Insolubility of Trapped Particle Motion in a Magnetic Dipole Field

ALEX J. DRAGT AND JOHN M. FINN[1]

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

Topological and numerical techniques are used to show that the problem of trapped charged particle motion in a magnetic dipole field is insoluble. Similar results hold for motion in the earth's magnetic field and are of interest for radiation belt phenomena. Pedagogical discussion is devoted to the subject of how it can happen that a classical mechanics problem is insoluble and in what sense. It is shown that the complete adiabatic magnetic moment series is divergent and that due to the existence of homoclinic points the solutions to the equations of motion are too complicated to be written in closed form. As a consequence, there is currently no rigorous theoretical explanation for the empirical success of adiabatic orbit theory, and a completely satisfactory mathematical justification will be far from easy.

1. INTRODUCTION

An essential ingredient for an understanding of magnetospheric and radiation belt phenomena is a knowledge of the motions of charged particles trapped in the earth's magnetic field. Trapped orbits can be approximately described in terms of the so-called magnetic moment, longitudinal, and flux invariants [*Northrop*, 1963; *Northrop and Teller*, 1960]. These quantities are not true invariants because they are not actual constants of motion over the course of an orbit. However, they are 'adiabatically invariant' in that they remain approximately constant for moderate lengths of time for low-energy orbits. Consequently, they can be used to predict particle motion only for moderately short times. In actual practice it is almost always tacitly assumed that the adiabatic invariant predictions hold good for arbitrarily long times. Such an assumption may be correct, but its use certainly requires discussion and eventual mathematical justification.

In view of the uncertainties associated with the adiabatic invariant predictions, it would be highly desirable to have an exactly soluble model problem which in some sense approximates the true problem. One obvious possibility is to replace the earth's actual magnetic field by that of a perfect dipole. We shall call the problem of determining the orbits in a pure magnetic dipole field the Størmer problem in honor of *Størmer* [[illegible]], who first considered it [*Rossi and Olbert*, 1970]. If the Størmer problem could be solved exactly, we could hope to treat the full problem by perturbation methods.

On the basis of past experience in other areas of physics, we might naively hope that the simplifications and symmetry induced by a pure dipole field do indeed lead to an exactly soluble problem. For example, the motion of a satellite about the earth becomes the exactly soluble Kepler problem if we simplify the earth's gravitational field by ignoring quadrupole and higher-moment terms. However, it has been well known to celestial mechanicians since the time of *Pioncaré* [1892] and others [*Whittaker*, 1937] that 'most' classical mechanics problems are 'insoluble.' The purpose of this paper is to show that the Størmer problem for trapped orbits belongs to this insoluble majority. We hope that our warning will spare aspiring graduate students and others the expense of spending long and fruitless hours in the hope that to the amazement of all, they might discover just the right canonical transformation which leads to an exact solution and lasting fame. At the same time, we hope that our discussion will prove instructive as to why and in what sense a classical mechanics problem can be insoluble, since this subject is not common knowledge among physicists and is rarely touched upon in mechanics courses or texts. Finally, of more interest from a practical point of view, we will show that insolubility implies that the complete adiabatic magnetic moment series diverges. Consequently, there is at present no rigorous mathematical explanation for the empirical success of adiabatic orbit theory in explaining magnetospheric phenomena and correlating spacecraft data. Rather, its success must currently be viewed at the very least as a plausible minor miracle.

We have tried to make our exposition as simple and nontechnical as possible. More detailed calculations will be presented elsewhere. Section 2 describes briefly the equations of motion for the Størmer problem and the use of dipolar coordinates. In section 3 we convert the Størmer problem into the determination of an area preserving map M. Section 4 describes various properties of area-preserving mappings including the possible existence of a 'homoclinic' point. Section 5 shows that if M has a homoclinic point, then the Størmer problem is insoluble. Here we also discuss what 'insolubility' means. We show in section 6 by direct numerical integration that the map M has homoclinic points. Our results and the current state of the Størmer problem are summarized in a final section.

2. EQUATIONS OF MOTION AND COORDINATES

A portion of a typical trapped Størmer orbit is illustrated in Figure 1. For a nonrelativistic particle of charge q and mass m the orbit is generated by the Hamiltonian

$$H = (\tfrac{1}{2}m)\{p_z^2 + p_\rho^2 + [(p_\phi/\rho) - qA_\phi]^2\} \qquad (1)$$

where ρ, z, ϕ are cylindrical coordinates and the vector potential $\mathbf{A}$ describing the dipole field is given by

$$\mathbf{A} = \hat{\phi}\mathfrak{M}\rho r^{-3} \qquad (2)$$

with $r^2 = \rho^2 + z^2$. Here $\mathfrak{M}$ is the magnitude of the dipole moment. (The relativistic case is also described by (1) if m is replaced by γm.) The motion consists of three parts: a gyration about a field line, a bouncing back and forth across the equatorial plane along the line between mirror points to form a kind of spiral, and a slow drift about the earth. For a fuller discussion of the motion and a fuller exposition of some of the points which are to follow, we refer the reader to an earlier article [*Dragt*, 1965].

[1] Now at Plasma Physics Laboratory, Princeton University, Princeton, New Jersey 08540

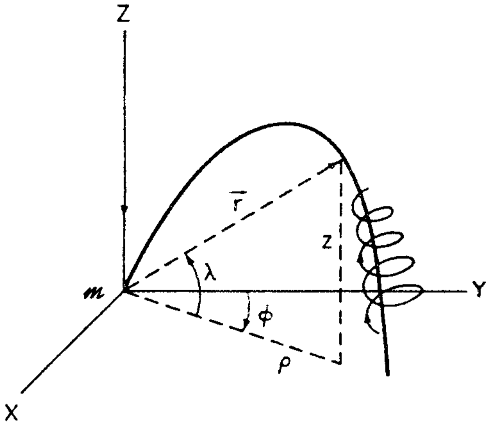

Fig. 1. Motion of a trapped particle.

Inspection of (1) and (2) shows that H is independent of ϕ as we might well expect from the axial symmetry of the problem. It follows that

$$\dot{p}_\phi = -\partial H/\partial\phi = 0 \tag{3}$$

and hence p_ϕ is a constant of motion. It can conveniently be written in the form

$$p_\phi = q\mathfrak{M}\Gamma \tag{4}$$

where Γ is an integration constant having the dimensions of a reciprocal length.

An analysis of the properties of H shows that trapped orbits can occur only if $\Gamma > 0$ and that in this case, particles with sufficiently low energy spiral about a field line satisfying

$$r = \Gamma^{-1}\cos^2\lambda \tag{5}$$

where, as in Figure 1, λ denotes geomagnetic latitude.

Applying Hamilton's equations we find

$$\dot{\phi} = \partial H/\partial p_\phi = (p_\phi - q\rho A_\phi)/(m\rho^2) \tag{6}$$

As far as the motion in ρ and z is concerned, we note that we may replace p_ϕ in H by its value given in (4). After this substitution is made, we can regard $H(p_\rho p_z; \rho z; p_\phi = q\mathfrak{M}\Gamma)$ as a 'reduced' Hamiltonian describing two-dimensional motion in the ρ, z plane. Once this motion is determined to give $\rho(t)$ and $z(t)$, we can easily determine $\phi(t)$ simply by integrating (6):

$$\phi(t) = \int dt\, (q\mathfrak{M}\Gamma - q\rho A_\phi)/(m\rho^2) \tag{7}$$

Therefore in what follows we shall concentrate on finding $\rho(t)$ and $z(t)$.

At this point it is convenient to introduce dimensionless space and time variables ρ', z', t' by the rules

$$\begin{aligned} z' &= z\Gamma \\ \rho' &= \rho\Gamma \\ t' &= t\Gamma^3 q\mathfrak{M}/m \end{aligned} \tag{8}$$

In these variables the trapped particle gyrates about a guiding field line obeying

$$r' = \cos^2\lambda \tag{9}$$

Also the particle has unit cyclotron frequency when it is in the equatorial plane. The motion is governed by the dimensionless Hamiltonian

$$H = \tfrac{1}{2}(p_z^2 + p_\rho^2) + V(\rho, z) \tag{10}$$

where

$$V(\rho, z) = \tfrac{1}{2}[(1/\rho) - (\rho/r^3)]^2 \tag{11}$$

For notational convenience we have omitted the primes in (10) and (11), and we shall continue to do so henceforth. All future references will be to dimensionless variables and dimensionless equations of motion.

We see from (10) that the motion in the ρ, z plane is the same as that of an imaginary particle of unit mass moving in the effective two-dimensional potential $V(\rho, z)$. Therefore we can get a qualitative picture of possible orbits by examining it. Figure 2 is a contour map showing level lines of V. The potential vanishes on the floor of the valley given by the line

$$\rho^2 = r^3 \tag{12}$$

usually called the thalweg (German: 'valley way'), and is positive elsewhere. (Note that (12) is the same as the guiding field line (9)). Also the walls of the potential become steeper as one proceeds from the equator, $z = 0$, down the thalweg toward either pole. Finally, there is a pass at $z = 0$, $\rho = 2$, where V has the value $\frac{1}{32}$.

From energy conservation we conclude that all orbits which begin in the valley and which satisfy $W_0^2 < \frac{1}{16}$, where W_0 is a dimensionless 'velocity' defined by

$$W_0^2 = 2H \tag{13}$$

cannot escape to infinity. These are the orbits which we have called 'trapped.'

Figure 3 shows a typical portion of a trapped orbit in the ρ, z plane obtained by numerical integration. The motion consists of oscillations about the thalweg superimposed upon motion along the thalweg. Because the walls of the potential become steeper as one proceeds down the thalweg, our imaginary particle experiences a retarding force. Consequently, the orbit eventually turns around. The oscillations about the thalweg correspond to gyrations about the guiding field line in the full three-dimensional orbit, and the motion back and forth along the thalweg corresponds to the bouncing motion between mirror points.

Since in first approximation the motion in the ρ, z plane consists of oscillations about the thalweg superimposed upon motion along the thalweg, it is convenient to take this feature into account by the introduction of orthogonal 'dipole' coordinates q_1 and q_2. They are defined in terms of ρ and z by the relations

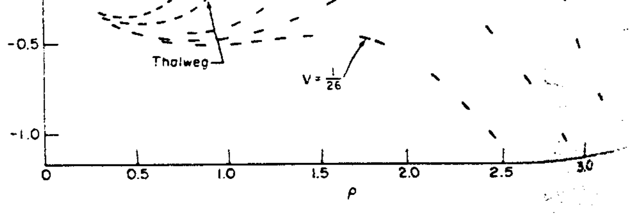

Fig. 2. Level lines of the effective potential $V(\rho, z)$.

$$q_1 = z/r^3 \qquad q_2 = (r^3/\rho^2) - 1 \tag{14}$$

Lines of constant q_2 correspond to dipole field lines. In particular, the line $q_2 = 0$ corresponds to the thalweg. Lines of constant q_1 are orthogonal to lines of constant q_2. Figure 4 shows the orbit of Figure 3 as it appears in terms of dipolar coordinates. The motion has now been separated out, in first approximation, into oscillations about $q_2 = 0$ superimposed upon motion along the q_1 axis.

We close this section by noting the form of the Hamiltonian in dipolar coordinates. Let p_1 and p_2 be momenta canonically conjugate to q_1 and q_2. Then after calculation one finds that the Hamiltonian is given (in mixed variables) by

$$H = \tfrac{1}{2}[(p_1/h_1)^2 + (p_2/h_2)^2 + r^{-3}(q_2 + 1)^{-1}q_2^2] \tag{15}$$

where

$$h_1^2 = r^8/(r^2 + 3z^2) \tag{16}$$

$$h_2^2 = \rho^6 r^{-4}/(r^2 + 3z^2) \tag{17}$$

In particular, for future reference we observe that H for trapped orbits can be expanded in a power series in the q's and p's of the form [*Contopoulos and Vlahos*, 1975]

$$H = H_2 + H_3 + H_4 + \cdots \tag{18}$$

where

$$H_2 = \tfrac{1}{2}(p_1^2 + p_2^2 + q_2^2) \tag{19}$$

$$H_3 = -2q_2^3 - 3q_2p_1^2 \tag{20}$$

$$H_4 = (\tfrac{1}{4})(6p_2^2q_1^2 + 21q_2^2p_1^2 + 10q_2^4 + 3q_1^2q_2^2 + 9q_1^2p_1^2) \tag{21}$$

3. CONVERSION INTO A MAPPING PROBLEM

The purpose of this section is to explain how the particle motion described in the previous section can be used to generate a mapping M. Our method involves the use of what is called a surface of section and dates back to Poincaré's celebrated work on celestial mechanics.

Consider the four-dimensional phase space consisting of the variables ρ, z, p_ρ, p_z. From Hamiltonian mechanics we know that *every orbit* in the configuration space ρ, z corresponds to a trajectory in phase space. Furthermore, there is a unique trajectory through each point in phase space, and trajectories never intersect unless they happen to close on themselves.

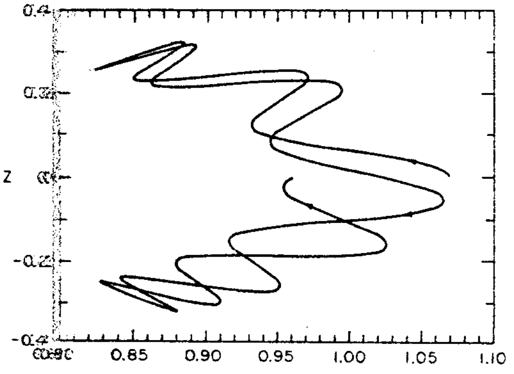

Fig. 3. A trapped orbit as seen in ρ, z coordinates. The initial conditions are $z = 0$, $\rho = 1.07$, $\dot\rho = 0$, $\dot z = 0.0355$, $W_0^2 = 0.005$.

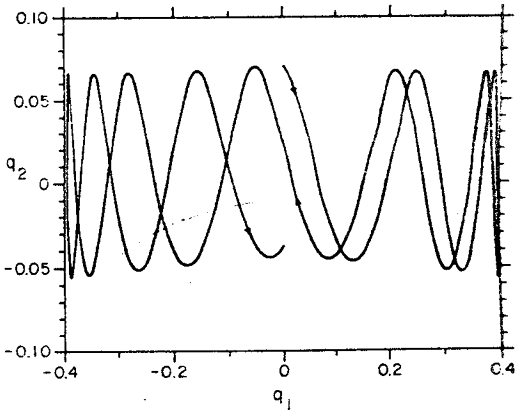

Fig. 4. The orbit of Figure 3 as it appears in dipolar coordinates.

Since $W_0^2 = 2H$ is a constant of motion, it is convenient to group together all trajectories with the same W_0^2. They evidently lie on a three-dimensional hypersurface. Now consider all trajectories which correspond to trapped orbits in configuration space and which have a fixed value of W_0^2. From our discussion in section 2 of the nature of orbits, it is intuitively obvious that each such phase space trajectory, when it is extended far enough forward or backward in time, must cross the hyperplane $z = 0$. That is, all orbits must cross the equatorial plane at least once. A rigorous proof has been given by *DeVogelaere* [1954]. Put another way, the 'surface' $z = 0$ cuts across every (trapped) trajectory and thus may be called a surface of section.

Let us record the values of ρ, $\dot\rho = p_\rho$, and $\dot z = p_z$ at the moment of crossing. Their values and the equations of motion derived from H allow a complete reconstruction of the whole trajectory. Futhermore, since we have agreed to fix W_0^2, the value of $\dot z$ is redundant because by solving (13) for $\dot z$ when $z = 0$ we find

$$\dot z = [W_0^2 - \dot\rho^2 - 2V(\rho, 0)]^{1/2} \tag{22}$$

Thus a trajectory is specified by the two numbers $\rho|_{z=0}$ and $\dot\rho|_{z=0}$, and the surface of section is effectively two-dimensional. There is, of course, an ambiguity in the sign of $\dot z$ as given by (22). So we should really say that each ρ, $\dot\rho$ pair in general specifies two trajectories, one with $\dot z|_{z=0} > 0$ and one with $\dot z|_{z=0} < 0$.

However, we observe that V is symmetric about the ρ axis: $V(\rho, z) = V(\rho, -z)$. Consequently if $[\rho(t) = f(t), z(t) = g(t)]$ is an orbit, so is its mirror image $[\rho(t) = f(t), z(t) = -g(t)]$. Thus the two possible choices of sign for $\dot z$ amount to choices between two mirror images.

We also need to make one further clarification. It is obvious from (11) and (22) that only a certain region of the ρ, $\dot\rho$ plane leads to real values of $\dot z$ and that outside this region, $\dot z$ is pure imaginary. The region of real $\dot z$, which we shall call the physical region, is the boundary and interior of the curve

$$\dot\rho^2 + [1/\rho - 1/\rho^2]^2 = W_0^2 \tag{23}$$

The boundary itself gives vanishing values of $\dot z$, and therefore all points on the boundary are points of that orbit which is confined to the ρ axis in configuration space. (This orbit corresponds to an equatorial orbit in the ρ, z, ϕ space.) Points inside the boundary correspond to orbits which leave the ρ axis and extend into either the northern or southern hemispheres.

We have seen that a ρ, $\dot\rho$ pair in the physical region in general specifies two orbits. We shall see now how the equations of motion can be used to generate a mapping M of the physical region into itself [*DeVogelaere*, 1958; *Godart*, 1970]: Select a ρ, $\dot\rho$ point in the interior of the physical region, and compute $\dot z$ by using (22) with a positive square root. Employing these values (plus $z = 0$) as initial conditions, compute an orbit. Analysis shows that this orbit will either go to the origin of the dipole by way of the northern hemisphere or it will turn around and recross the ρ axis; i.e., $z = 0$. If the trajectory recrosses the ρ axis, record the crossing values ρ', $\dot\rho'$. Now if the orbit specified by the initial values ρ, $\dot\rho$ does not go to the origin, then we may define a mapping M by the rule

$$M\colon\quad (\rho, \dot\rho) \rightarrow (\rho', \dot\rho') \tag{24}$$

For example, the orbit in Figure 3 was launched with $W_0^2 = 0.005$, $z = 0$, $\rho = 1.070$, $\dot\rho = 0$ which, by using (22) gives $\dot z = 0.0355$. It returned to $z = 0$ with $\rho = 1.023$ and $\dot\rho = 0.053$. Thus we have for M the result

$$M\colon\quad (1.070, 0.0) \rightarrow (1.023, 0.53)$$

Figure 5 shows the physical region in the ρ, $\dot\rho$ plane for $W_0^2 = 0.005$ and the launch values of ρ, $\dot\rho$. We also show the action of M and its powers for the above example.

To continue our discussion, we use the fact that the orbit to the origin by way of each hemisphere is unique [*Braun*, 1970*a*]. Therefore M has been defined everywhere within the interior of the physical region except for one point. Let O be the ρ, $\dot\rho$ point which generates the orbit to the origin. This trajectory never returns. However, we can extend the definition of M to O by defining the return orbit to be the outgoing orbit retraced backwards. Thus if O has coordinates ρ_0 and $\dot\rho_0$, the point MO has coordinates ρ_0 and $-\dot\rho_0$. It can be shown that this extension preserves continuity. Finally, we extend the definition of M to points on the boundary of the physical region by again invoking continuity. In conclusion, we have defined a mapping M of the entire physical region into itself. In particular, the boundary is mapped into itself, and the interior is mapped into itself.

In defining the mapping M from ρ, $\dot\rho$ to ρ', $\dot\rho'$ we used the positive square root in (22). Let us call the orbit thus generated $[\rho_+(t), z_+(t)]$. If t_R is the time at which the orbit again returns to cross the ρ axis, we have

$$\begin{aligned} z_+(t_R) &= 0 \\ \rho_+(t_R) &= \rho' \\ \dot\rho_+(t_R) &= \dot\rho' \end{aligned} \tag{25}$$

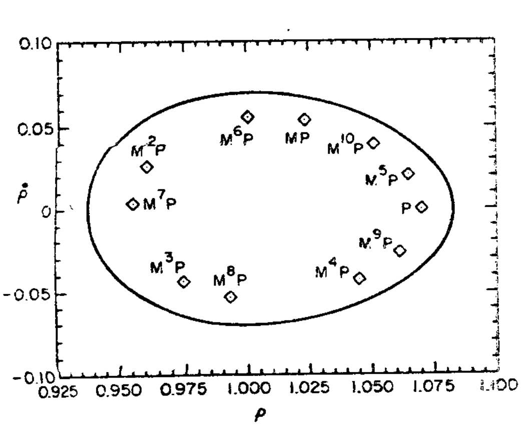

Fig. 5. An illustration of the action of M in the ρ, $\dot\rho$ plane when $W_0^2 = 0.005$. The oval curve is the boundary of the physical region. The point P has coordinates (1.07, 0) corresponding to the launch values of ρ, $\dot\rho$, and M^nP denotes the result of n actions of M on P.

Suppose we had used the negative sign instead. Then we would have generated the orbit $[\rho_-(t), z_-(t)]$, where according to our earlier discussion,

$$\begin{aligned} \rho_-(t) &= \rho_+(t) \\ z_-(t) &= -z_+(t) \end{aligned} \tag{26}$$

In particular, at $t = t_R$ we find

$$\begin{aligned} z_-(t_R) &= -z_+(t_R) = 0 \\ \rho_-(t_R) &= \rho_+(t_R) = \rho' \\ \dot\rho_-(t_R) &= \dot\rho_+(t_R) = \dot\rho' \end{aligned} \tag{27}$$

Thus the definition of M is independent of the sign of the square root.

What is M good for? In brief, a knowledge of M is equivalent to a knowledge of the long-time behavior of orbits. For suppose, as in Figure 5, we launch an orbit from some point P in the physical region into, say, the northern hemisphere ($\dot z$ > 0). The orbit will return through the surface of section at the point MP. It will then continue on into the southern hemisphere only to return again through the point MMP. Further successive crossings generate the points M^3P, M^4P, etc. Thus a description of the behavior of M^n for large n is equivalent to a description of the long-time behavior of all orbits.

4. Properties of Area-Preserving Maps

The mapping M has three simple properties which follow almost directly from its definition and which also serve as starting points for deeper investigation:

1. M has an inverse M^{-1}. M also is continuous and differentiable. That is, if P is some point in the ρ, $\dot\rho$ plane and $Q = MP$ is its image under M, then small changes in P result in small changes in Q. In fact, if P is not the point O which leads to the origin, then the coordinates of $Q = MP$ can be differentiated with respect to the coordinates of P. These assertions follow from the fact that the solution to a differential equation depends continuously and differentiably on the initial conditions provided the solution does not pass through a point (such as the origin) where the differential equation contains singular terms. The existence of the inverse mapping M^{-1} follows from the observation that an orbit can always be traced backwards in time.

2. M is area preserving. This means that if R is a region in the ρ, $\dot\rho$ plane and if R' is its image under the action of M, then R and R' have the same area. The result is a consequence of the fact that the equations of motion are derivable from a Hamiltonian [*Arnold and Avez*, 1968].

3. M has fixed points. A fixed point is a point which is sent into itself under the action of M. Evidently, fixed points of M, or of powers of M, correspond to periodic orbits in the ρ, $\dot\rho$ plane. The existence of such orbits is intuitively obvious and can be proven rigorously [*DeVogelaere*, 1958].

Properties (1) and (2) can be combined to give a classification of the fixed points discussed in property (3). Our discussion will apply to area-preserving maps in general. To make our notation more precise, we shall now denote points in the ρ, $\dot\rho$ plane by specifying a two-component vector. Let $\mathbf{a}$ be a point in the physical region and let $\mathbf{b}$ be its image under the action of M. We write $\mathbf{b} = M\mathbf{a}$. Next consider the action of M on the nearby point $\mathbf{a} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ denotes a small vector. Since M is differentiable, we may expand $M(\mathbf{a} + \boldsymbol{\epsilon})$ in a power series in the components of $\boldsymbol{\epsilon}$ to get an expression of the form

$$M(\mathbf{a} + \epsilon) = \mathbf{b} + L_a\epsilon + O(\epsilon^2) \tag{28}$$

where L_a denotes a 2 × 2 matrix. We shall call L_a the linear part of M at the point $\mathbf{a}$.

The matrix L_a has two important properties. First, it is obviously real, since M acts on real vectors to produce real vectors. Secondly, L_a has unit determinant

$$\det L_a = 1 \tag{29}$$

This result follows from property (2): Let us consider some small region R centered around $\mathbf{a}$ and transform it to a small region R' centered around $\mathbf{b}$ by the application of M. Then the ratio of the areas of these two regions is given by the Jacobian of M which is just det L_a. Since M is area preserving, the ratio of the areas must be 1, and hence (29) follows.

Now suppose $\mathbf{a}$ is a fixed point; i.e., $M\mathbf{a} = \mathbf{a}$. Then for any nearby point $\mathbf{a} + \epsilon$ we have

$$M(\mathbf{a} + \epsilon) = \mathbf{a} + L_a\epsilon + O(\epsilon^2)$$

$$M^2(\mathbf{a} + \epsilon) = M(\mathbf{a} + L_a\epsilon) + O(\epsilon^2) = \mathbf{a} + L_a^2\epsilon + O(\epsilon^2)$$

and in general for any power,

$$M^n(\mathbf{a} + \epsilon) = \mathbf{a} + L_a^n\epsilon + O(\epsilon^2) \tag{30}$$

We see that in first approximation the behavior of M^n at a fixed point is governed by its linear part. Therefore we should explore what can be said about L_a and its powers.

The behavior of L_a is characterized by its eigenvectors and eigenvalues. Let us call the eigenvalues λ_1 and λ_2. Then in view of (29) we must have

$$\lambda_1 \lambda_2 = 1 \tag{31}$$

because the determinant of a matrix equals the product of its eigenvalues. Also if λ_1 happens to be complex, λ_2 must be its complex conjugate, since L_a and hence its characteristic equation are real. Combining these two properties, we see after a little analysis that there are only five possibilities:

1. According to the hyperbolic possibility, one eigenvalue, say λ_1, is real and greater than 1. Then we have $\lambda_1 = \lambda$ and $\lambda_2 = 1/\lambda$ with $\lambda > 1$.

2. According to the elliptic possibility, both eigenvalues are complex and lie on the unit circle. Then $\lambda_1 = e^{i\phi}$ and $\lambda_2 = e^{-i\phi}$ where ϕ is some real angle different from zero or a multiple of π.

3. According to the inversion hyperbolic possibility, both eigenvalues are real and negative. Then $\lambda_1 = -\lambda$ and $\lambda_2 = -1/\lambda$ with $\lambda > 1$.

4. According to the parabolic possibility, both eigenvalues equal +1.

5. According to the inversion parabolic possibility, both eigenvalues equal −1.

For our purposes we shall be particularly interested in the first possibility, the hyperbolic case.

Suppose $\mathbf{h}$ is a hyperbolic fixed point. That is, L_h is hyperbolic. We are going to explore the effect of powers of M on nearby points $\mathbf{h} + \epsilon$. In view of (30) we begin by studying the effects of L_h^n on ϵ. Since $\mathbf{h}$ is hyperbolic, λ_1 and λ_2 are real and different. Therefore L_h has two real linearly independent eigenvectors which we denote by $\mathbf{v}_1$ and $\mathbf{v}_2$. Let us expand ϵ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$ by writing

$$\epsilon = \epsilon_1 \mathbf{v}_1 + \epsilon_2 \mathbf{v}_2 \tag{32}$$

It follows that

$$L_h^n\epsilon = \epsilon_1\lambda_1^n\mathbf{v}_1 + \epsilon_2\lambda_2^n\mathbf{v}_2 \tag{33}$$

We see that if ϵ has components ϵ_1 and ϵ_2, then $L_h^n\epsilon$ has components $\epsilon_1\lambda_1^n$ and $\epsilon_2\lambda_2^n$. In particular, in view of (31), the product of the components of $L_h^n\epsilon$ is $\epsilon_1\epsilon_2$, independent of the value of n. Let us regard the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ as a set of (usually) oblique axes. Then the set of points given by (32) with the product $\epsilon_1\epsilon_2$ put equal to a constant is easily recognized as a hyperbola. It follows from (33) that the points $L_h^n\epsilon$ for fixed ϵ and variable n all lie on the same hyperbola and the action of L_h is to move points either along hyperbolas or along the axes $\mathbf{v}_1$ and $\mathbf{v}_2$.

The case of points on the axes is particularly simple. Then we have either $\epsilon_2 = 0$ or $\epsilon_1 = 0$ and hence either

$$L_h^n\epsilon = \lambda^n\epsilon \tag{34a}$$

or

$$L_h^n\epsilon = \lambda^{-n}\epsilon \tag{34b}$$

respectively. Here we have used our convention $\lambda_1 = \lambda > 1$ and $\lambda_2 = 1/\lambda$. We see that the action of L_h is to move points on $\mathbf{v}_1$ along $\mathbf{v}_1$ away from the origin and points on $\mathbf{v}_2$ along $\mathbf{v}_2$ into the origin. Figure 6 illustrates the action of L_h both for this case and for points off the axes.

This completes our exploration of L_h and its action. Returning to (30), we see that 'in the small,' when higher powers in ϵ are neglected, the effect of M itself is to move points near $\mathbf{h}$ along on hyperbolas or their asymptotes, the axes. Another way of describing this situation is to say that neglecting higher-order terms in ϵ, there is a set of curves near $\mathbf{h}$, namely, hyperbolas and their axes, which are each invariant under M.

It is a remarkable result that the full map M, with no powers of ϵ neglected, also possesses invariant curves in the neighborhood of a hyperbolic fixed point. Naturally enough, near the fixed point these curves look like the hyperbolas and their asymptotes that we obtained by examining L_h. The existence of invariant curves which pass through the fixed point which are the analogs of the $\mathbf{v}_1$ and $\mathbf{v}_2$ axes was proved by *Hadamard* [1901]. Formal series expressions for these curves and the invariant curves analogous to hyperbolas were obtained by *Birkhoff* [1920]. Finally, *Moser* [1956] established that these series actually converge near the hyperbolic fixed point and thus proved the existence in general of invariant curves near hyperbolic fixed points.

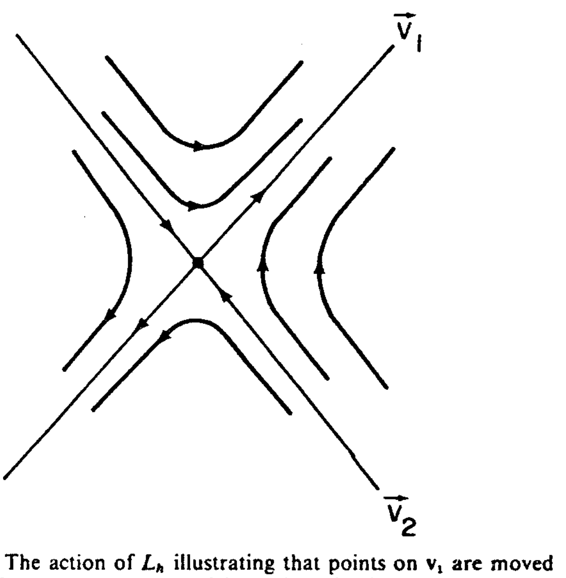

Fig. 6. The action of L_h illustrating that points on $\mathbf{v}_1$ are moved outward, those on $\mathbf{v}_2$ are moved inward, and others are moved on hyperbolas.

Let us now focus our attention on two particular invariant curves of M, namely, those which pass through the fixed point $\mathbf{h}$ itself. As was mentioned earlier, these two curves are the analogs of the $\mathbf{v}_1$ and $\mathbf{v}_2$ axes that we found when we neglected the higher-order terms in ϵ. By (30) they are tangent to $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, as they pass through $\mathbf{h}$. We shall call the curve which is analogous to the $\mathbf{v}_1$ axis the unstable manifold and denote it by the symbol W_u. Similarly, we will denote by W_s the curve which is analogous to the $\mathbf{v}_2$ axis and call it the stable manifold. They are given these names because in analogy to (34) it can be shown that they have the properties

$$W_s = \forall \ \mathbf{p} \text{ near } \mathbf{h} \qquad (35a)$$

such that

$$\lim_{n\to\infty} M^n\mathbf{p} = \mathbf{h}$$

and

$$W_u = \forall \ \mathbf{p} \text{ near } \mathbf{h} \qquad (35b)$$

such that

$$\lim_{n\to\infty} M^{-n}\mathbf{p} = \mathbf{h}$$

That is, W_s consists of all points $\mathbf{p}$ which are ultimately moved into $\mathbf{h}$ under repeated action of M. Hence we get the name stable. Note that by (34b), points on the $\mathbf{v}_2$ axis have this property if we consider only the action of L_h. By contrast, W_u consists of all points which go into $\mathbf{h}$ under repeated action of M^{-1}, and hence they are moved away from $\mathbf{h}$ under repeated action of M. Evidently, from (34a), this is analogous to the behavior of points on the $\mathbf{v}_1$ axis under the repeated action of L_h^{-1} and L_h.

To get an idea of how our discussion works out in a specific case, let us consider what might be regarded as the simplest nonlinear area-preserving map. It is a map consisting of just linear and quadratic terms. If we select x and y as Cartesian coordinates in the plane, our simple example is given by the rule

$$\begin{aligned} x' &= \lambda[x + (x - y)^2] \\ y' &= \lambda^{-1}[y + (x - y)^2] \end{aligned} \qquad (36)$$

where, for the moment, λ is an adjustable parameter. We shall denote this mapping by the symbol M_c in honor of Cremona, who was an early student of polynomial maps.

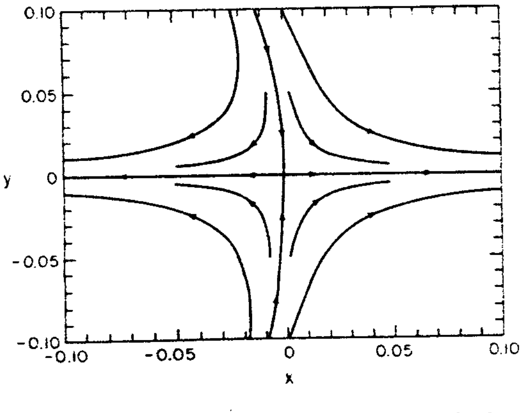

Fig. 7. Invariant curves for the Cremona map M_c in the case $\lambda = 3$. The curves through the origin are the invariant manifolds W_u and W_s.

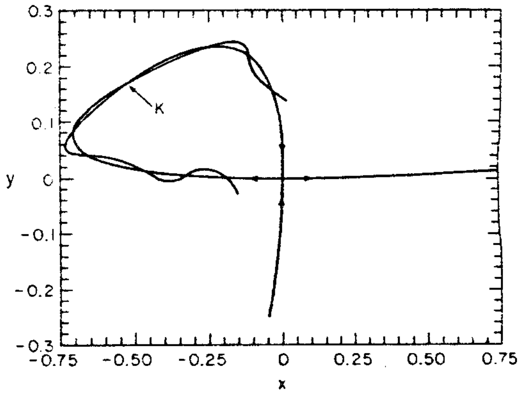

Fig. 8. The extension of W_u and W_s of Figure 7 resulting in a homoclinic point K for M_c.

A short calculation shows that the linear part L_{xy} of M_c at the point x, y is given by the matrix

$$L_{xy} = \begin{pmatrix} \lambda[1 + 2(x - y)] & -2\lambda(x - y) \\ 2(x - y)/\lambda & [1 - 2(x - y)]/\lambda \end{pmatrix} \qquad (37)$$

It is easily verified that L_{xy} has determinant +1, so our example is indeed area preserving as advertised.

Next we observe that M_c has the origin as a fixed point. That is, the point $x = y = 0$ is sent into itself. Moreover, at the origin, L_{xy} has the form

$$L_{00} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \qquad (38)$$

which shows that λ and λ^{-1} are the eigenvalues associated with the fixed point. Therefore if we take for λ some number greater than 1, the origin is a hyperbolic fixed point. Finally, in this case the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ lie along the x and y axes, respectively. We therefore expect that the unstable manifold W_u for M_c will be tangent to the x axis at the origin and the stable manifold W_s will be tangent to the y axis.

Figure 7 shows various invariant curves, including W_u and W_s, in the neighborhood of the origin for M_c in the case $\lambda = 3$. These curves were obtained by analytical and numerical means which we will publish in detail elsewhere [*Finn*, 1974]. We see that they indeed have the expected hyperbolic structure near the origin.

The stage is now set for us to make a fundamental observation first about M_c and then the general case. Suppose we apply successive powers of M_c and M_c^{-1} to the invariant curves that we have obtained near the origin. This operation will have the effect of extending the curves away from the origin. Moreover, the extended curves will still be invariant, for by construction if $\mathbf{p}$ is a given point on an original curve or its extension, $M_c\mathbf{p}$ and $M_c^{-1}\mathbf{p}$ will also be on the curve or its extension.

In particular, let us extend in this fashion the unstable and stable manifolds, W_u and W_s. Figure 8 shows the result obtained numerically for $\lambda = 3$. We see that the piece of W_u corresponding to positive x and the piece of W_s corresponding to negative y extend off to infinity. However, the other two pieces of W_u and W_s approach each other and eventually meet. Indeed, they cross each other with a nonzero angle.

The possibility that the W_u and W_s emanating from a hyperbolic fixed point $\mathbf{h}$ of an area-preserving map might intersect with a nonzero angle (rather than joining smoothly) was first

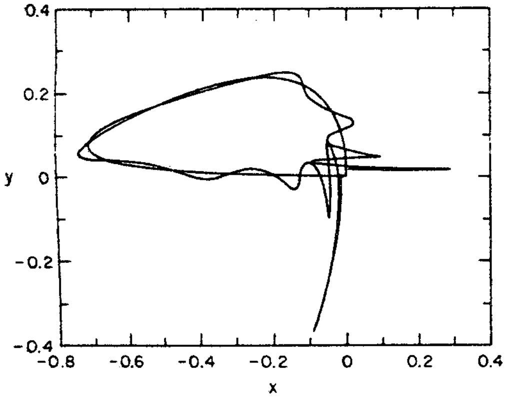

Fig. 9. Successive homoclinic intersections and oscillations for M_c. The other halves of W_u and W_s, those pieces that go off to infinity, are not shown.

envisioned by *Poincaré* [1892]. He called such a point of intersection a homoclinic point and showed that if a homoclinic point existed, W_u and W_s must have an extremely complicated oscillatory structure.

Poincaré reasoned roughly as follows: Let us denote a homoclinic point of an area-preserving map by the symbol K. Now apply powers of the map and its inverse to K. In the case of our example, the Cremona map M_c, this means that we should compute the points M_cK, $M_c^{-1}K$, M_c^2K, $M_c^{-2}K$, etc. Then, by the construction of W_u and W_s, these points must also lie on both W_u and W_s, since K belonged to both W_u and W_s! Therefore the curves W_u and W_s must intersect each other over and over again. In fact, they must intersect infinitely often, and if the map is differentiable as we have been assuming, all the angles of intersection must be nonzero. Thus the existence of a single homoclinic point implies the existence of an infinite set of homoclinic points. And in order to achieve this infinite set, W_u and W_s must oscillate around each other infinitely often.

Figure 9 shows this intersection and oscillation for the map M_c, again with $\lambda = 3$. Two properties are immediately apparent. First, the spacing between successive homoclinic points becomes finer and finer as one approaches the hyperbolic fixed point either along W_u or W_s. This is to be expected because the behavior of the map is governed by its linear part near the fixed point, and then (34) comes into play. Second, the amplitude of oscillation increases as one approaches the fixed point. This occurs because, since the map M_c is area preserving, the areas 'under' successive oscillations must all be the same. In order to preserve area in the face of decreased spacing, the amplitude must increase.

The net effect of these two properties is that near the hyperbolic fixed point the oscillations of W_u about W_s must intersect the oscillations of W_s about W_u to produce even more homoclinic points. The result is that the hyperbolic fixed point is actually the corner of an ever denser 'cloud' of homoclinic points. This property is illustrated for M_c in Figure 10.

A moment's reflection on the reader's part now will show that because of the generality of our arguments, the existence of this cloud of homoclinic points is not peculiar to just the mapping M_c. It will in fact occur for any area-preserving map which possesses a homoclinic point.

5. Homoclinic Points and Insolubility

What does all this have to do with the Størmer problem? In the next section we will show numerically that the mapping M for the Størmer problem has a homoclinic point. In this section we will show that the existence of a homoclinic point means that the Størmer problem is 'insoluble.'

Consider any dynamical system having n degrees of freedom, canonical coordinates $q_1, \cdots, q_n$ and $p_1, \cdots, p_n$ (collectively denoted by q and p), and a time development generated by a time independent Hamiltonian $H(p,q)$. Such a system will always possess $2n$ independent constants of motion $C_1(p, q, t)$, $\cdots$, $C_{2n}(p, q, t)$ which in general depend on the canonical variables p, q and the time t. By a constant of motion we mean a function which satisfies

$$(d/dt)C_i(p, q, t) = 0 \qquad (39)$$

along every trajectory. Indeed, given any point (q, p, t) in state space, we can always trace back the unique trajectory through this point to a fixed reference time t° and then record the $2n$ numbers p°, q°. These quantities, which we write as $p^\circ(p, q, t)$, $q^\circ(p, q, t)$, are obviously constant along a trajectory by construction. Hence they, or any $2n$ functionally independent functions of them, provide $2n$ constants of motion.

By the very generality of the argument that we have just given, the existence of $2n$ constants of motion places very little restriction on the trajectories generated by H. Indeed it seems quite possible and is in fact assumed in statistical mechanics or ergodic theory that for some Hamiltonians there may be some trajectories which, when traced forward and backward in time, wander arbitrarily near any point in phase space.

One way to preclude such a complicated behavior is to demonstrate the existence of what we shall call integrals of motion. For a given Hamiltonian $H(p, q)$ we define an integral of motion $I(p, q)$ to be a single-valued analytic time independent function on phase space which, like a constant of motion, also satisfies $dI/dt = 0$ along every trajectory generated by H. By our hypotheses about the nature of I, equations of the form $I(p, q) =$ const describe a set of disjoint hypersurfaces in phase space. And if H does have I as an integral of motion, each trajectory is confined to one of these hypersurfaces. Thus complicated 'wanderings' which would take a trajectory from one hypersurface to another are ruled out. For this reason, the integrals that we have defined are sometimes called isolating integrals [*Contopoulos*, 1963].

The motion on a given hypersurface may still be very complicated. However, suppose that there exist further independent integrals I_2, I_3, etc. in addition to $I = I_1$. Then trajectories must lie on intersections of families of hypersurfaces, and the

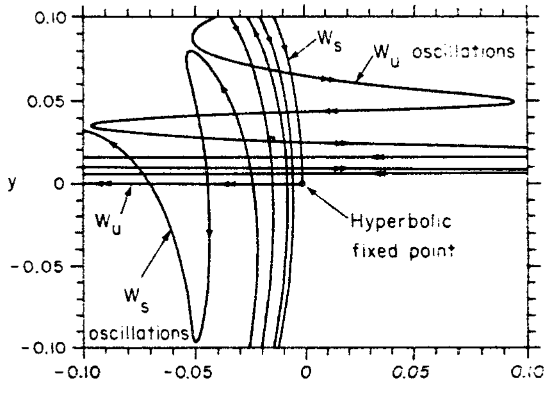

Fig. 10. A continuation of Figure 9 near the origin showing the formation of a grid of intersecting lines. The spacing of the grid becomes finer and finer as it approaches the hyperbolic fixed point. Each grid intersection is a homoclinic point. The result of all these intersections is an ever denser cloud of homoclinic points which has the hyperbolic fixed point as a limit point.

motion is consequently further restricted. In general, the more integrals a system has, the simpler it is to describe its motion. And the more integrals we know for a given system, the more we know about its motion.

At this point the ingenious reader may object that every Hamiltonian system with $2n$ degrees of freedom will automatically possess $2n - 1$ integrals. One merely takes the $2n$ constants of motion $C_i(p, q, t)$ and functionally eliminates the time from among them. The result of this elimination will be $2n - 1$ functions of just the variables p, q, and these functions will remain constant along trajectories. Hence why not call them integrals?

Some such argument is made more or less explicitly in almost every advanced mechanics text. However, all that this argument proves is the existence of what we might call local integrals of motion. There is no guarantee that these local integrals extend to global single-valued analytic functions. In fact, if the motion happens to exhibit various features of ergodicity, such an extension must be impossible.

Let us return to the Størmer problem. It obviously possesses p_ϕ and H as integrals. The use of p_ϕ enabled us to reduce the three-dimensional problem to one with two dimensions, and the use of H enabled us to group together trajectories according to the value of W_0^2. We will now show that if the mapping M has a homoclinic point, then there are no further global integrals of motion. This means that there is no further single-valued analytic function $I(p, q)$ on phase space which satisfies the Liouville equation

$$\dot{I} = [I, H] = 0 \tag{40}$$

where the brackets denote the Poisson bracket operation.

In practical calculations we begin with differential equations whose terms are analytic and thus are naturally led to work with analytic functions, or uniformly convergent (and hence again analytic) sums of analytic functions. The nonexistence of an analytic integral means that the trapped Størmer problem cannot be handled in this manner and therefore is entitled to be called 'insoluble.' This does not mean that one cannot compute Størmer trajectories. Indeed, numerical integration is always possible unless the trajectory goes to the origin. Nor does it mean that further progress cannot be made on the Størmer problem. But it does mean that further progress will be quite different from what one normally thinks about when one speaks of 'solving' a classical mechanics problem. It may in fact even be the case that Størmer trajectories are ergodic in certain regions of phase space.

We are ready to present an argument as to why the existence of a homoclinic point rules out the existence of any global analytic integral beyond p_ϕ and H itself. Suppose I is a further integral. Let $\mathbf{p}$ be a point in the $\rho, \dot{\rho}$ plane. We have seen that such a point, for a fixed value of W_0^2, defines a unique trajectory in phase space. Let $I(\mathbf{p})$ denote the value of the integral along this trajectory. Now apply powers of M to $\mathbf{p}$. Since M is generated by following trajectories and I is constant along a given trajectory, we must have

$$I(M^n\mathbf{p}) = I(\mathbf{p}) \qquad n = 0, \pm 1, \pm 2, \cdots \tag{41}$$

We see that I is invariant under the action of M.

Suppose $\mathbf{p}$ is some point on the stable manifold W_s belonging to a hyperbolic fixed point $\mathbf{h}$. Then in view of (35) and (41) we must have

$$I(\mathbf{p}) = I(\mathbf{h}) \qquad \mathbf{p} \in W_s \tag{42}$$

Here we have used the fact that I is assumed analytic and hence must be continuous. Suppose $\mathbf{p}$ is some point on the unstable manifold W_u. Then we again get

$$I(\mathbf{p}) = I(\mathbf{h}) \qquad \mathbf{p} \in W_u \tag{43}$$

by the same argument. It follows that I has one and the same value, namely, $I(\mathbf{h})$, everywhere on the curves W_u and W_s. In particular, the directional derivative of I must vanish along both the curves W_u and W_s.

Next suppose that W_u and W_s intersect at a homoclinic point K located at $\mathbf{k}$. Then we must have

$$I(\mathbf{k}) = I(\mathbf{h}) \tag{44}$$

since $\mathbf{k}$ belongs to both W_u and W_s. But even more can be said: We know that the directional derivative of I is zero along both W_u and W_s. However, at the point $\mathbf{k}$ the tangents to W_u and W_s are linearly independent vectors, since by definition W_u and W_s intersect there at an angle. Because the directional derivative of I now is zero along two linearly independent directions, we conclude that the gradient of I must vanish at any homoclinic point:

$$\nabla I(\mathbf{k}) = 0 \tag{45}$$

where $\mathbf{k}$ is any homoclinic point. The same argument applies to the hyperbolic point $\mathbf{h}$, where W_s and W_u also meet at an angle.

We are almost done. We saw earlier that the fixed point $\mathbf{h}$ was the corner of an ever denser cloud of homoclinic points arranged in such a way that $\mathbf{h}$ is an accumulation point of the cloud along several different paths. We now also know that $\nabla I = 0$ at each of these points. Finally, we have assumed that I and hence ∇I are analytic functions. It follows from the well-known uniqueness theorem for analytic functions [*Titchmarsh,* 1939] by a simple extension to two variables that ∇I must vanish identically at every point in the physical region. Therefore I must have the same value, namely, $I(\mathbf{h})$, everywhere. That is, the integral is just a constant function everywhere and is therefore useless! Thus we conclude that the existence of a homoclinic point precludes the existence of any global analytic integral except for useless constant functions [*Moser,* 1973].

We close this section with a remark about soluble classical mechanics problems such as appear in textbooks. How do they fit into our discussion? What is special about any mapping M that they might generate? Briefly, the answer to these questions is that soluble problems never produce a homoclinic point. Their fixed points are either elliptic or parabolic, so that there are no stable and unstable manifolds, or there are a few hyperbolic fixed points whose stable and unstable manifolds either never meet or, if they do meet, join smoothly without intersecting. Thus no homoclinic points are ever formed.

6. Numerical Evidence for Homoclinic Points

Figure 11 shows a periodic orbit for the Størmer problem in the case $W_0^2 = 0.01$. It corresponds to a fixed point $\mathbf{h}$ of M given by $\rho_h = 1.11494632$, $\dot{\rho}_h = 0$. Numerical calculation shows that this point is hyperbolic with an eigenvalue given by $\lambda = 2.49$ and that the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ of L_h are $\mathbf{v}_1 = (1, 53)$ and $\mathbf{v}_2 = (1, -53)$. The location of $\mathbf{h}$ and the arrangement of $\mathbf{v}_1$, $\mathbf{v}_2$, W_u, and W_s in this case are shown schematically in Figure 12. Here, unlike the case in Figure 8, no pieces of W_u or W_s go off to infinity. Instead, the piece of W_u extending from $\mathbf{h}$ into the upper half plane meets on the ρ-axis with that piece of W_s which approaches $\mathbf{h}$ from the lower half plane. The other piece of W_u also meets the other piece of W_s on the ρ axis with a slightly larger value of ρ. In anticipation of future results we

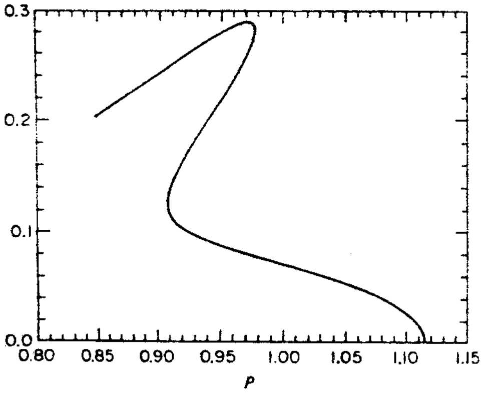

Fig. 11. A periodic Størmer orbit corresponding to a hyperbolic fixed point of M. The orbit is symmetric about the ρ axis and retraces itself.

have indicated that both intersections of W_u and W_s occur with a nonzero angle. That is, we have depicted two homoclinic points.

Figure 13 shows the results of computing $M^n(\mathbf{h} \pm \delta \mathbf{v}_{1,2})$ foı $n = -10, -9, \cdots, +10$ and $\delta = 1. \times 10^{-6}, 1.2 \times 10^{-6}, \cdots 2.6 \times 10^{-6}$. Since δ is very small, the points that we obtain in this manner must lie very near W_u and W_s. It is evident from the figure that the respective branches of W_u and W_s either meet or intersect near the points (0.933, 0) and (0.945, 0), but we cannot be immediately sure what happens; i.e., the angles of intersection may be zero. They are, at any rate, rather small.

By a technique to be published elsewhere [*Finn*, 1974] we have been able to obtain the first few terms of a power series $I(p_1 p_2 q_1 q_2)$ which satisfies term by term the equation

$$[I, H] = 0 \tag{46}$$

Here H is the Hamiltonian given by (18). In addition we have been able to give an algorithm for computing arbitrarily many terms in the series. Our algorithm can be applied to any Hamiltonian of the form (18) providing the leading piece H_2 has the structure given by (19). In actual practice the algebraic expressions involved in the calculation soon become too lengthy for human manipulation. For that reason we have programmed our algorithm into a digital computer and let it do the algebra. Even the answer itself is rather complicated. For the Størmer problem we find, by using the same notation as in (18), that the first few terms are given by

$$I_2 = (p_2^2 + q_2^2)/2 \tag{47a}$$

$$I_3 = -3p_1^2 q_2 - 2q_2^3 \tag{47b}$$

$$\begin{aligned} I_4 = {} & (\text{[illegible]})p_1^2 p_2^2 + (\tfrac{117}{8})p_1^2 q_2^2 \\ & + (\tfrac{3}{4})p_2^2 q_1^2 + (\tfrac{15}{4})p_2^2 q_2^2 - (\tfrac{3}{4})q_1^2 q_2^2 \\ & - (\tfrac{3}{2})p_1 p_2 q_1 q_2 + (\tfrac{9}{2})p_1^4 + (\tfrac{15}{8})p_2^4 + (\tfrac{55}{8})q_2^4 \end{aligned} \tag{47c}$$

In general it is best not to write down the answer on paper. Rather, one should transfer it directly from one computer program to another, since because of length, only computers can make use of the answer anyhow!

Three things should be said about the power series I. First, examination of the series shows that it contains all the terms which one would obtain by expanding the usual adiabatic magnetic moment invariant in a power series. It also contains additional terms which correspond to as yet unknown (and perhaps forever unknown due to algebraic complexity) higher-order corrections to the magnetic moment invariant. Second, the series cannot be convergent if M has a homoclinic point. For if the series did converge, it would provide an additional integral of motion in contradiction to the results of the last section. Third, if we truncate the series I at an appropriate level, we obtain a quantity I^T, which is nearly constant along trajectories since I formally satisfies (46). This means that we may use the truncated integral I^T as a sort of 'magnifying glass' to examine the behavior of trajectories in great detail. The idea is that if, for example, we look at a set of points $M^n\mathbf{p}$, we may not be able to detect small homoclinic oscillations because we need a rather coarse scale in order to just plot the points. However, if we instead study $I^T(M^n\mathbf{p})$, then most of the variation in I^T may be due to homoclinic oscillations.

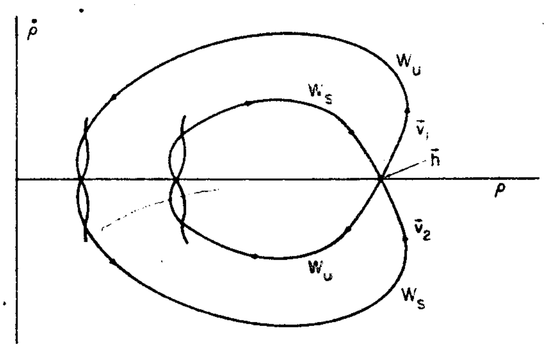

Fig. 12. A schematic drawing showing the location of a hyperbolic fixed point **h** and the arrangement of $\mathbf{v}_1$, $\mathbf{v}_2$, W_u, and W_s for the Størmer map M.

To display how constant I^T is in practice, Figure 14 presents a set of points **p** satisfying

$$I^T(\mathbf{p}) = I^T(\mathbf{h}) \tag{48}$$

when I^T is the polynomial obtained by truncating the series I beyond terms of sixth order. We have replotted the points of Figure 13 to show that the outer branches of W_u and W_s lie remarkably close to the curve (48).

We next select a point **g** given by $\mathbf{g} = (1.115, 0)$. It lies just slightly to the right of **h**. Consequently, if we compute the points $M^n\mathbf{g}$ for $n = 0, \pm 1, \pm 2, \cdots$, we should get points very near the outer branches of W_u and W_s. We have carried out this calculation for n between −80 and +80. Figure 15 shows the results: the points are indistinguishable in behavior from those in the outer portion of Figure 13 and appear to lie on a smooth curve.

Suppose now that the outer branches of W_u and W_s intersect at a finite angle near (0.933, 0) and then go into oscillation about each other. This oscillation, if it exists, should be reflected in the behavior of the points $M^n\mathbf{g}$, since they lie near

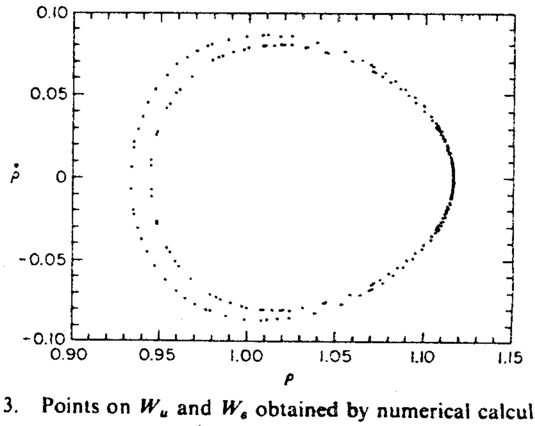

Fig. 13. Points on W_u and W_s obtained by numerical calculation.

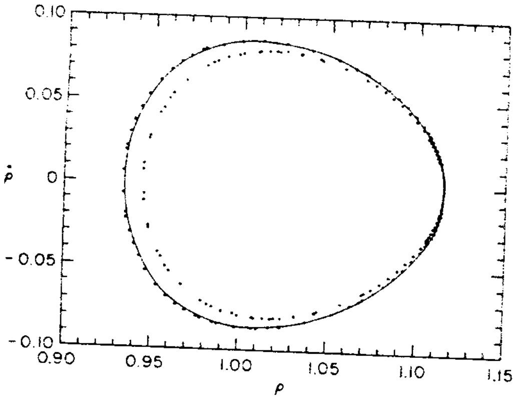

Fig. 14. The quantity I^T is nearly constant along trajectories. The line consists of points satisfying $I^T(\mathbf{p}) = I^T(\mathbf{h})$. The remaining points are those of Figure 13.

$W_u$and W_s. In fact, an extension of the theory in section 4 shows that near W_u and W_s there are further invariant curves and that these curves will oscillate about each other just as W_u and W_s do.

Figure 16 shows $I^T(M^n\mathbf{g})$ plotted as a function of the ρ component of $M^n\mathbf{g}$. It is now evident, because we can use a finer scale, that there is considerable scatter in the points. Also the scatter is largest for ρ near ρ_h, the location of the hyperbolic fixed point. This is exactly what is to be expected from homoclinic oscillations. The value of I^T should jump about erratically, and the jumps should be largest near the hyperbolic point where, as we saw in section 4, the homoclinic oscillations should be largest. This effect is further illustrated in Figure 17, where we have plotted $I^T(M^n\mathbf{g})$ versus the $\dot{\rho}$ component of $M^n\mathbf{g}$ for all those points lying in the right half plane ($\rho \geq 1$) of Figure 15. This latter mode of presentation is particularly useful because it spreads out points near the hyperbolic fixed point. Again, it is clear that there is erratic behavior and that this erratic behavior is greatest near the hyperbolic fixed point. We therefore conclude that the outer portions of W_u and W_s intersect near (0.933, 0) in a nonzero angle, although this angle is too small to be seen directly. Finally, we have made a similar study of the inner portions of W_u and W_s and find that they also intersect in a nonzero angle.

At this point the reader may have several questions. First, how do we know that our results are not due to inaccuracies in numerical integration? To rule out this possibility, we have checked for truncation errors by halving the step size, and we have checked for round-off errors by first integrating a trajectory forward in time and then backward to see if we can

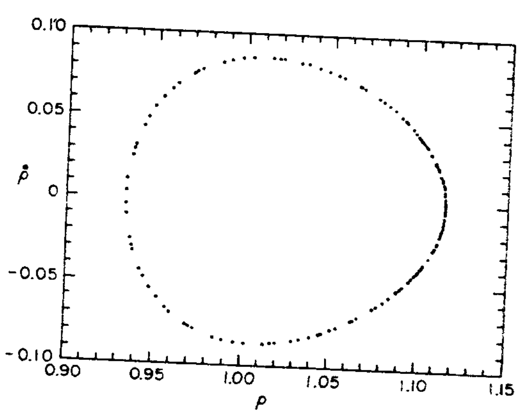

Fig. 15. Points very near the outer branches of W_u and W_s. There is no visible sign of homoclinic oscillation.

regain the initial conditions. We find that the points u
making the figures in this paper are accurate to at leas
significant figures. By contrast, the homoclinic oscill
appearing in Figures 16 and 17 occur in the third signi
figure.

Second, why in preparing Figures 15 through 17 did w
points $M^n\mathbf{g}$ just outside of W_u and W_s rather than the
points in Figure 13, which are much closer to W_u and W
in fact first did exactly that by computing the points M^n
$10^{-6} \times \mathbf{v}_1$) for large n and found to our surprise that for la
we began to get points not only on the outside portions c
and W_s but also points on the inside portions! This behavi
first sight seems confusing, but its explanation is that for
n the homoclinic oscillations build up to such a large ar
tude that they cause 'transitions' between the inner and c
portions of W_u and W_s. In fact, even for the case of Figu
we found that we began to get points near the inner portio
W_u and W_s if we used values of n greater than 84 in abso
value. The occurrence of these transitions, which inciden
take place in both directions, is another proof of the exist
of homoclinic oscillations.

The fixed point corresponding to the orbit in Figure 11 is
hyperbolic fixed point of M (for $W_0^2 = 0.01$) which is closes
the boundary of the physical region. It is but the first in a se
of a hundred or more (but finite in number!) hyperbolic fi
points which lie on the ρ axis ($\dot{\rho} = 0$) and extend toward ρ_0
launch value of ρ for the orbit to the origin. (Numer
integration gives $\rho_0 = 1.03607169$, $\dot{\rho}_0 = -5.67675 \times 10^{-3}$
$W_0^2 = 0.01$.) We find that as ρ tends toward ρ_0, each fi
point is more hyperbolic (i.e., λ is larger) than the last.
have also found that each hyperbolic point examined
manifolds W_u and W_s which intersect in two homocli
points. Finally, the homoclinic angles, as judged by the a
between oscillations, become ever larger the closer the pare
hyperbolic point is to ρ_0. Thus there is conclusive numeri
evidence that M for the Størmer problem has not only
homoclinic point with its consequences but also has seve
distinct families of pairs of homoclinic points. The first fami
is near the boundary of the physical region. Subsequent fam
lies lie ever closer to ρ_0. As stated earlier, this series of famili
eventually terminates at a finite distance from ρ_0.

7. DISCUSSION AND CONCLUSION

We learned in section 5 that the existence of a single homo
clinic point implied insolubility. We now have also learne
that the Størmer problem is almost unbelievably rich in homo
clinic points. It follows that the Størmer problem is insolubl
in the sense that there are no further global analytic integral
of motion.

We have also found that the complete adiabatic magneti
moment series is divergent. Therefore the magnetic moment
series cannot be used to infer the long-time behavior of orbits.
This circumstance places us in an uncomfortable position: Just
what can be said about long-time behavior?

First, by assuming albedo neutron decay to be a source and
atmospheric scattering to be a loss mechanism, one can infer
experimentally that there are radiation belt protons whose
lifetimes must be at least of the order of 50 yr. These same
protons have a cyclotron period of about 0.02 s and a bounce
period of about 0.2 s. Consequently, we need to deal with
orbits consisting of approximately 10^{11} gyrations and 10^{10}
bounces! A similar calculation for radiation belt electrons
shows that they make even more gyrations.

Second, if we follow these orbits numerically, in the dipole approximation to the earth's field, we find that the quantity $E_\perp/B$ (the first term in the magnetic moment series and the expression usually assumed to be constant in practical calculations) typically changes by 1%–50% between equatorial crossings. The truncated magnetic moment series I^T, with a truncation chosen to minimize variations, typically varies by 0.01%. See, for example, Figures 16 and 17. Taking the best case of a 0.01% change per crossing and hoping that successive changes accumulate randomly (rather than additively as a pessimist would assume), we conclude that the magnetic moment could change by 100% after 10^8 crossings. Thus even in the best case under the most optimistic assumptions we are at least 2 orders of magnitude away from the confinement times that apparently are required. It seems that successive changes in $E_\perp/B$ or I^T accumulate neither additively nor randomly but instead must be highly self cancelling. We stress that this conclusion is based on our present understanding of radiation belt measurements. It currently has no rigorous mathematical or numerical justification for orbits of physical interest. Later in our discussion we will see that from the work of Braun there is a proof of long-term confinement for protons having very small energies.

Second, if we try to follow orbits directly for long times by numerical integration just to see what happens, we are even less successful. Typically, with high-speed computers one can integrate from 1 to 10 times faster than protons of interest move in real time. Thus we need from 5 to 50 yr of computer time to simulate reality! But even if one could afford the cost this would entail, the accuracy of the numerical solution so obtained would be destroyed early on by truncation and round-off errors.

We conclude that there is currently no rigorous mathematical or numerical justification for the use of adiabatic invariants to predict long-time behavior. (Note that use of the longitudinal and flux invariants for long times presupposes the convergence of the magnetic moment series. Moreover, even if the magnetic moment series converged for some particular problem, the convergence of the longitudinal and flux invariant series would still be suspect.) Its empirical success for the Van Allen radiation and for laboratory mirror machines where a similar situation holds [*Gibson et al.*, 1960] is truly remarkable.

We might inquire at this point whether the full physical problem might be soluble even though the Størmer idealization is not. This seems very unlikely for the following reason. Suppose we denote the coefficients of the earth's magnetic field in a vector spherical harmonic expansion by the symbols β_1, β_2, etc. Then if the *full* problem is soluble, there will be an integral of motion of the form $I(\phi\rho z; p_\phi p_\rho p_z;\ \beta)$. If this integral can be found, we expect that it will be an analytic function of the parameters β_i, and hence I can be evaluated when all $\beta_i =$ 0. But then we have found an additional integral for the Størmer problem which we know is impossible. So either the full problem is also insoluble, or its integrals are not analytic in the parameters β_i. The first possibility rules out the existence of an integral, and the second probably rules out the discovery of an integral.

We should also discuss the fact that our definition of insolubility is perhaps not the first one that would come to the reader's mind. The question of what types of problems should be called soluble or integrable has had a long history [*Birkhoff*, 1927; *Whittaker*, 1937; *Wintner*, 1947; *Arnold and Avez*, 1968]. Originally, a problem was viewed as soluble if its solution could be written down, perhaps only implicitly, in terms of definite integrals (quadratures). The definition was later extended when Liouville showed that a problem could be reduced to quadratures if one could find n (for a system of n degrees of freedom) integrals in involution (i.e., satisfying $[I_i, I_j] = 0$). Thus the determination of integrals of motion entered into the notion of solubility. See, for example *Hagihara* [1970].

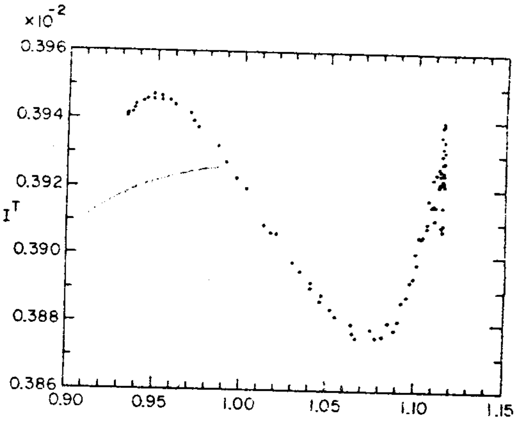

Fig. 16. The use of I^T as a magnifying glass. The points of Figure 15 are here replotted to show $I^T(\mathbf{p})$ as a function of the ρ component of $\mathbf{p}$. The effect of homoclinic oscillations is now obvious near ρ = 1.115.

This point of view also brings to the fore the Hamilton-Jacobi equation

$$H(\partial W/\partial q, q) = E \tag{49}$$

or its variants. Its complete solution for the transformation function $W(\alpha_1, \cdots, \alpha_n, q_1, \cdots, q_n)$ would lead to new coordinates $Q_i(p, q)$ which would be ignorable, and hence the canonically conjugate $P_i(p, q)$ would be integrals of motion in involution.

At this point it is essential to make again the distinction between local and global integrals. As was mentioned earlier, local integrals always exist. (Strictly speaking, one must not be at an equilibrium point where all derivatives of H with respect to the p's and q's are zero, and hence all p's and q's are constant. For a proof, see *Abraham* [1967].) This means, among other things, that locally the Hamilton-Jacobi equation always has a complete solution depending on n integration constants $\alpha_1, \cdots, \alpha_n$. Further, the solution satisfies det $(\partial^2 W/\partial\alpha\partial q) \neq 0$, so that the desired transformation to new variable can actually be accomplished.

However, we are interested in global integrals, or at least in integrals that are sufficiently global that they exist in all re-

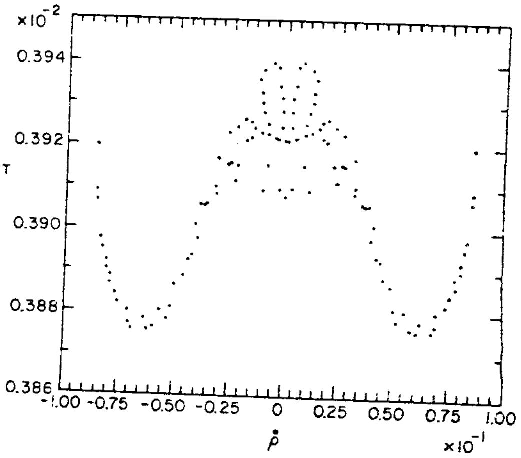

Fig. 17. The points of Figure 15 satisfying $\rho \geq 1$ are replotted here to display $I^T(\mathbf{p})$ versus the $\dot{\rho}$ component of $\mathbf{p}$. The result of homoclinic oscillations near $\mathbf{h}$ ($\dot{\rho}$ = 0) is now even more apparent

gions of phase space visited by a class of trajectories of physical importance. We have seen that no such global integral beyond p_ϕ and H exists for the Stǿrmer problem because of homoclinic points. It follows that there is also no complete solution W to the Hamilton-Jacobi equation which leads to a global analytic transformation to new coordinates, such as action-angle variables, for such a solution would provide a global analytic integral.

There is also the possibility of having integrals which are not global but whose domain is more than local. Such integrals might be called 'regional.' Three such situations occur for the Stǿrmer problem in very different ways. The first involves untrapped orbits (J. Moser, unpublished manuscript, 1963). So far we have only discussed trapped orbits which are constrained by energy conservation and initial conditions to lie in the valley to the left of the pass in Figure 2. Now consider orbits which are launched outside the valley, again with $W_0^2 < \frac{1}{16}$. By energy conservation these orbits can never enter the valley. Further, we find that

$$d^2/dt^2\,(\mathbf{r}^2) = [2(\dot{\rho}^2 + \dot{z}^2) - 2\mathbf{r}\cdot\nabla V] > 0 \tag{50}$$

since by direct computation or examination of Figure 2

$$(-\mathbf{r}\cdot\nabla V) > 0 \tag{51}$$

in the region of interest. It follows that all these orbits eventually approach infinity for both large positive and negative time.

As the orbit approaches $r = \infty$ in the ρ, z plane for large (say positive) time, it becomes nearly a straight line, and we can write

$$\mathbf{r}(t) = \mathbf{v}t + \mathbf{u} + O(1/t) \tag{52}$$

as $t \to +\infty$, where, since the potential V vanishes at infinity, $\mathbf{v}$ satisfies the relation

$$\mathbf{v}^2 = W_0^2 \tag{53}$$

We also require that $\mathbf{u}$ be orthogonal to $\mathbf{v}$ to make its definition unique:

$$\mathbf{u}\cdot\mathbf{v} = 0 \tag{54}$$

The two two-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$ assign to each trajectory four unique numbers. It is also clear that these numbers depend analytically on the coordinates q and momenta p at any point of a phase space trajectory and hence are integrals of motion. However, these integrals are not all independent. Because we are dealing with trajectories of a fixed energy, the magnitude of $\mathbf{v}$ is redundant by (53). Further, by (54) we are only interested in $u_\perp$, the component of $\mathbf{u}$ perpendicular to $\mathbf{v}$. (In the language of scattering theory, $u_\perp$ can be viewed as the 'impact parameter' with respect to the origin in the absence of the potential $V(\rho, z)$.) We conclude that (52) through (54) assign to each trajectory two independent quantities, the direction of $\mathbf{v}$ and $u_\perp$, and that these quantities are integrals of motion. These two quantities and H provide three analytic integrals of motion for the two-dimensional orbits in the ρ, z plane. This is just the maximum number to be expected for two degrees of freedom. A simple extension of the argument to the full orbits in three dimensions shows that one can find five integrals in this case. Thus in what we might call the 'untrapped' or 'scattering' region the Stǿrmer problem is completely integrable.

Suppose we try to continue analytically the integrals found in the scattering region into the trapped region. This continuation must fail, for we know that there are no further global integrals beyond p_ϕ and H in the trapped region. Thus the Stǿrmer scattering problem is integrable and hence soluble, while the trapped Stǿrmer problem is insoluble. It is interesting to speculate about the status of the problem for the case $W_0^2 > \frac{1}{16}$ in which case at least some orbits in the valley can also escape to infinity!

A second type of regional integrability occurs for the case of trajectories sufficiently near a periodic trajectory corresponding to a hyperbolic fixed point of M. We have already seen that the map M has invariant curves in the surface of section near a hyperbolic fixed point $\mathbf{h}$. These curves look like distorted hyperbolas and their asymptotes near $\mathbf{h}$, and indeed *Moser* [1956] has shown that sufficiently near $\mathbf{h}$ there exists an analytic invertible transformation to new variables ξ, η such that the invariant curves take the form

$$\xi\eta = \text{const} \tag{55}$$

Consequently, any function of the product $\xi\eta$, say, $f(\xi\eta)$, becomes an invariant function when it is written in terms of the original variables. It is easy to see that this invariant function can be 'promoted' to a regional integral by assigning to every trajectory sufficiently near the periodic trajectory the value $f(\xi(\mathbf{p})\eta(\mathbf{p}))$, where, as before, $\mathbf{p}$ denotes the point in the $\rho, \dot{\rho}$ plane at which the trajectory crosses the surface of section.

However, if there is a homoclinic point, we know that the regional integral cannot be extended to a global integral, nor can the integral be extended indefinitely along a trajectory. For if we follow a trajectory which starts near the periodic trajectory, we will find that it wanders away from the periodic trajectory for a while, since it must repeatedly intersect the surface of section at points near W_u. It must then again return to the vicinity of the periodic trajectory, since it must also repeatedly intersect the surface of section at points near W_s. But, because of homoclinic oscillations, it will return in general with a different value of the local integral, and hence the local integral cannot be extended globally.

The third case of regional integrability is even more complicated. So far we have not discussed elliptic fixed points. It can readily be verified that if $\mathbf{e}$ is an elliptic fixed point, the effect of L_e is to 'twist' points around on ellipses just as L_h moved points about on hyperbolas. We might suspect that in this case M would have as invariant curves a family of slightly distorted ellipses concentric about $\mathbf{e}$. *Moser* [1962] and *Arnold* [1961] have shown that under quite general conditions, closed invariant curves analogous to distorted ellipses do exist in every neighborhood of $\mathbf{e}$. This means that a periodic orbit corresponding to an elliptic fixed point is completely stable in the sense that an orbit started near the periodic orbit will always remain near this orbit. For if it were to wander away, it would produce points in the surface of section lying both inside and outside the invariant curve. This is forbidden by topological arguments, since M is invertible and continuous.

However, apart from exceptional cases, the invariant curves do not belong to a continuous family as one might have incorrectly guessed. Instead they are isolated. *Zehnder* [1973] has shown that quite generally there are hyperbolic fixed points of high powers of M between any two closed invariant curves, and these hyperbolic fixed points have manifolds W_u and W_s which intersect at an angle to produce homoclinic points! Thus there are also homoclinic points in every neighborhood of an elliptic fixed point in the general case!

Elliptic fixed points of M have been found numerically for the Stǿrmer problem. Indeed, there is one at (0.93913263, 0) in Figure 13 between the two pieces of W_u and W_s. It is expected that Zehnder's result will also hold in this case, but this surmise has not been explored numerically.

Braun [1970*b*] has used the Moser 'twist map' theorem to show that for sufficiently low energies (unfortunately too small by many orders of magnitude at present for physical applications) there are also closed invariant curves of M around O, the point corresponding to the orbit to the origin. These curves have the general shape of the boundary of the physical region; e.g., they are expected to look like the shape portrayed in Figure 15. In fact, there may be a closed invariant curve near the points shown. This conjecture cannot be proven numerically, for it is impossible to rule out 'homoclinic like' oscillations which are too small to be detected.

Braun's result is very important, for again M cannot map points from the exterior of a closed invariant curve into the interior. We note that points in the ρ, $\dot{\rho}$ plane near the boundary of the physical region correspond to nearly equatorial orbits, while points near O correspond to orbits which mirror far down the thalweg. Thus the existence of a closed invariant curve shows that orbits corresponding to points outside the invariant curve must always mirror at latitudes less than a fixed latitude for all time. Consequently, it is possible to infer long-time behavior for a whole class of orbits (for sufficiently small W_0^2) even though the magnetic moment series is divergent. Braun's result is of importance because it shows that in the Størmer idealization, classes of orbits in the radiation belts are bounded away from hitting the earth's atmosphere and hence are eternally confined. If his results could be extended to larger values of W_0^2, one would have a significant beginning of a theory of long-time behavior for orbits of physical interest.

Suppose that M does have a closed invariant curve. What does this mean for integrability and for the Hamilton-Jacobi equation? First, if we consider all trajectories corresponding to points in the ρ, $\dot{\rho}$ plane on the invariant curve, it is clear that they will form the surface of a torus in the four-dimensional phase space. The closed invariant curve is the intersection of the torus with the surface of section.

Second, it is possible to find a function I(p) which is constant on the invariant curve. It is again easy to see that this function can be promoted to an integral $I(p, q)$ on the torus just as we did earlier for the invariant function near the hyperbolic point. However, from Zehnder's result we expect $I(p, q)$ will in general satisfy the Liouville equation (40) only on the torus and not nearby. Thus in this case we have at most an integral in the two-dimensional region formed by the torus in the four-dimensional phase space. Similarly, it can be shown that the Hamilton-Jacobi equation has a periodic solution of the action-angle type for certain values of α_1, α_2 corresponding to the torus and no periodic solution nearby [*Moser*, 1969].

There is one last point to be discussed. We have learned that the Størmer problem is not globally integrable. Might it not still be the case that someday someone will write down in explicit form a pair ρ, z of functions of the initial conditions and of time which satisfy the Størmer equations of motion? This somewhat ill-defined possibility seems very unlikely. We know that the functions $\rho(t)$ and $z(t)$ must produce an area-preserving map M with homoclinic points. No one has ever explicitly exhibited any pair of functions with this property and which also arise from equations of motion derived from any Hamiltonian, let alone the Størmer Hamiltonian. Furthermore, it has been shown [*Smale*, 1965; *Nitecki*, 1971; *Moser*, 1973] that if a mapping has a homoclinic point, then it is possible to embed topologically within its action a sequence shift σ.

The definition of a sequence shift requires a few introductory words: Let S be the set of all doubly infinite sequences

$$s = (\cdots, s_{-1}, s_0, s_1, \cdots) \tag{56}$$

whose entries are elements of some countable set A. We will say that two separate sequences s and s' are close by if $s_n = s_n'$ for $|n| < N$, where N is a large number. This notion of closeness introduces a topology into S. A sequence shift σ is now defined by the rule

$$\sigma(s)_k = s_{k-1} \tag{57}$$

that is, σ shifts any given sequence s one notch to the right.

The sequence shift has served as a model for random behavior in ergodic theory [*Billingsley*, 1965]. For example, let s and s'' be two arbitrary sequences. Then it is easy to construct an 'interpolating' sequence s' which is near s in the context of the topology defined in the preceding paragraph and which also has the property that the result of $2N$ shifts, $\sigma^{2N}(s')$, is a sequence near s''. The fact that the shift can be topologically embedded within M can be used to show that the action of high powers of M is exceedingly complicated near a homoclinic point. Correspondingly, the motion in the Størmer problem near hyperbolic periodic orbits must be exceedingly complicated to the point that it defies explicit long-time representation.

Acknowledgments. The first author is indebted to the Institut des Hautes Etudes Scientifiques, where this work was begun. He wishes to thank N. Kuiper and L. Michel for their fine hospitality. We are both indebted to the University of Maryland Computer Science Center, which provided the computer time essential to this research. A Tektronix Graphics terminal purchased under grant AEC-AT(40-1)3443 was employed to prepare most of the figures for this paper. It greatly facilitated our work. Finally, we wish to thank G. Contopoulos and J. Moser for many helpful discussions. This work was supported in part by the National Science Foundation under grant GP-41822X.

The Editor thanks D. J. Southwood and D. P. Stern for their assistance in evaluating this paper.

REFERENCES

Abraham, R., *Foundations of Mechanics*, p. 142, W. A. Benjamin, New York, 1967.

Arnold, V. I., On the stability of positions of equilibrium of a Hamiltonian system of ordinary differential equations in the general elliptic case, *Sov. Math.*, *2*, 247, 1961.

Arnold, V. I., and A. Avez, *Ergodic Problems of Classical Mechanics*, appendices 26 and 31, W. A. Benjamin, New York, 1968.

Billingsley, P., *Ergodic Theory and Information*, John Wiley, New York, 1965.

Birkhoff, G. D., Surface transformations and their dynamical applications, *Acta Math.*, *43*, 1, 1920.

Birkhoff, G. D., *Dynamical Systems*, p. 255, American Mathematics Society, New York, 1927.

Braun, M., Structural stability and the Størmer problem, *Indiana Univ. Math. J.*, *20*, 469, 1970*a*.

Braun, M., Particle motion in a magnetic field, *J. Differential Equations*, *8*, 294, 1970*b*.

Contopoulos, G., A classification of the integrals of motion, *Astrophys. J.*, *138*(4), 1297, 1963.

Contopoulos, G., and L. Vlahos, Integrals of motion and resonances in a dipole magnetic field, *J. Math. Phys.*, *16*, 1469, 1975.

DeVogelaere, R., Surface de section dans le problème de Stōrmer, *Bull. Cl. Sci. Acad. Roy. Belg.*, *40*, 705, 1954.

DeVogelaere, R., *Contributions to the Theory of Nonlinear Oscillations*, vol. 4, edited by S. Lefshetz, p. 53, Princeton University Press, Princeton, N. J., 1958.

Dragt, A. J., Trapped orbits in a magnetic dipole field, *Rev. Geophys. Space Phys.*, *3*, 255, 1965. (Correction, *Rev. Geophys. Space Phys.*, *4*, 112, 1966.)

Finn, J. M., Integrals of canonical transformations and normal forms for mirror machine Hamiltonians, Ph.D. thesis, Univ. of Md., College Park, Md., 1974.

Gibson, G., W. Jordan, and E. Lauer, Containment of positrons in a mirror machine, *Phys. Rev. Lett.*, *5*,(4) 141, 1960.

Godart, O., *Periodic Orbits, Stability, and Resonances, Proceedings of Symposium at University of Sao Paulo*, edited by G. E. Giacaglia, p. 56, D. Reidel, Dordrecht, Netherlands, 1970.

Hadamard, J., Sur l'itération et les solutions asymptotiques des equations differentielles, *Bull. Soc. Math. France*, *29*, 224, 1901.

Hagihara, Y., *Celestial Mechanics*, vol. 1, p. 310, MIT Press, Cambridge, Mass., 1970.

Moser, J., The analytic invariants of an area preserving mapping near a hyperbolic fixed point, *Commun. Pure Appl. Math.*, *9*, 673, 1956.

Moser, J., On invariant curves of area-preserving mappings of an annulus, *Nachr. Akad. Wiss. Goettingen Math. Phys., Kl.*, 1, 1962.

Moser, J., *Proceedings of the International Conference on Functional Analysis and Related Topics*, p. 60, University of Tokyo Press, Tokyo, 1969.

Moser, J., *Stable and Random Motions in Dynamical Systems*, p. 46, Princeton University Press, Princeton, N. J., 1973.

Nitecki, Z., *Differentiable Dynamics, An Introduction to the Orbit Structure of Diffeomorphisms*, p. 151, MIT Press, Cambridge, Mass., 1971.

Northrop, T. G., *The Adiabatic Motion of Charged Particles*, John Wiley, New York, 1963.

Northrop, T. G., and E. Teller, Stability of the adiabatic motion of charged particles in the earth's field, *Phys. Rev.*, *117*(1) 215, 1960.

Poincaré, H., *Les Methodes Nouvelles de la Mechanique Celeste*, vol. 3, chap. 33, Gauthier-Villars, Paris, 1892. (Reprinted by Dover, New York, 1957.)

Rossi, B., and S. Olbert, *Introduction to the Physics of Space*, p. 45, McGraw-Hill, New York, 1970.

Smale, S., *Differential and Combinatorial Topology*, edited by S. Cairns, p. 63, Princeton University Press, Princeton, N. J., 1965.

Størmer, C., *The Polar Aurora*, Oxford at the Clarendon Press, London, 1955.

Titchmarsh, E. C., *Theory of Functions*, p. 88, Oxford University Press, New York, 1939.

Whittaker, E. T., *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, chap. 14, Cambridge University Press, London, 1937.

Wintner, *Analytical Foundations of Celestial Mechanics*, p. 144, Princeton University Press, Princeton, N. J., 1947.

Zehnder, E., Homoclinic points near elliptic fixed points, *Commun. Pure Appl. Math.*, *26*, 131, 1973.

(Received July 30, 1975;
accepted November 26, 1975.)