 221-251.
- [De 2] . Non-regularizability of the anisotropic Kepler problem. J. Diff. Equations. 29 (1978), 253-268.
- [De 3] . Transverse heteroclinic orbits in the anisotropic Kepler problem. In The Structure of Attractors in Dynamical Systems. Springer-Verlag Lecture Notes in Math No. 668. New York: Springer-Verlag (1978), 67-87.
- [De 4] . Morse-Smale singularities in simple mechanical systems. To appear in J. Diff. Geom.
- [De 5] . Triple collision in the planar isosceles three body problem. To appear in Inventiones Math.
- [De 6] . The baker transformation and a mapping associated to the restricted three body problem. To appear.
- [De 7] . Three area preserving mappings exhibiting stochastic behavior. To appear in Classical Mechanics and Dynamical Systems. New York: Marcel Dekker.
- [DN] , and Nitecki, Z. Shift automorphisms and the Hénon mapping. Commun. Math. Phys. 67 (1979) 137-146.

- [Ea 1] Easton, R. Isolating blocks and symbolic dynamics.
J. Diff. Equations 17 (1975), 96-118.
- [Ea 2] _____ . Regularization of vector fields by surgery.
J. Diff. Equations 10 (1971), 92-99.
- [EM] _____ , and McGehee, R. Homoclinic phenomena for orbits doubly asymptotic to an invariant three sphere.
To appear.
- [Eu] Euler, L. De motu rectilineo trium corporum se mutuo attrahentium. *Novi Comm. Acad. Sci. Imp. Petrop.* 11 (1767), 144-151.
- [Fr] Franks, J. Homology and Dynamical Systems. To appear in CBMS Regional Conference Series.
- [Gu 1] Gutzwiller, M. The anisotropic Kepler problem in two dimensions. *J. Math. Phys.* 14 (1973), 139-152.
- [Gu 2] _____ . Bernoulli sequences and trajectories in the anisotropic Kepler problem. *J. Math. Phys.* 18 (1977), 806-823.
- [Gu 3] _____ . Periodic orbits in the anisotropic Kepler problem. To appear.
- [H1] Hénon, M. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* 50 (1976), 69-77.
- [H2] _____ . Etude générale de la transformation. Mimeo-graphed notes.
- [KS] Katok, A. and Strelcyn, J.-M. Invariant manifolds for smooth maps with singularities. To appear.
- [La] Lagrange, J.L. Oeuvres. Vol. 6. Paris (1873), 272-292.
- [L1] Lacomba, E. and Losco, L. Triple collision in the isosceles three body problem. *Bull. AMS* 3 (1980), 710-714.
- [L2] _____ . Quadruple collision in the trapezoidal four body problem. To appear in Classical mechanics and dynamical systems. New York: Marcel Dekker.
- [MM] Mather, J. and McGehee, R. Solutions of the four body problem which become unbounded in finite time. Lecture Notes in Physics. Vol. 38. Springer-Verlag (1975), 573-597.
- [McG1] McGehee, R. Triple collision in the collinear three body problem. *Inventiones Math.* 27 (1974), 191-227.

- [McG2] McGehee, R. Recent Developments in Celestial Mechanics.
To appear in CBMS Regional Conference Series.
- [Mo] Moser, J. Stable and Random Motions in Dynamical Systems.
Annals of Math Studies, No. 77. Princeton Univ. Press,
(1973).
- [Ni] Nitecki, Z. Differentiable dynamics. MIT Press, (1971).
- [Pe] Pesin, Ja. B. Families of invariant manifolds corresponding to non-zero characteristic exponents. Math. USSR Izvestija 10 (1976), 1261-1305.
- [Po] Pollard, H. Mathematical Introduction to Celestial Mechanics. Carus Mathematical Monographs, No. 18. Math. Asso. Amer., (1976).
- [Sa] Saari, D. Singularities and collisions of Newtonian gravitational systems. Arch. Rational Mech. Anal. 49 (1973), 311-320.
- [SM] Siegel, C. and Moser, J. Lectures on Celestial Mechanics. Berlin: Springer-Verlag (1971).
- [Sim 1] Simo, C. Analysis of triple collision in the isosceles problem. To appear in Classical Mechanics and Dynamical Systems. New York: Marcel Dekker.
- [Sim 2] _____. Masses for which triple collision is non-regularizable. To appear in Celestial Mechanics.
- [Sm] Smale, S. Diffeomorphisms with many periodic points. In Differential and Combinatorial Topology. Princeton University Press (1965), 63-80.
- [Su] Sundman, K. Mémoire sur le problème des trois corps. Acta Math. 36 (1912), 105-179.
- [Sz] Szebehely, V. Instabilities in Dynamical Systems. Dordrecht: D. Reidel (1979).
- [W] Waldvogel, J. Stable and unstable manifolds in planar triple collision. In Instabilities in Dynamical Systems. Dordrecht: D. Reidel (1979).
- [Wi] Wintner, A. The Analytical Foundations of Celestial Mechanics. Princeton University Press (1941).

Robert L. Devaney
Department of Mathematics
Boston University
Boston, MA 02215