CONTEMPORARY MATHEMATICS

81

Hamiltonian Dynamical Systems

AMS-IMS-SIAM Joint Summer Research Conference on Hamiltonian Dynamical Systems
June 21–27, 1987
University of Colorado

Kenneth R. Meyer
Donald G. Saari
Editors

American Mathematical Society

http://dx.doi.org/10.1090/conm/081

Recent Titles in This Series

164 **Cameron Gordon, Yoav Moriah, and Bronislaw Wajnryb, Editors,** Geometric topology, 1994

163 **Zhong-Ci Shi and Chung-Chun Yang, Editors,** Computational mathematics in China, 1994

162 **Ciro Ciliberto, E. Laura Livorni, and Andrew J. Sommese, Editors,** Classification of algebraic varieties, 1994

161 **Paul A. Schweitzer, S. J., Steven Hurder, Nathan Moreira dos Santos, and José Luis Arraut, Editors,** Differential topology, foliations, and group actions, 1994

160 **Niky Kamran and Peter J. Olver, Editors,** Lie algebras, cohomology, and new applications to quantum mechanics, 1994

159 **William J. Heinzer, Craig L. Huneke, and Judith D. Sally, Editors,** Commutative algebra: Syzygies, multiplicities, and birational algebra, 1994

158 **Eric M. Friedlander and Mark E. Mahowald, Editors,** Topology and representation theory, 1994

157 **Alfio Quarteroni, Jacques Periaux, Yuri A. Kuznetsov, and Olof B. Widlund, Editors,** Domain decomposition methods in science and engineering, 1994

156 **Steven R. Givant,** The structure of relation algebras generated by relativizations, 1994

155 **William B. Jacob, Tsit-Yuen Lam, and Robert O. Robson, Editors,** Recent advances in real algebraic geometry and quadratic forms, 1994

154 **Michael Eastwood, Joseph Wolf, and Roger Zierau, Editors,** The Penrose transform and analytic cohomology in representation theory, 1993

153 **Richard S. Elman, Murray M. Schacher, and V. S. Varadarajan, Editors,** Linear algebraic groups and their representations, 1993

152 **Christopher K. McCord, Editor,** Nielsen theory and dynamical systems, 1993

151 **Matatyahu Rubin,** The reconstruction of trees from their automorphism groups, 1993

150 **Carl-Friedrich Bödigheimer and Richard M. Hain, Editors,** Mapping class groups and moduli spaces of Riemann surfaces, 1993

149 **Harry Cohn, Editor,** Doeblin and modern probability, 1993

148 **Jeffrey Fox and Peter Haskell, Editors,** Index theory and operator algebras, 1993

147 **Neil Robertson and Paul Seymour, Editors,** Graph structure theory, 1993

146 **Martin C. Tangora, Editor,** Algebraic topology, 1993

145 **Jeffrey Adams, Rebecca Herb, Stephen Kudla, Jian-Shu Li, Ron Lipsman, and Jonathan Rosenberg, Editors,** Representation theory of groups and algebras, 1993

144 **Bor-Luh Lin and William B. Johnson, Editors,** Banach spaces, 1993

143 **Marvin Knopp and Mark Sheingorn, Editors,** A tribute to Emil Grosswald: Number theory and related analysis, 1993

142 **Chung-Chun Yang and Sheng Gong, Editors,** Several complex variables in China, 1993

141 **A. Y. Cheer and C. P. van Dam, Editors,** Fluid dynamics in biology, 1993

140 **Eric L. Grinberg, Editor,** Geometric analysis, 1992

139 **Vinay Deodhar, Editor,** Kazhdan-Lusztig theory and related topics, 1992

138 **Donald St. P. Richards, Editor,** Hypergeometric functions on domains of positivity, Jack polynomials, and applications, 1992

137 **Alexander Nagel and Edgar Lee Stout, Editors,** The Madison symposium on complex analysis, 1992

136 **Ron Donagi, Editor,** Curves, Jacobians, and Abelian varieties, 1992

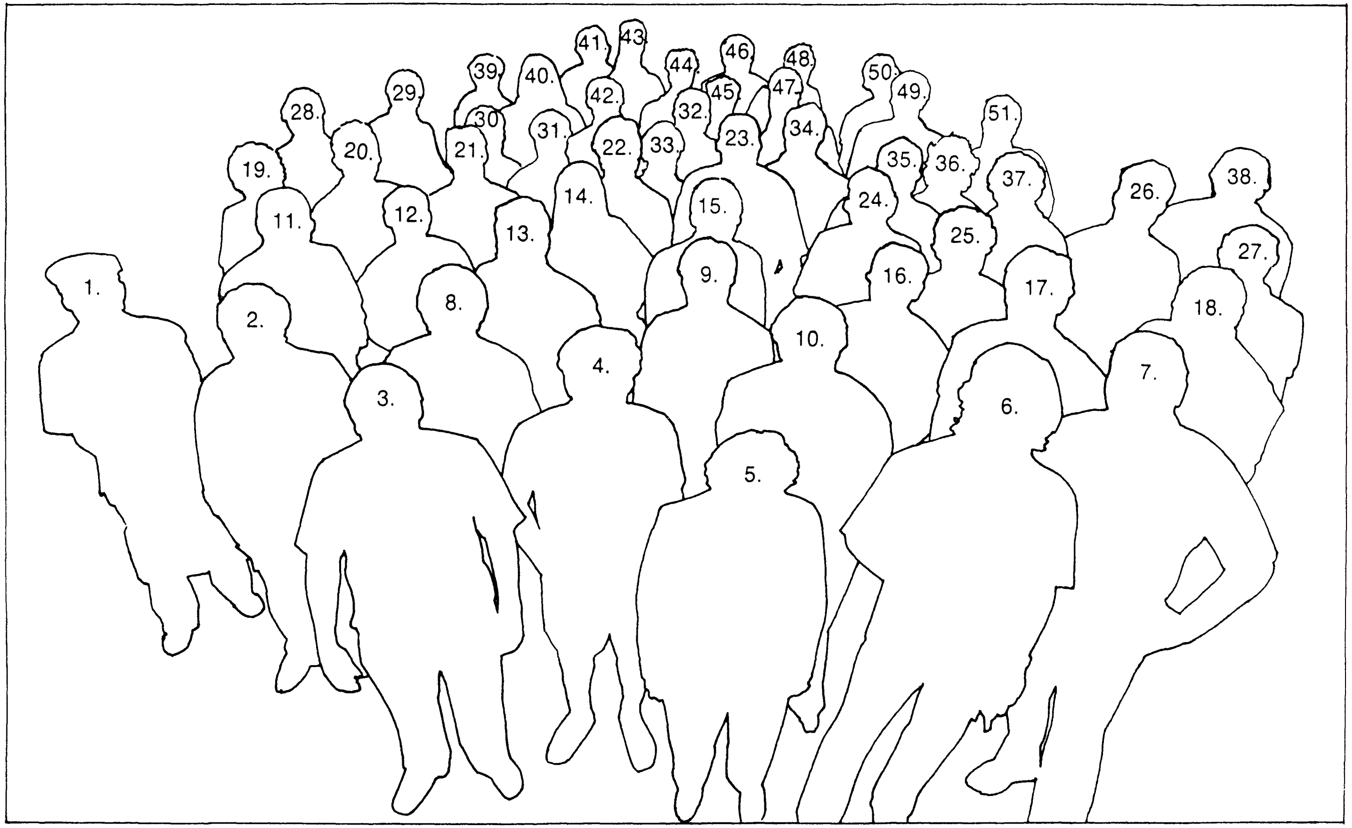
135 **Peter Walters, Editor,** Symbolic dynamics and its applications, 1992

134 **Murray Gerstenhaber and Jim Stasheff, Editors,** Deformation theory and quantum groups with applications to mathematical physics, 1992

(*Continued in the back of this publication*)

This page intentionally left blank

Hamiltonian Dynamical Systems



1. Jung Rno, 2. J. D. Fernandez, 3. Neal Hulkower, 4. Alberto Baider, 5. Carol Williams, 6. Marc Murison, 7. Ken Meyer, 8. En-Bing Lin, 9. Pau Atela, 10. Andre Deprit, 11. Mark Levi, 12. Paul Rabinowitz, 13. Mark Adler, 14. Dusa McDuff, 15. Anthony Bloch, 16. Sebastian Ferrer, 17. Rick Moeckel, 18. Bob Devaney, 19. E. Perez-Chavela, 20. Mohamed Elbialy, 21. James Montaldi, 22. Dan Offin, 23. Don Saari, 24. Mort Brown, 25. David Fried, 26. Clark Robinson, 27. Bob Easton, 28. John Bryant, 29. Anand Sivaranakrishahn, 30. Unknown, 31. David Rod, 32. H. G. Poinkhoff, 33. Dick Cushman, 34. John Franks, 35. Jim Reineck, 36. Roger Kraft, 37. Chris McCord, 38. Robert Winters, 39. Larry Bates, 40. Christopher Cotter, 41. Richard Churchill, 42. Chris Gole, 43. Dieter Schmidt, 44. Dan Goroff, 45. Dick Hall, 46. Greg Buck, 47. Jan Sanders, 48. Joel Robbin, 49. Phil Boyland, 50. Steve Pelikan, 51. David Hart

This page intentionally left blank

CONTEMPORARY MATHEMATICS

81

Hamiltonian Dynamical Systems

AMS-IMS-SIAM Joint Summer
Research Conference on
Hamiltonian Dynamical Systems
June 21–27, 1987
University of Colorado

Kenneth R. Meyer
Donald G. Saari
Editors

American Mathematical Society
Providence, Rhode Island

The AMS-IMS-SIAM Joint Summer Research Conference in the Mathematical Sciences on Hamiltonian Dynamical Systems was held at the University of Colorado, Boulder on June 21–27, 1987, with support from the National Science Foundation, Grant DMS-8613199.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 58F, 70H, 70F.

Library of Congress Cataloging-in-Publication Data

AMS-IMS-SIAM Joint Summer Research Conference in the Mathematical Sciences on Hamiltonian Dynamical Systems (1987: University of Colorado)

Hamiltonian dynamical systems: proceedings of the AMS-IMS-SIAM joint summer research conference held June 21–27, 1987 with support from the National Science Foundation/Kenneth R. Meyer and Donald G. Saari, editors.

p. cm.—(Contemporary mathematics, ISSN 0271-4132; v. 81) "The AMS-IMS-SIAM Joint Summer Research Conference in the Mathematical Sciences on Hamiltonian Dynamical Systems was held at the University of Colorado, Bounder, Colorado, from June 21–June 27, 1987, with support from the National Science Foundation, grant DMS-8613199"–T.p. verso.

Bibliography: p.

ISBN 0-8218-5086-5 (alk. paper)

1. Hamiltonian systems—Congresses. 2. Mechanics, Celestial-Mathematics—Congresses. I. Myer, Kenneth R. (Kenneth Ray), 1937– . II. Saari, D. (Donald) III. American Mathematical Society. IV. Title. V. Series: Contemporary mathematics (American Mathematical Society); v. 81.

QA614.83.A47 1988 88-26831
515.3′52—dc 19 CIP

Printed in the United States of America.

The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.

Printed on recycled paper.

This volume was printed from copy prepared by the author.

10 9 8 7 6 5 4 3 2 99 98 97 96 95 94

CONTENTS

PREFACE

In June, 1987, a one-week conference on Hamiltonian Dynamics was held at the University of Colorado in Boulder, Colorado. This conference was one of the series organized under the auspices of the AMS-SIAM-IMS Conference Series. The organizing committee consisted of Ken Meyer (Co-Chair), Don Saari (Co-Chair), Richard Hall, Tudor Ratiu, and Alan Weinstein. The articles in this volume are contributions by participants whose papers, after a review, were viewed as contributing to this research area and representing the general thrust of the conference. The papers vary from being essentially expository descriptions of recent developments to being fairly technical with new results. Collectively, they provide a good survey of contemporary work in this area.

It is highly appropriate that this conference was held during the summer of 1987 -- the three hundredth anniversary of the publication of I. Newton's *Principia Mathematica.* (We would like to claim that this timing was due to careful planning and design; in fact, it was purely by coincidence.) *Principia*, which developed and applied the science of dynamics to the emerging problems of orbital mechanics, gave birth to the field of celestial mechanics and, subsequently, to Hamiltonian dynamics. The area of celestial mechanics was well represented at this conference. Many of the talks emphasized either topics directly concerned with the Newtonian n-body problem, the three body problem, the artificial earth satellite, etc., or those dynamical issues, such as integrability, KAM, and extensions of the Poincare-Birkhoff results, that emerged from celestial mechanics and extend to wider classes of dynamical systems.

In addition to topics related to celestial mechanics, this conference brought together researchers from a wide spectrum of areas of contemporary research in Hamiltonian dynamics. Just a small sample includes the existence of periodic orbits with variation methods, twist and annulus maps, stable manifold theory, almost periodic motion, heteroclinic and homoclinic orbits, etc. It is our hope that by bringing together papers from such a diverse range of topics will serve as a stimulant for further development in Hamiltonian dyamics.

Kenneth R. Meyer
Cincinnati, Ohio

Donald G. Saari
Evanston, Illinois

March, 1988

http://dx.doi.org/10.1090/conm/081/986254

Contemporary Mathematics
Volume 81, 1988

Some Qualitative Features of the Three-Body Problem

Richard Moeckel*

1.Introduction. This paper is a survey of certain features of the three-body problem that I find particularly appealing. The emphasis will be on presenting the "big picture" of what is going on in the three-body problem. It is shown in figure 1.

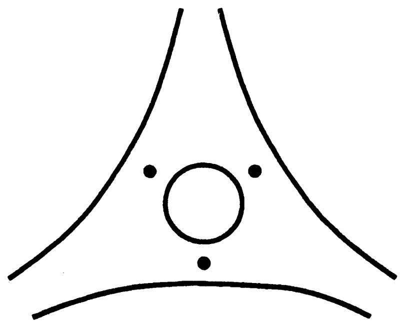


FIGURE 1: The Three-Body Problem

This picture was first drawn for me by Charles Conley (probably on a napkin at the Gourmandaise restaurant in Madison, Wisconsin) when I was a graduate student and I have spent a good deal of time since then trying to figure it out. It turns out that it is somewhat oversimplified but it captures the main features !

Looking at figure 1, the expert will wonder which three-body problem it depicts. We will consider only the planar three-body problem with unrestricted masses. Many interesting results about the restricted problem are omitted. One of the main purposes of this paper and of the lecture from which it derives is to show how little is really known about this problem. Thus as a counterpoint to the theorems we will list many open problems (some of which may actually be solvable).

2.The Equations. The planar three-body problem concerns the motion of three point masses in a plane under the influence of their mutual gravitational attraction. We let $q_j \in \mathbf{R}^2$

* Research supported by the National Science Foundation

stand for the position of the j^{th} point, $p_j \in \mathbf{R}^2$ for its momentum and $m_j \in \mathbf{R}^+$ for its mass. The system is governed by the Hamiltonian function:

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

where $q = (q_1,q_2,q_3) \in \mathbf{R}^6$, $p = (p_1,p_2,p_3) \in \mathbf{R}^6$, $M = \mathrm{diag}(m_1,m_1,m_2,m_2,m_3,m_3)$ and:

$$U(q) = \frac{m_1 m_2}{|q_1 - q_2|} + \frac{m_1 m_3}{|q_1 - q_3|} + \frac{m_2 m_3}{|q_2 - q_3|} \; .$$

U(q) is minus the Newtonian gravitational potential energy. Hamilton's differential equations are:

(2.1)
$$\dot{q} = M^{-1} p$$
$$\dot{p} = \nabla U(q) \; .$$

These equations define a dynamical system in $\mathbf{R}^{12}$. However, it is possible to reduce the problem to a five-dimensional system by making use of the well-known integrals of motion.

The first integral is the total momentum. We assume without loss of generality that:

$$p_1 + p_2 + p_3 = 0 \; .$$

This assumtion implies that the center of mass will be constant and we can take it to be the origin:

$$m_1 q_1 + m_2 q_2 + m_3 q_3 = 0 \; .$$

These equations restrict the momentum vector,p, and position vector,q, to four-dimensional subspaces of $\mathbf{R}^6$ so together they reduce the dimension of the system by 4.

Next we consider angular momentum. The equations 2.1 are invariant under simultaneous rotation of all positions and momenta in $\mathbf{R}^2$. As a result, total angular momentum is constant:

$$p_1 \times q_1 + p_2 \times q_2 + p_3 \times q_3 = \omega \; .$$

Here we view the cross products as scalars. This reduces the dimension of the system by 1. Since the system is symmetric under rotations, we can pass to a quotient space in which all vectors (q,p) which differ only by a simultaneous rotation of all q_j and p_j are identified. This eliminates 1 more dimension.

Finally, the Hamiltonian itself is the total energy of the system and is conserved:

$$H(q,p) = \frac{1}{2}p^T M^{-1} p - U(q) = h \; .$$

This eliminates 1 more dimension. All of these equations together define a five-dimensional manifold, $M(h,\omega)$, the quotiented energy and angular momentum manifold.

The topology of these manifolds depends on the energy, h, the angular momentum, ω, and the masses, m_j [Eas,Sm].

To facilitate the geometrical discussion below, it is convenient to introduce a coordinate system discovered by McGehee [Mc1]. Since the center of mass is at the origin, the moment of inertia about the origin plays a central role:

$$\mathfrak{I} = q^T M q = m_1 |q_1|^2 + m_2 |q_2|^2 + m_3 |q_3|^2 .$$

The variable $r = \sqrt{\mathfrak{I}}$ will be the radial variable of a kind of polar coordinate system in $\mathbf{R}^6$. It is a measure of the size of the triangle formed by the three point masses; in particular, r = 0 represents a triple collision at the origin. The normalized position vector, $s = \frac{q}{r}$ measures the shape and angular position of the triangle. Note that by definition, s satisfies: $s^T M s = 1$. In the quotient manifold we lose information about the angle and we think of s as representing only shape. It turns out to be advantageous to normalize momentum differently. Define $z = \sqrt{r}\, p$. The variables (r,s,z) are superior to (q,p) because of their behavior near the triple collision singularity. The energy and angular momentum equations in these coordinates are:

$$H(s,z) = \frac{1}{2} z^T M^{-1} z - U(s) = h r \tag{2.2}$$

$$z_1 \times s_1 + z_2 \times s_2 + z_3 \times s_3 = \omega \sqrt{r} .$$

We are able to factor the r dependence out of the Hamiltonian because of the homogeneity of the Newtonian potential function. The possibility of writing the energy equation in this way motivates the choice of scaling for the momentum. When the differential equations are expressed in these coordinates it is found that they contain a singular common factor of $r^{-\frac{3}{2}}$. Multiplying through by a factor of $r^{\frac{3}{2}}$ changes only the parametrization of solution curves; the result is:

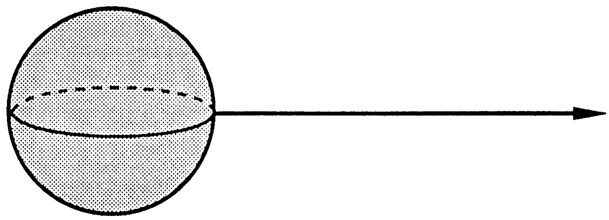
$$\begin{aligned} r' &= v r \\ s' &= z - \tfrac{1}{2} v s \\ z' &= \nabla U(s) + \tfrac{1}{2} v z \end{aligned} \tag{2.3}$$

where $v = s \cdot z$ and ' denotes differenttiation with respect to the new parameter. Note that the last two equations, describing the rate of change of the shape and normalized momentum, are independent of r.

3. Hill's Regions. We have reduced the dimension of the dynamical system describing the planar three-body problem from twelve dimensions to five. Unfortunately, five is still too many to sketch. Since the behavior of the size and shape of the triangle formed by the three bodies has a more direct intuitive meaning than the behavior of the momenta we will focus attention on the configuration space. Define:

$$C = \{ (r,s) : r \geq 0, s^T M s = 1, m_1 s_1 + m_2 s_2 + m_3 s_3 = 0 \} / S^1$$

the space of all admissable configurations with the rotation symmetry quotiented out. It is not difficult to see that this space is homeomorphic to $R^+ \times S^2$; the two equations in s define a three-dimensional ellipsoid in $\mathbf{R}^6$ and the quotient space of this ellipsoid under the circle action is homeomorphic to a two-sphere. We can visualize this as in figure 2.



r = 0 (triple collision)

FIGURE 2: Configuration Space

Once again we note that r represents the size of the triangle formed by the three bodies while s represents its shape. A ray represents a family of similar triangle of varying size. We will draw the shape two-sphere in more detail (figure 3). There are several interesting features. First, the collinear "triangles" form a circle (depicted here as the equator) in the two-sphere. The isosceles triangles form three circles distinguished by which mass lies on the axis of symmetry. These three circles meet at the equilateral triangle configurations (shown here as the poles); note that there are two rotationally inequivalent equilateral triangles with the masses 1,2,3 appearing in either clockwise or counterclockwise fashion. Each circle of isosceles triangles intersects the collinear circle in two points; one represents a collinear configuration with one mass at the midpoint of the other two and the other represents a double collision configuration.

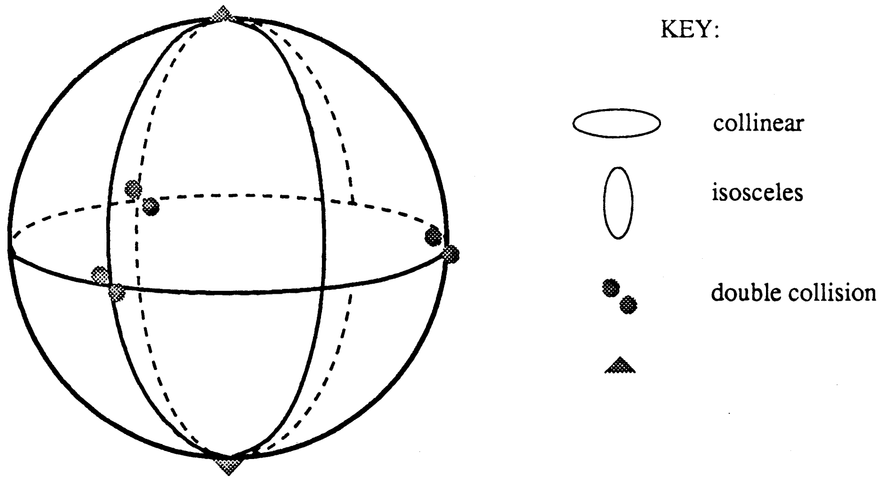


FIGURE 3 : The Shape Sphere

It was G.W. Hill who first realized that the energy and angular momentum integrals impose constraints on the configuration [H]. Although he worked in the restricted three-body problem, the idea is fruitful in the planar problem as well. Define the Hill's regions:

$$C(h,\omega) = \{ (r,s) \in C : \text{for some } z \in \mathbf{R}^6, (r,s,z) \in M(h,\omega) \} .$$

Thus $C(h,\omega)$ is the projection onto the configuration space of the integral manifold $M(h,\omega)$.

The Hill's regions, $C(h,\omega)$, will provide an organizing center for this paper. We will study how they vary as the energy and angular momentum are changed. For each choice of h and ω we will describe some features of the dynamical system on $M(h,\omega)$ and how they look in $C(h,\omega)$. The shapes of the Hill's regions will suggest several open problems.

We do not need to survey the entire two-parameter family of Hill's regions. First, we can restrict attention to negative energies: $h < 0$. If $h \geq 0$ all orbits scatter to infinity in both time directions, so no recurrence is possible and the problem holds little interest. Second, it is easy to show that the dynamics depends only on the quantity $\lambda = - h\,\omega^2$ (of course the dynamics also depends on the choice of masses). Thus to see the whole story it suffices to fix some $h < 0$ and let ω vary over $[0,\infty)$.

Before turning to the case by case description we derive the inequalitites characterizing Hill's regions. The energy and angular momentum equations 2.2 are the key to deriving these. The energy equation alone imposes some restrictions on the configuration; namely, since the kinetic energy term, $\frac{1}{2}z^T M^{-1} z$, is non-negative we have:

$$U(s) \geq |h|\, r \tag{3.1}$$

and so for any fixed shape s_0, the size is restricted to the range $0 \leq r \leq \frac{U(s_0)}{|h|}$. Thus all configurations arising from a state with energy h lie in the region shown in figure 4.

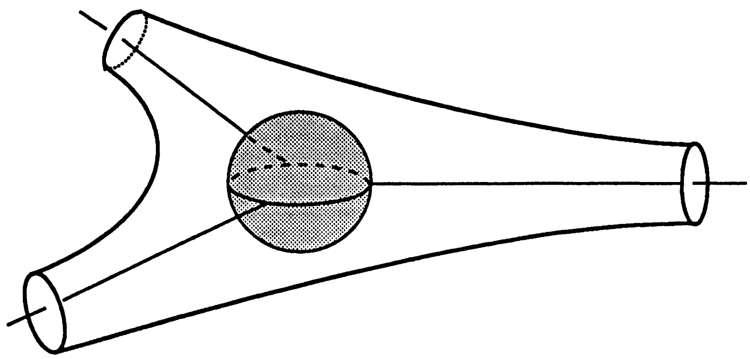


FIGURE 4 : Constraints due to energy

The potential function U(s) on the shape sphere has maximum value ∞ at the double collision configurations and attains minima at the equilateral configurations. Thus the energy imposes no restriction on the size of a double collision configuration but rules out all sufficiently large triangles with any other shape, the greatest restrictions being on the equilateral triangles.

The inequality 3.1 was derived from the observation that the kinetic energy is non-negative. One can show without difficulty that when the angular momentum is fixed the following sharper estimate holds:

$$\frac{1}{2}z^T M^{-1} z \geq \frac{1}{2}\frac{\omega^2}{r}.$$

When this is plugged into the energy equation we find:

$$U(s) \geq |h|\, r + \frac{\omega^2}{2r} \tag{3.2}$$

which characterizes $C(h,\omega)$.

$C(h,\omega)$ is a solid region in $R^+ \times S^2$. Its boundary is given by the equality in 3.2. Since this is a quadratic equation for r given s we see that $\partial C(h,\omega)$ lies in two sheets over some subset of the shape two-sphere. The projection of $\partial C(h,\omega)$ to the two-sphere is the set of all s such that 3.2 holds for some $r \geq 0$. Minimizing the right side of 3.2 gives:

$$U(s) \geq \sqrt{2\,|h|\,\omega^2} = \sqrt{2\,\lambda} \tag{3.3}$$

which defines the projection. Thus the Hill's region lies over a region of the shape two-sphere bounded by an equipotential curve. Its boundary lies in two sheets over the projection and these two sheets come together over the equipotential curve. These observations will underlie the pictures which follow.

4. Large Angular Momentum. We will begin our survey with the case of large ω, or equivalently, large λ. Inequality 3.3 forces the shape, s, to lie in one of three disks around the double collision configurations (where $U(s) = \infty$). This means that two of the bodies are very close relative to their distance from the third body; we call this a tight binary configuration.The Hill's region consists of three lobes over the disks. It is shown in figure 5.

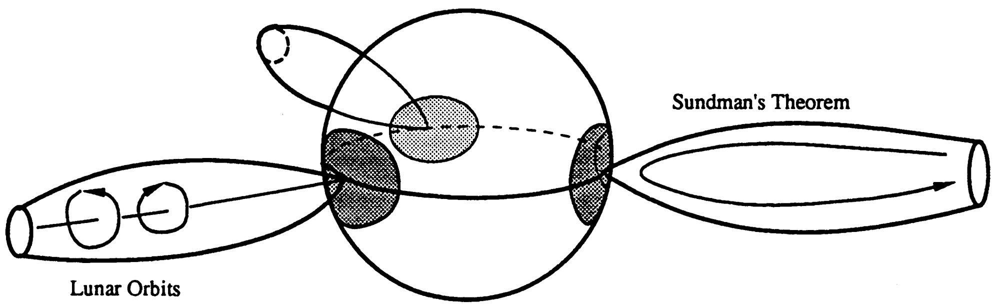


FIGURE 5 : Hill's Region for Large Angular Momentum

The lobes touch triple collision ($r = 0$) and infinity over the double collision configurations. The behavior of orbits near these two extremes of r is similar for all non-zero angular momenta so we will describe this before turning to the features specific to the large angular momentum case.

One of the nicest results in the whole theory is:

Theorem: Let the angular momentum be non-zero. Then any orbit passing sufficiently close to triple collision is of the following type: the configuration is a tight binary for all time and the short side of the triangle remains bounded while the other two sides tend to infinity in both forward and backward time.

This theorem is due to Sundman [Su] with refinements due to Birkhoff [Bir]. In particular, it implies that triple collision is impossible for $\omega \neq 0$. An orbit obeying Sundman's theorem is shown in the right lobe of figure 5. Although we will not draw them, such orbits occur in all of the pictures referring to non-zero angular momenta.

At the other end of the lobe we have the two-body problem at infinity. As $r \rightarrow \infty$, 3.1 shows that the configuration is forced into tight binary. Although it is conceivable a priori that all three sides of the triangle could become infinite, an appeal to the unrescaled energy equation shows that the short side remains bounded while the other two become infinite. It stands to reason that the influence of the third mass on the binary will become negligible and that the binary will behave essentially as a two-body problem. In fact, using rescalings similar to those in section 2 it is possible to paste a copy of the two-body problem onto each lobe of $M(h,\omega)$ at infinity [Mc2,Mc-Eas,Rob]. Actually, there are many two-body problems at infinity distinguished by the asymptotic speed of separation of the binary from the third mass. Intuitively, there are three cases. Either the third mass has just enough energy to escape from the binary and so reaches infinity with zero asymptotic speed (parabolic case) or it has plenty of energy and reaches infinity with positive asymptotic speed (hyperbolic case) or it does not have enough energy and returns for another approach to the binary (elliptic case). Clearly the parabolic case separates the other two. It is shown in the references above that the set of orbits tending parabolically to infinity forms a four-dimensional invariant manifold in $M(h,\omega)$ which we call the stable manifold of infinity (even though a whole open set of orbits tends to infinity hyperbolically). Similarly there is a four-dimensional unstable manifold of parabolic infinity. These can be viewed as the stable and unstable manifolds of an invariant three-sphere pasted onto $M(h,\omega)$, the parabolic two-body problem at infinity. These invariant sets are present for all angular momenta, even $\omega = 0$. Several open problems which can be posed for all ω concern these manifolds. Do there exist orbits homoclinic to parabolic infinity ? Do there exist orbits which tend parabolically to infinity in one time direction but which remain bounded in the other (capture or escape orbits) ? Do there exist orbits which oscillate to infinity, i.e., orbits with $\overline{\lim}\, r(t) = \infty$ but $\underline{\lim}\, r(t) < \infty$? Such orbits have been found in special cases of the three-body problem [Sit,Mos]. It is shown in [Mc-Eas] that if there are favorable homoclinic intersections of the invariant manifolds of parabolic infinity in the planar problem then capture/escape orbits and oscillation orbits also exist.

We turn now to those features which are specific to large angular momenta. First, recall that 3.3 forces the masses into a tight binary configuration from which they cannot escape. Thinking of the binary as the earth and moon and of the third body as the sun we see that the moon will always be a bounded distance from the earth and much closer to the earth than it is to the sun. This is a planar version of Hill's proof of the stability of the earth-moon system [H], one of the first applications of qualitative, geometrical reasoning to mechanics !

Continuing the earth-moon analogy we could look for "lunar" periodic orbits, that is, orbits such that the two bodies in the binary move in nearly circular orbits around their center of mass while the binary and the third mass move in nearly circular orbits around their center of mass. Such an orbit is called prograde if both circular motions have the same orientation and it is called retrograde if the orientations are different.

Theorem: For all sufficiently large angular momenta there is at least one prograde lunar orbit and at least one retrograde lunar orbit in each lobe of $M(h,\omega)$.

This result is due to Hill [H] in the restricted three-body problem and to Moulton [Mou1] in the planar problem. A nice proof can be found in [Mey]. A prograde and a retrograde lunar orbit are depicted in the front lobe of figure 5. The reason for requiring large ω is that if the configuration is a very tight binary then the third body can be viewed as a perturbation on the binary. It is an open question whether these orbits persist to lower angular momenta. In the restricted problem Conley [C] used the lunar orbits as the boundaries for an annular cross-section to the three-dimensional phase space. It is not clear how to generalize his work to the five-dimensional planar setting.

5. The First Critical ω. As we lower ω, the Hill's regions behave continuously until we reach a certain critical value. From the description in section 3 of how the Hill's regions lie over their projections to the shape sphere it is clear that bifurcations of the Hill's regions arise from bifurcations of the equipotential curves in the shape sphere. Thus critical values of ω

correspond to critical values of U(s). Critical points of U(s) are called central configurations or relative equilibria. We have already pointed out that the equilateral configurations are minima of U(s). There are also three collinear configurations which are saddle points of U(s). These are distinguished by the order in which the three bodies appear along the line; the exact spacing depends in a complicated way on the masses [Mou2]. The five central configurations of the three-body problem are shown in figure 6.

Each central configuration determines a restpoint in $M(h,\omega)$ for the corresponding critical ω. Actually these represent periodic orbits of the three-body problem for which the triangle formed by the three masses rotate rigidly around the center of mass; they appear as restpoints in $M(h,\omega)$ because we have quotiented out the rotational symmetry. These periodic orbits, which were discovered by Lagrange [Lag], are also shown in figure 8.

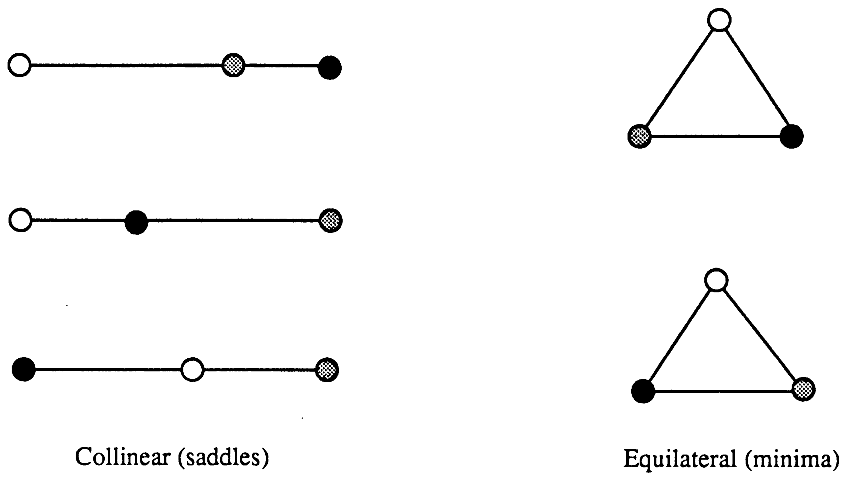
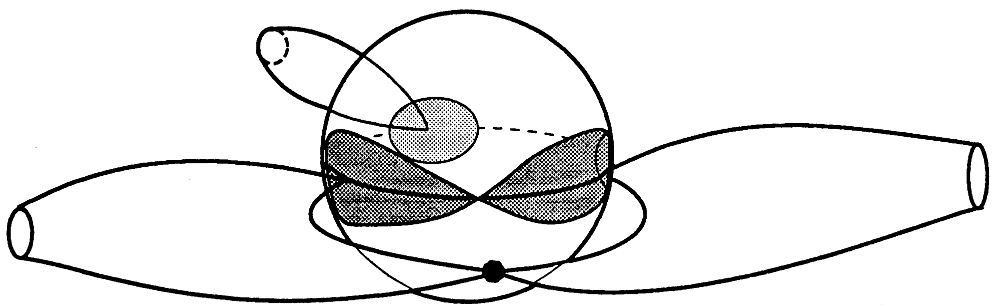


FIGURE 6 : Central Configurations

As we reach the first critical value of ω, two of the three disks around the double collisions in the shape sphere meet at one of the collinear saddle points,s_c. Over this saddle point at radius $\frac{U(s_c)}{2|h|}$, two of the three lobes of the Hill's region meet at the point representing the configuration of the Lagrangian periodic orbit. As we pass through the critical ω a tunnel opens between the two lobes. The critical Hill' region and tunnel are shown in figure 7. Other than the Lagrangian periodic orbit and the orbits arising from Sundman's theorem, not much is known about the dynamics for the critical angular momentum. However, an interesting invariant set lives in the tunnel. As we pass through the critical ω a hyperbolic invariant three-sphere bifurcates from the restpoint; this invariant set has four-dimensional stable and unstable manifolds. The existence of this invariant set follows from a linear analysis of the restpoint together with standard perturbation results for hyperbolic invariant manifolds. Inside the invariant three-sphere there is at least one periodic orbit, the elliptical orbit of Lagrange. This is the continuation to lower angular momenta of the circular orbit described above. The configuration remains similar to the central configuration but instead of rigidly rotating, the size expands and contracts as the three bodies orbit on similar ellipses around the center of mass obeying Kepler's laws for the two-body problem (figure 8). Since the shape remains constant, such an orbit appears in the Hill's region as a radial line segment over the central configuration; this segment runs completely across the tunnel. The elliptical Lagrange orbit appears in figure 7 along with a crazy orbit from the invariant three-sphere.



Lagrange Orbit

ω just below
critical

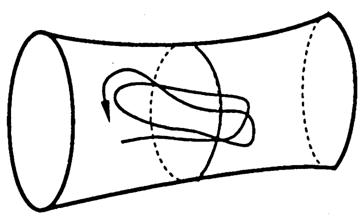


FIGURE 7 : First Critical Hill's Region and the Tunnel

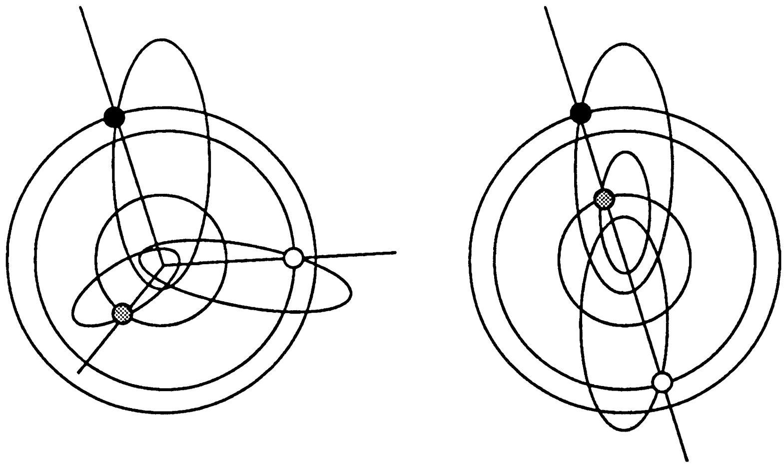


FIGURE 8 : Circular and Elliptical Lagrange Orbits

The shape of the Hill's region for ω below the critical level suggests a problem: do there exist binary exchange orbits, that is, orbits heteroclinic between the two lobes ? Such an orbit would exhibit different tight binary configurations in forward and backward time. More specifically, one could ask for heteroclinic orbits connecting the two parabolic infinities. Perhaps the invariant three-sphere in the tunnel is involved in such a network of homoclinic and heteroclinic orbits. An orbit connecting a three-sphere at infinity to the three-sphere in the tunnel would be an interesting type of capture orbit. A final open problem concerns the persistance of the invariant three-sphere or at least of some large invariant set as the angular momentum is lowered.

6. Below the Third Critical ω. The other two collinear central configurations are associated with bifurcations similar to the one described in section 5. At the critical levels, circular periodic orbits appear and develop into hyperbolic invariant three-spheres as the angular momentum is lowered further. New tunnels develop connecting the third lobe to the two which were already joined. After the third collinear orbit has developed the projection of the Hill's region is an equitorial band on the shape sphere and we have a Hill's region as in figure 9.

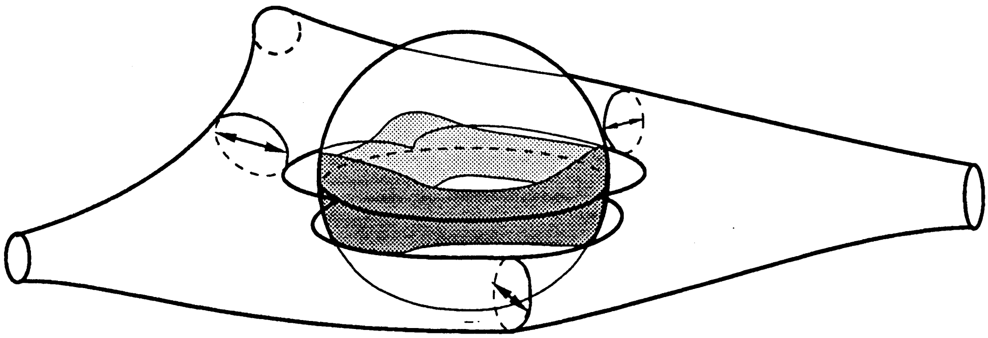


FIGURE 9 : Hill's Region Below the Third Critical ω

Very little is known about the dynamics for these intermediate values of angular momentum. About all that can be said is that there are three collinear, elliptical Lagrangian periodic orbits; these are the radial line segments crossing the tunnels in the figure. For parameters near the critical values there will also be invariant three-spheres but as mentioned in section 5, their persistance is in doubt. If there are interesting invariant sets in the three tunnels which behave in some sense (Conley index?) like hyperbolic invariant three- spheres, then there would be the possibility of heteroclinic connections. A less daunting open problem is suggested by the topology of the Hill's region. Do there exist periodic orbits which are homotopically nontrivial in the sense that they run around the Hill's region passing through all three tunnels to form a noncontractible closed curve ?

7. Below the Last Critical ω. As we lower ω still further the equitorial band on the shape sphere becomes wider until at the last critical value of ω it finally covers the north and south poles (the equilateral central configurations). At the critical level two restpoints develop in M(h,ω) corresponding to the two circular, equilateral Lagrangian periodic orbits. As we lower ω further these become elliptical just as in the collinear case. A detail of the bifurcation of the Hill's region over one of the equilateral points is shown in figure 10. After the bifurcation, the elliptical Lagrange orbits appear as radial line segments connecting the two sheets. Globally, the boundary of the Hill's region splits into two two-spheres (figure11). This topology persists all the way down to $\omega = 0$. As $\omega \rightarrow 0$, the inner surface converges to the triple collision sphere $r = 0$ and the Hill's region tends to the region of figure 4.

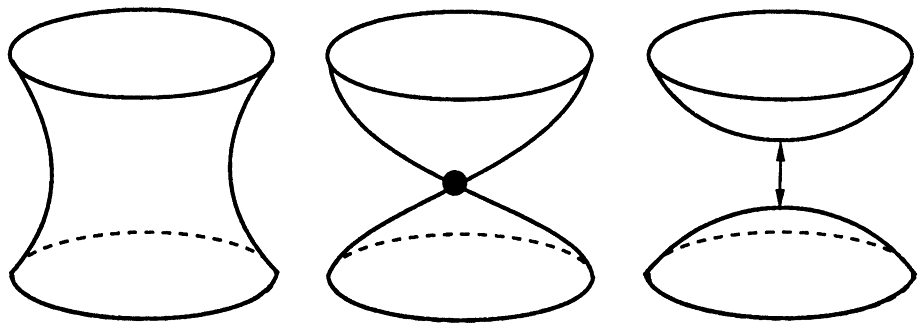


FIGURE 10 : Bifurcation over the Poles

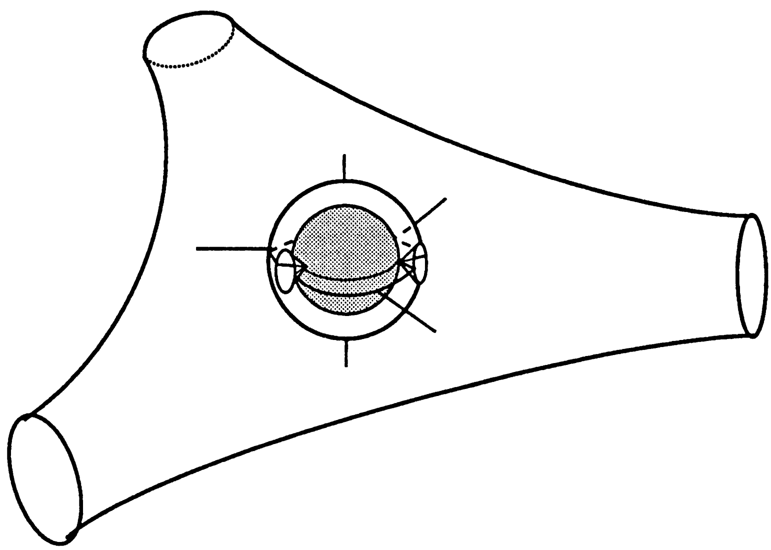


FIGURE 11 : Hill's Region after the Last Bifurcation

For all nonzero ω below the last bifurcation, there will be five elliptical Lagrangian orbits, one for each central configuration. For ω just below the critical level there may also be complicated invariant sets near the equilateral orbits. This depends on the choice of the masses; for some masses (the minority) KAM theory applies near the equilateral restpoints and there will be invariant tori, long-period orbits, etc. For most choices of the masses, however, the equilateral periodic orbits are born hyperbolic with three-dimensional stable and unstable manifolds in $M(h,\omega)$. In this case they are isolated invariant sets.

Since the dimension of $M(h,\omega)$ is so large, the KAM theory does not imply stability and this would be the place to look for Arnold diffusion in the three-body problem. If the equilateral orbits are hyperbolic it is natural to look for transverse homoclinic and heteroclinic points connecting them. We will see in section 8 that such orbits do exist for small non-zero ω, but for the nearly circular case the problem is open.

8. Low Angular Momentum. In this section we will consider the case $\omega = 0$ and the case of small non-zero ω. The interesting dynamics involves orbits which pass near the triple collision singularity. Thanks to the coordinate system of McGehee, which blows up the singularity into the invariant set $\{r = 0\}$ of 2.3, one can effectively study orbits passing near the singularity. As a result, more can be said about this case than about all the others combined. The study of triple collision began with Sundman[Su] and was carried on by Siegel [S-M]. McGehee's study of the collinear three-body [Mc3] problem led to much further work. The isosceles three-body problem, a subsystem present when two of the three masses are equal, has been studied by a number of authors [Dev1,Dev2,Si,L-L,M1,M2]. Finally [M3,M4] treat the planar case.

We will begin with the case $\omega = 0$. The Hill's region is reproduced in figure 12. Sundman's theorem about orbits near $r = 0$ no longer applies and triple collisions are possible. Quite a lot can be said about orbits which begin or end in triple collision. First of all, such orbits exist. In fact the elliptical orbits of Lagrange become more and more eccentric as $\omega \to 0$ and in the ellipses degenerate into line segments (figure 8). The limiting orbits are homothetic expansions and contractions; the shape is always the central configuration while the size increases from zero to some maximum and back to zero again. Thus these orbits are homoclinic to triple collision. In figure 12 they appear as the five line segments from the sphere $r = 0$ to the outer surface. It turns out that every triple collision orbit must approach one of the five central configurations. Some other triple collision orbits are shown in the figure. The orbits which tend to a given central configuration form a smooth submanifold of M(h,0) of dimension 2 in the collinear case and 3 in the equilateral case. The manifold of collinear collision orbits consists entirely of orbits whose configuration is collinear for all time. Such orbits would lie in the equitorial plane in figure 12. However, there are orbits ending at the equilateral collision which look nearly collinear until the last moment when the middle mass slips out from between the other two to form the required equilateral triangle. Such an orbit is shown in the figure.

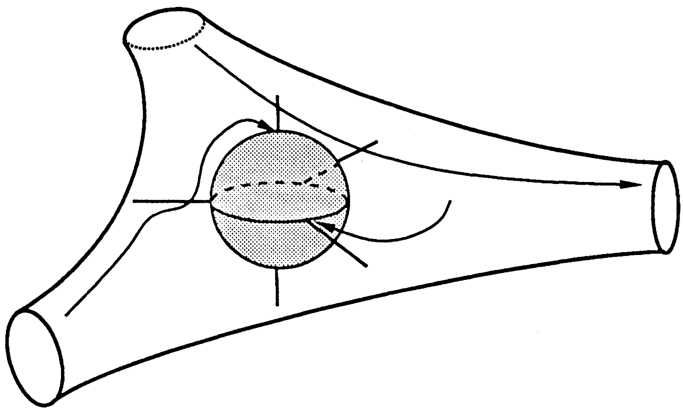


FIGURE 12 : Some Zero Angular Momentum Orbits

Among the orbits tending to triple collision in one time direction there are orbits which tend parabolically to each of the three infinities in the other time direction. By making use of these connections from infinity to triple collision one can show that there are binary exchange orbits; the exchange is carried out during a close approach to triple collision.

We already mentioned that the limiting Lagrange orbits can be viewed as orbits homoclinic to triple collision. There are infinitely many other orbits homoclinic to the two equilateral triple collisions and heteroclinic between them (for technical reasons this is only a theorem for masses in a certain open set in mass space but it probably holds for all choices of masses). Some of these pass very near to the collinear Lagrange orbits switching to the equilateral configurations only very near triple collision.

Finally we mention another kind of oscillation orbit which is known to exist in the isosceles subsystem. There are orbits which approach arbitrarily close to triple collision without actually colliding; in fact they converge to one of the collinear Lagrangian orbits. They feature infinitely many increasingly close approaches to collision between which they expand nearly homothetically like the Lagrange orbit. Such an orbit satisfies $\underline{\lim}\, r(t) = 0$ but $\overline{\lim}\, r(t) > 0$. Such an orbit is shown in figure 13 along with some of the orbits homoclinic to triple collision.

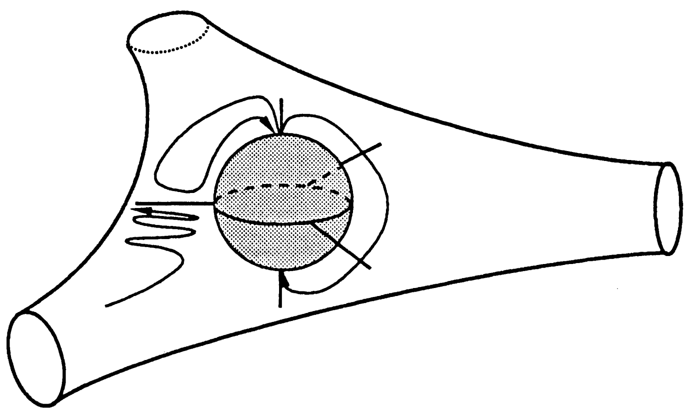


FIGURE 13 : Orbits Homoclinic to Triple Collision

The case of small non-zero angular momentum combines close approaches to triple collision with the recurrence of the highly elliptical Lagrange orbits to produce chaotic results. We will concentrate on the equilateral Lagrange orbits. When ω is sufficiently small, these orbits are hyperbolic with three-dimensional stable and unstable manifolds. These two orbits are connected by heteroclinic orbits to one another and to the three infinities (at least for masses chosen in a suitable open set as mentioned above). These are shown in figure 14.

First there are orbits running from the equilateral Lagrange orbits to parabolic infinity. Depending on the direction in which they run they are either capture orbits or escape orbits. The capture orbits, for example, evolve as follows: the particles are in a tight binary configuration, but the third mass approaches the binary, interacts closely with it and begins a very regular bounded motion which approaches an equilateral elliptical Lagrangian periodic orbit as $t \to \infty$.

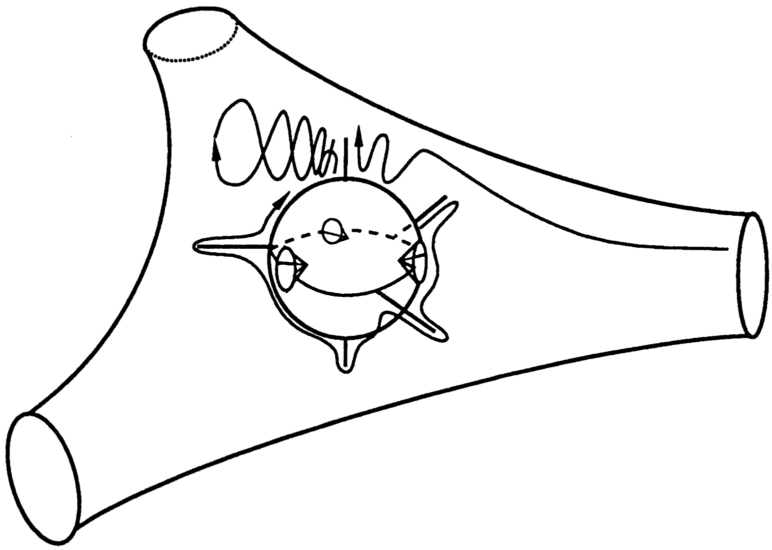


FIGURE 14 : Some Low Angular Momentum Orbits

There are transverse homoclinic and heteroclinic orbits connecting the two equilateral Lagrangian orbits. In fact, there are infinitely many distinct connecting orbits, some of which pass very near to the collinear Lagrangian orbits. The presence of homoclinic orbits produces all of the usual chaos. There are wild orbits which change shape abruptly after each close encounter with triple collision. In fact, if the masses are nearly equal, one can arrange orbits which imitate all five of the Lagrangian behaviors in turn (see figure 15); one can even arrange for such orbits to be periodic.

Roughly speaking, the reason for the existence of all of these orbits in the low angular momentum case is that the combination of the recurrence of the equilateral Lagrange orbits with the stretching and spiralling which orbits experience while passing close to triple collision produces a large invariant set describable by the methods of symbolic dynamics.

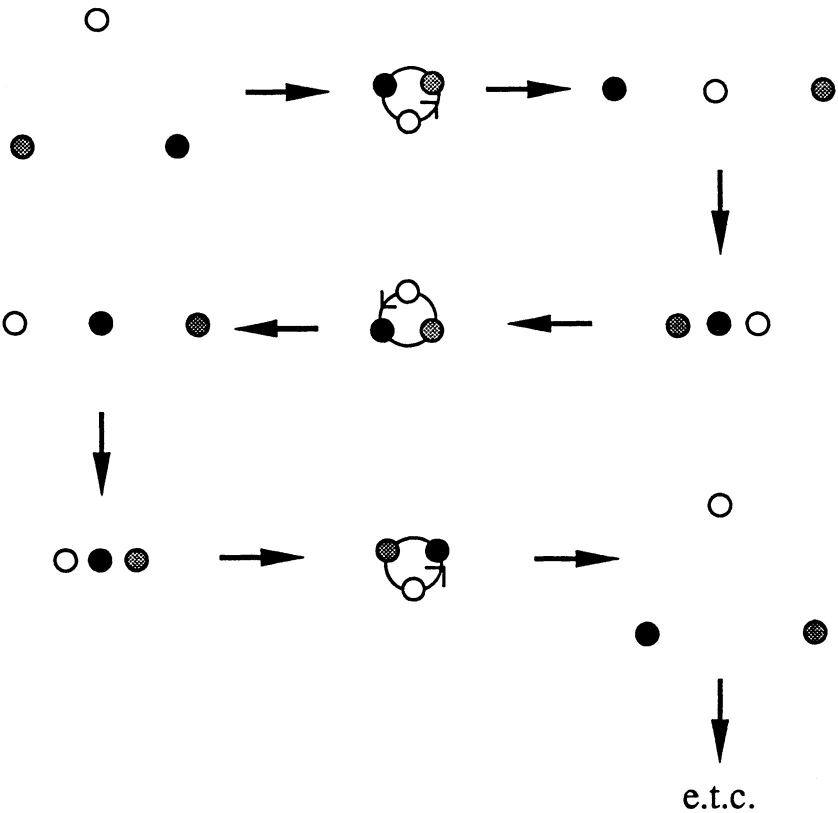


FIGURE 15 : A Low Angular Momentum Orbit

There are a number of open questions. It would be nice to incorporate the infinities into the symbolic dynamics; currently, one can get out from near triple collision to parabolic infinity but then one cannot necessarily get back for another close approach. This would enlarge the invariant set to include oscillation orbits as described in section 4. Another question concerns the collinear Lagrange orbits. These are not necessarily hyperbolic for small ω. What is going on in the neighborhood of these orbits as $\omega \to 0$? Finally, a question which could be posed for any ω is: what is happening to the angle that was quotiented out ? Do the orbits described above rotate systematically as they perform their wild changes of shape or do they emerge from the close approaches at random angles ?

9. Conclusion. We will close this tour of the three-body problem with another look at figure 1, the "big picture". There are really very few landmarks in the phase space of the three-body problem. The most important are triple collision, infinity and the periodic orbits of Lagrange. These are the elements of figure 1. A program for understanding the three-body problem is first to conduct local studies of these features and then to find out how they are connected to one another. As we have seen, a little progress has been made in the three centuries since Newton formulated the problem but a genuine understanding of what is possible remains a distant goal.

References:

[Bir] G.D.Birkhoff: Dynamical Systems, AMS Coll.Pub.,vol.9 (1927).

[C] C.Conley: On some new long periodic solutions of the plane restricted three-body problem. Comm.Pure Appl.Math. 16,449-467 (1963).

[Dev1] R.Devaney: Triple collision in the isosceles problem, Inv.Math. 60,70-88 (1980).

[Dev2] ----------: Singularities in classical mechanics, in Ergodic Theory and Dynamical Systems,Birkhauser, Boston (1981).

[Eas] R.Easton: Some topology of the three-body problem, JDE 10,371-377 (1971).

[H] G.W.Hill: Researches in the lunar theory, Am.J.Math.1,5-26,129-147,245-260 (1878).

[Lag] J.L.Lagrange: Ouvres,vol.6,272-292,Paris (1873).

[L-L] E.Lacomba and L.Losca: Triple collision in the isosceles three-body problem, Bull.AMS 3,B489-B492 (1980).

[Mc1] R.McGehee: Singularities in classical mechanics, Proc.Int.Cong.Math.,827-834, Helsinki (1978).

[Mc2] -----------: A stable manifold theorem for degenerate fixed points with applications to celestial mechanics, JDE.

[Mc3] -----------: Triple collision in the collinear three-body problem,Inv.Math.27,191-227 (1974).

[Mc-Eas] R.McGehee and R.Easton: Homoclinic phenomena for orbits doubly asymptotic to an invariant three-sphere, Ind.U.Math.J. 28,211-239 (1979)

[Mey] K.Meyer: Periodic solution of the N-body problem, JDE 39, no.1, 2-38 (1981).

[M1] R.Moeckel: Orbits of the three-body problem which pass infinitely close to triple collision, Amer.J.Math. 103,6 70-88 (1981).

[M2] ----------: Heteroclinic phenomena in the isosceles three-body problem, SIAM J.Math.Anal. 15,5 857-876 (1984).

[M3] ----------: Orbits near triple collision in the three-body problem, Ind.U.Math J. 32,2 221-240 (1983)

[M4] ----------: Chaotic dynamics near triple collision, submitted to Arch.Rat.Mech.

[Mos] J.Moser: Stable and Random Motions in Dynamical Systems,Ch.III,Annals of Math.Studies,vol.77,Princeton Univ.Press (1973).

[Mou1] F.R.Moulton: A class of periodic solutions of the problem of three bodies with applications to lunar theory,Trans.AMS 7,537-577 (1906).

[Mou2] ------------: An Introduction to Celestial Mechanics,Dover, New York (1970).

[Rob] C.Robinson: Homoclinic orbits an oscillation for the planar three-body problem, Ind.U.Math.J. 28,211-239 (1979).

[S-M] C.Siegel and J.Moser: Lectures on Celestial Mechanics,Grundlehren der Math. Wiss.187, Springer-Verlag, New York (1971).

[Si] C.Simo: Analysis of triple collision in the isosceles problem, Classical Mechanics and Dynamical Systems, Marcel Dekker, New York (1980).

[Sit] K.Sitnikov: Existence of oscillating motions for the three-body problem, Dokl.Akad.Nauk. USSR,133,no.2,303-306 (1960).

[Sm] S.Smale:Topology and mechanics, Inv.Math.10,305-331 (1970).

[Su] K.F.Sundman: Memoire sur le problem des trois corps, Acta Math. 36,105-179 (1913).

http://dx.doi.org/10.1090/conm/081/986255

Contemporary Mathematics
Volume 81, 1988

SYMMETRY IN n-PARTICLE SYSTEMS

Donald G. Saari [1]

ABSTRACT. According to Noether's theorem, symmetry in Hamiltonian systems translates into integrals of motion. Some of the methods used to extract information about the dynamics from the integrals are outlined in Section 2. In this paper a conceptually simple approach is introduced to subsume and extend many of these efforts. The basic ideas are introduced with the integrals of angular momentum for n-particle systems, and the utility of this approach is indicated with some new results. This approach extends to all symmetry integrals for systems of the general form $r'' = \nabla U(r)$.

1. ANGULAR MOMENTUM. For a n-particle system in a d-dimensional physical space, $\mathbf{r}_i \in R^d$, $i = 1, 2, .., n$, is the position vector of the i^{th} particle. Usually, $d = 2, 3$. (Treat $d = 2$ as the x-y plane in R^3.) The equations of motion are

$$(1.1) \qquad m_i \mathbf{r}_i'' = \nabla_i U(\mathbf{r}_1,..,\mathbf{r}_n) \quad i = 1, .. ,n$$

where $m_i \neq 0$, U is defined on a domain D in $(R^d)^n$, and ∇_i is the gradient with respect to $\mathbf{r}_i$. For instance, if $m_i > 0$ and $U = \Sigma_{i<j} m_i m_j / |\mathbf{r}_i - \mathbf{r}_j|$, then Eq. 1.1 is the Newtonian n-body problem where the domain requirements are that $\mathbf{r}_i \neq \mathbf{r}_j$ for $i \neq j$.

Equations (1.1) admit the energy integral

$$(1.2) \qquad T = (1/2)\, \Sigma_i\, m_i \mathbf{v}_i^2 = U + h$$

where $\mathbf{v}_i = \mathbf{r}'_i$ is the velocity of the i^{th} particle. If U, the *self-potential*, depends on the distances between particles, then the invariance of U with respect to translations admit the 2d integrals that fix the "center of mass" of the system,

$$(1.3) \qquad \Sigma_i\, m_i\, \mathbf{r}_i = \mathbf{A}t + \mathbf{B}, \quad \Sigma_i\, m_i \mathbf{v}_i = \mathbf{A},$$

where, if $m_i > 0$, the usual choice is $\mathbf{A} = \mathbf{B} = \mathbf{0}$. The integrals restrict the orbits to a linear subspace of $(R^d)^n \times (R^d)^n$ that

---

1980 Mathematics Subject Classification (1985 Revision). 58F05, 70F10.

1. Supported by a grant from the NSF.

we denote by $R^{d(n-1)} x R^{d(n-1)}$. Also, U is invariant with respect to rotations, so we find the d(d-1)/2 integrals (angular momentum)

(1.4) $\mathbf{c} = \Sigma_i m_i \mathbf{r}_i(t) \times \mathbf{v}_i(t)$.

Without loss of generality, assume that the constant of integration is $\mathbf{c} = c\mathbf{e}_3$.

The integrals reduce the system to degree 2dn-(d+1)(d+2)/2 where Eqs. 1.2, 1.4 restrict the orbits to the level sets of

(1.5) $F: R^{d(n-1)} x R^{d(n-1)} \dashrightarrow R^{1+(d(d-1)/2)}$;

$F = (T-U, \Sigma_i m_i \mathbf{r}_i x \mathbf{v}_i)$.

How can we exploit these integrals? The obvious approach is to use Eqs. 1.2, 1.4 to solve for some of the velocity terms, but the mixed, quadratic form of Eq. 1.4 complicates this analysis when n>2 and d = 3. Nevertheless, this "implicit function" approach has served us well for d = 2.

1. *Topological classification.* An insight into the orbit structure follows from the topology of the level sets of F; i.e., $M_{h,c} = F^{-1}((h,c))$. In particular, we want to characterize the bifurcations where the topology changes. For the coplanar (d=2) n-body problem, S. Smale [13] and R. Easton [2] made important contributions, and partial statements for d = 3 are given by H. Cabral [1]. These results are somewhat specific to the functional form of U. I'll indicate how to obtain "best possible" results for all d for a wide class of potentials.

2. *Restrictions on configurations.* The configurations formed by the particles are identified with the *system position vector* $\mathbf{r} = (\mathbf{r}_1,..,\mathbf{r}_n) \in R^{d(n-1)}$. Do the integrals restrict which configurations the dynamics admits? To find such restrictions on $\mathbf{r}$ one might project $M_{h,c}$ to configuration space. Using a different approach, this program was carried out for the coplanar Newtonian n-body problem by C. Marchal and D. G. Saari [4]. We used the Sundman inequality,

(1.6) $c^2 \leq 4\ IT = 4I(U + h)$,

where $2I = \Sigma_i m_i \mathbf{r}_i^2$ is the polar momentum of inertia. ($I^{1/2}$ measures the "radius" of the universe.) Consequently,

(1.7) $c^2/4I^{1/2} - hI^{1/2} \leq C(\mathbf{r}) = I^{1/2}U$.

The last term in this inequality, the *configurational measure*, is invariant with respect to scalar change (i.e., $C(\mathbf{r})$ is homogeneous of degree zero) and rigid body rotations. Thus, $C(\mathbf{r})$ is invariant with respect to elements in an Eulerian similarity class. In this manner Eq. 1.7 proves that the configurations, as measured by $C(\mathbf{r})$, are restricted by $I^{1/2}$ and constants of the system. For the important case $h<0$, the left hand side of Eq. 1.7 is bounded below. Thus, a configuration $\mathbf{r}$ is admitted iff

(1.8) $\quad c\,|h|^{1/2} \le C(\mathbf{r})$.

These restrictions on admissible configurations are "best possible" for coplanar problems. As true for the topological classification, they are not best possible if $d = 3$ because Sundman's inequality is not sharp for $d \ge 3$. I'll indicate how to obtain "best possible" results and why the above approach is equivalent to projecting $M_{h,c}$ to configuration space.

3. *Analytic results.* Analytic techniques have long been used to analyze orbits. As just some examples, recall that by a clever use of canonical changes of variables for the Newtonian three body problem, C. L. Siegal [12] showed that if the system suffers a total collapse at $\mathbf{0}$, then $\forall\ i$, $\mathbf{r}_i/I^{1/2}$ approaches a definite limit: the three particles can't collide with an "infinite spin." N. Hulkower [3] used similar techniques to discuss a related problem for the expanding three - body systems. Using different techniques, Sundman proved that if the Newtonian n-body problem has a complete collapse, then $c = 0$. (This result may have been known by Weierstrass. (See [7,p66].) Weierstrass showed for the Newtonian three body problem that if $c = 0$, then the motion is restricted to a fixed plane in R^3 for all time. Laplace and others determined restrictions on motion where the configuration doesn't change. I'll extend these results.

4. *Plane of motion.* Three particles define a plane passing through $\mathbf{0}$. A practical problem of astronomy is to understand how this "plane of motion" changes with the dynamics. The approach used here completely determines this motion in terms of the configurations formed by the particles and constants of the system.

2. DECOMPOSITION OF THE VELOCITY.

To introduce the basic idea, consider n-particle systems given by Eq. 1.1 where U admits the integrals Eqs. 1.2, 1.3, 1.4. The approach is to use the integrals to determine an orthogonal decomposition of the *system* *velocity vector* $v = (v_1, \dots, v_n)$. This decomposition is defined by the *system inner product* $\langle , \rangle : (R^d)^n x (R^d)^n \to R$, where for $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$ and the inner product $(-,-)$ on R^d,

(2.1) $\langle a, b \rangle = \Sigma_i \; m_i (a_i, b_i)$.

The system inner product is positive definite iff $m_i > 0$. If the m_i's differ in sign, $\langle -,- \rangle$ is a signed inner product with degenerate vectors. To simplify the exposition, I'll assume that $m_i > 0$, but occasionally I'll mention what differences arise when the m_i's differ in sign. The gradient defined by the system inner product admits the equations of motion

(2.2) $r'' = \nabla U(r)$,

where "∇" $= ((1/m_1)\nabla_1, \dots, (1/m_n)\nabla_n)$. Also,

(2.3) $2T = \langle v, v \rangle, \quad 2I = \langle r, r \rangle$.

For $a, b \in (R^d)^n$, let $a \times b = (a_1 x b_1, \dots, a_n x b_n)$. Let $E_i = (e_i, \dots, e_i) \in (R^d)^n$ where e_i is the unit vector in R^d with unity in the i^{th} component. With this notation, the integrals in Equation 1.3 become

(2.4) $\langle E_i, r \rangle = 0, \quad \langle E_i, v \rangle = 0$.

The angular momentum integral, Eq. 1.4, is

(2.5) $\langle E_i \times r, v \rangle = (c, e_i)$,

where $i = 3$ for $d = 2$, and $i = 1,2,3$, for $d = 3$. To see this, note that $\langle E_i \times r, v \rangle = \Sigma\, m_i (e_i x r_i, v_i) = (e_i, \Sigma m_i r_i x v_i) = (e_i, c)$.

An immediate consequence of Eq. 2.5 is that if $m_i > 0 \;\; \forall\, i$, then Sundman's inequality, Eq. 1.6, follows from Cauchy's inequality as

(2.6) $c^2 = (e_3, c)^2 = \langle E_3 x r, v \rangle^2 \le \langle E_3 x r, E_3 x r \rangle \langle v, v \rangle \le$
$\langle r, r \rangle \langle v, v \rangle$.

Thus Sundman's inequality *never* is sharp for motion out of the x-y plane because the last inequality never achieves equality. This explains the difficulty in extending coplanar conclusions derived from Sundman's result.

According to Eq. 2.5, the component of v along the line $E_i x r$ is a constant. This suggests the strategy of orthogonally decomposing the system velocity vector, v, into two parts. The first part, W_1, is the projection of v to the subspace spanned by $\{E_i x r\}$, while the second part, w_2, is what remains. To interpret the projection W_1, start with r and let $M_r = \{\Omega(r) = (\Omega(r_1), \dots, \Omega(r_n)) \mid \Omega \in SO(d)\}$. M_r, the orbit of a group action, is a smooth manifold of dimension $d(d-1)/2$ except if r is collinear and d = 3. For this singular situation, $\dim(M_r) = 2$. The tangent space

(2.7) $T_r M_r = \mathrm{span}\{E_3 x r\}$, d=2; $\mathrm{span}\{E_i x r\}$, i = 1,2,3, d = 3.

Thus, W_1 is the *component of the system velocity corresponding to a rigid body rotation*. Furthermore, W_1 is uniquely determined by the basis and Eq. 2.5, so $W_1 = \Sigma S_i E_i x r$ where $S = (S_1, S_2, S_3)$ is the instantaneous axis of rotation. This proves the first part of Theorem 1.

Theorem 1. *The rotational velocity component, W_1 is uniquely determined by r and c; conversely, W_1 and r uniquely determine c.*

Proof of the second part. $v = W_1 + w_2$ where w_2 is orthogonal to $T_r M_r$. The conclusion now follows from Eq. 2.5.

Thus, the component $W_1(r,c)$ is equivalent to the angular momentum integral. The remaining component of v, w_2, is in N_r, the normal bundle (in $R^{d(n-1)}$) to $T_r M_r$. Indeed, using Theorem 1, it follows for *any* V in N_r that $(r, W_1(r,c)+V)$ is on the integral surfaces defined by Eqs. 1.3 and 1.4. It now is easy to geometrically or topologically characterize these surfaces for all choices of d, and the characterization does not involve U(r).

More is possible; the orthogonal decomposition separates the kinetic energy, so Eq. 2.3 becomes

(2.8) $2T = \langle v,v\rangle = \langle W_1, W_1\rangle + \langle w_2, w_2\rangle = 2(U(r) + h)$.

Let $f(r,c,h) = 2(U(r) + h) - \langle W_1(c,r), W_1(c,r)\rangle$. Equation 2.8 shows that the combination of the integrals forces

(2.9) $\langle w_2, w_2\rangle = f(r,c,h)$,

which means that integrals determine $\langle w_2, w_2\rangle$. Conversely, it follows from the decomposition of v that for *any* $V \in N_r$ satisfying Eq. 2.9, (r, W_1+V) is in $F^{-1}(h,c)$. Thus we can characterize $M_{h,c}$ for any value of d. For instance, this characterization requires w_2 to be on the sphere in N_r with center 0 and radius $f^{1/2}$. If f = 0, then $w_2 = 0$. Since $\langle , \rangle$ is positive definite, no motion occurs where f<0. So, $M_{h,c}$ is a pinched sphere bundle over the space of all possible, admissible configurations $AC_{h,c} = \{r \mid f(r,c,h) \geq 0\}$, where "pinched" refers to the degenerate sphere, point 0, that occurs whenever f=0. (If the m_i's differ in sign, $\langle , \rangle$ is a signed inner product, so the spheres in the fibers are replaced by hyperbolids. Here there are no restrictions on the base space -- on the admissible configurations -- because <,> and f admit all signs. However, the form of the hyperbolid changes with sign(f), and this forces certain velocity components to dominate others.) Thus the decomposition of **v** leads to a simple method to recapture and extend the first two approaches described in Section 1.

To further develop these ideas for d = 2, let U be a positive, homogeneous function of degree α; i.e., $U(\tau\mathbf{r}) = \tau^{\alpha}U(\mathbf{r})$. An advantage of this restriction is that $C_{\alpha}(\mathbf{r}) = I^{-\alpha/2}U$ is a positive, homogeneous function of degree zero that is independent of rotations. Thus, this *configurational measure* determines which Eulerian (similarity) configurations are admissible.

Theorem 2. *Let d = 2 and let c and h be given. The set of admissible configurations is given by* $AC_{h,c} = \{\mathbf{r} \mid f(\mathbf{r},c,h) \geq 0\}$. *If* $\alpha<-2$ *or* $\alpha>0$, *then any Eulerian configuration is admitted for some value of I. If* $-2<\alpha<0$, *then an Eulerian configuration is admitted for certain values of I iff* $h < 0$ *and*

(2.10) $$C_{\alpha}(\mathbf{r}) \geq D(\alpha)\, c^{-\alpha}|h|^{1+(\alpha/2)},$$
$$D(\alpha) = 2^{\alpha}|\alpha|^{\alpha/2}/(2+\alpha)^{\alpha/2}\{1-(\alpha/(2+\alpha))\}.$$

If $\alpha = 0$, *then* $C_{\alpha} \geq -h$; *if* $\alpha = -2$, *then* $C_{\alpha} \geq c^2/4$. *If* $-2<\alpha<0$, *the topology of* $AC_{h,c}$ *changes for those values of c and h and at* $\mathbf{r}^*$ *where Eq. 2.10 is an equality and*

(2.11) $$\nabla C_{\alpha}(\mathbf{r}^*) = 0.$$

Definition. A configuration $\mathbf{r}^*$ is a *central configuration* iff $\nabla C_{\alpha}(\mathbf{r}^*) = 0$.

By using Euler's theorem for homogeneous functions, it follows that $\mathbf{r}^*$ *is a central configuration iff* $\tau\mathbf{r}^* = \nabla U(\mathbf{r}^*)$ where $\tau = \alpha U(\mathbf{r}^*)/2I$. Namely, the force vector lines up with the position vector. As we see from Eq. 2.11, one reeason central configurations are important is that they characterize the topological bifurcations. For the Newtonian 3-body problem, the central configurations are three collinear configurations and the equilateral triangle.

Proof. Because d = 2, it follows from a simple computation that $\langle \mathbf{W}_1, \mathbf{W}_1 \rangle = c^2/2I$. Because <-,-> is positive definite, Eq. 2.8 can be expressed as

(2.12) $T_a(I) = c^2/4I^{(2+a)/2} - hI^{-a/2} \leq C_a(\mathbf{r})$. [2]

This means that the configurational measure is bounded below by $T_a(I)$. If $a<-2$ or $a>0$, then Range(T_a) includes $(0,\infty)$, so there are no constraints on the configurations.[3] The same situation occurs for $-2<a<0$ if $h\geq 0$. On the other hand, if $-2<a<0$ and $h<0$, then $T_a(I) \geq D(a)c^{-a}|h|^{1+(a/2)}$ with equality when $I = (2+a)c^2/4ah$.

The boundary of $AC_{h,c}$ corresponds to a level set of $C_a(\mathbf{r})$, so the topology of the region changes at critical points of C_a. These critical points are the central configurations.

As an example, consider the Newtonian three body problem. The center of mass is fixed at 0, so $2MI = \Sigma_{j<k} m_j m_k (\mathbf{r}_j - \mathbf{r}_k)^2$ where $M = \Sigma m_i$. If we view $m_j m_k$ as "weights" and $r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|$ as variables, then $I^{1/2}$ and U^{-1} are, respectively, scalar multiples of the weighted geometric and harmonic means of the variables. According to a classical theorem, $C(\mathbf{r}) = I^{1/2}U$ has a global minimum $C(\mathbf{r}_e)$ iff all of the variables are equal; the particles form an equilateral triangle, $\mathbf{r}_e$. (This assertion holds for any U that is a weighted mean. For other choices of U, the central configurations can differ.) The three remaining central configurations are collinear, and they are distinguished by which mass is in the middle. Label these configurations as $\mathbf{r}^j$, $j = 2, 3, 4$, where $C(\mathbf{r}_e) < C(\mathbf{r}^2) \leq C(\mathbf{r}^3) \leq C(\mathbf{r}^4)$. Consequently, if $c|h|^{1/2} < C(\mathbf{r}_e)$, all configurations can occur. If $c|h|^{1/2} = C(\mathbf{r}_e)$, then a restriction occurs at equilateral configurations. Continuing, if $c|h|^{1/2}$ is between $C(\mathbf{r}_e)$ and $C(\mathbf{r}^2)$, then the configurations never resemble an equilateral triangle. Finally, restrictions are imposed on which collinear configurations can occur. These conclusions are depicted in Figure 1 where the lines indicate different boundary restrictions on configurations.

S2. Equation 2.12 generalizes Eq. 1.7, so the approach in Sec. 1 of using Sundman's inequality is equivalent to the projection of $M_{h,c}$ to configurations space. These results are for $C_a > 0$, there are related results for $C_a < 0$. Here, Eq. 2.2 imposes an *upper bound on admissible configurations.*

---

3. On the other hand, it follows immediately from Eq. 2.12 that if $c \neq 0$, there always are values of I that create constraints.

We can characterize $M_{h,c}$ because the remaining component of v, w_2, is on a sphere of radius $f^{1/2}$ in the fiber N_r. For instance, in Figure 1, if r is not on the boundary of the admissible configurations, then the distance of r from the boundary determines the value of f > 0. In turn, this determines radius of the sphere of values for w_2. Let $S^k{}_\beta$ be the k dimensional sphere in R^{k+1} with center at the origin and with radius β.

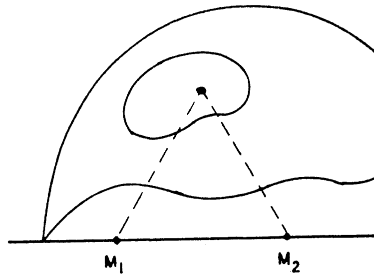


Figure 1

Theorem 3. $M_{h,c} = AC_{h,c} \times S^{2n-4}{}_\beta$ *where* $\beta^2 = f(r,c,h)$.

Differences arise when this approach is used to determine $M_{h,c}$ for d = 3. First, $\dim(N_r) = 3n-7$ if r is *not* a collinear configuration and 3n-6 if r is a collinear configuration. Thus, the sphere containing w_2 changes dimension should r pass through a collinear configuration. This is caused by the change in $\dim(T_r M_r)$ created by the singularity of the group action: rotating a collinear configuration about its axis doesn't change anything. To explain this, if r(t) defines a configuration that spans at least a d-1 dimensional space in R^d, then r can be used to describe the motion of a frame. In particular, the axis of rotation, S, is uniquely determined by r and c. On the other hand, if r is collinear for d=3, then only two components of S are determined. The third component is the added dimension in N_r, and it corresponds to a limiting change in the frame as the particles pass through a collinear configuration. A second difference is that $\langle W_1, W_1\rangle$ does not have the simple form $c^2/2I$ for d≥3, but a representation for W_1 follows from linear algebra. The third difference is in the description of A_{hc}. When d=2, as shown in Section 2, this set is characterized by I and the configurational measure $C_a(r)$. For d=3, other variables, such as the orientation of the configuration relative to the invariable *plane* (i.e., the plane orthogonal to c), play a role. We correctly might expect further constraints to be imposed on the admissible configurations. Finally, because other variables are involved, we should expect that the bifurcation points for the configurations need not be the critical points of C_a; configurations other than central configurations are involved. All of this happens. (A description of $AC_{h,c}$ for the Newtonian three body problem is in [9].)

3. NOETHER'S THEOREM. All of the results in Sections 2 and 4 are based on the orthogonal decomposition of **v** obtained by representing the angular momentum integral as $\langle E_i \times \mathbf{r}, \mathbf{v}\rangle = (\mathbf{c}, e_i)$, $i = 1,2,3$ (Eq. 2.5). The details differ, but whenever U admits integrals of the form

(3.1) $\langle \mathbf{H}(\mathbf{r}), \mathbf{v}\rangle$,

the same kind of program applies. Namely, let $\{\mathbf{H}_j(\mathbf{r})\}$ be the vectors defining the integrals given by Eq. 3.1. These vectors span an *integral subspace* of $T_r R^{d(n-1)}$. Orthogonally decompose the system velocity vector where one component is in the integral subspace. For the same reasons as given in Section 2, the component of v in the integral subspace is equivalent to the integrals, so this component is uniquely determined by **r** and the constants of integration. Because the decomposition is orthogonal, use the energy integral to determne the magnitude of the remaining component of v in terms of **r** and constants of the sytem. Thus, if the system inner product is positive (or negative) definite, this last component is on a sphere in the fiber where the radius is determined by **r** and constants of the system. (With a signed inner product, the component is on a hyperbolid.) So, to carry out this program, we need to know which integrals admit the form Eq. 3.1.

Theorem 4. *Let* $\mathbf{r}'' = \nabla U(\mathbf{r})$ *be given where* $\mathbf{r} \in R^k$ *and the gradient is defined by an inner product* $\langle -,- \rangle$. *Let G be a smooth mapping* $G: R^k \times (-\tau,\tau) \to R^k$ *satisfying:*

a. $G(\mathbf{r},0) = \mathbf{r}$ *for all* $\mathbf{r}$.

b. $U(G(\mathbf{r},s)) = U(\mathbf{r})$ *for* $s \in (-\tau,\tau)$.

c. *If DG is the Frechet derivative with respect to the first variable, then* $\langle v,v\rangle = \langle DG(r,s)(v), DG(r,s)(v)\rangle$ *for* $s \in (-\tau,\tau)$.

The system has the integral $\langle (d/ds)G(\mathbf{r},s), \mathbf{v}\rangle|_{s=0} = \mathbf{b}$.

In other words, Theorem 4 asserts that all of the Noether symmetry integrals can be expressed as Eq. 3.1, so an orthogonal decomposition can be used to re-express these integrals as components of v. (The manifold $M_{\mathbf{r}}$ is replaced with the orbit of G.) As an example, if U depends on the distances between particles, then $G(\mathbf{r},s) = \mathbf{r} + sE_i$ satisfies the three conditions. This means that $\langle (d/ds)G(\mathbf{r},s), \mathbf{v}\rangle|_{s=0} = \langle E_i, \mathbf{v}\rangle$ is constant on orbits; this is Eq. 2.4. A second choice is $G(\mathbf{r},s) = \Omega(s)(\mathbf{r}) = (\Omega(s)\mathbf{r}_1, \ldots, \Omega(s)\mathbf{r}_n)$ where

$$\Omega(s) = \begin{bmatrix} \cos(s) & \sin(s) & 0 \\ -\sin(s) & \cos(s) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

By matrix calculus, it follows that $\langle (d/ds)G(\mathbf{r},s), \mathbf{v}\rangle|_{s=0} = \langle \mathbf{E}_3 x \mathbf{r}, \mathbf{v}\rangle = c$; this is Eq. 2.5. Many earth satellite models are of the form $U = U(x^2+y^2, z)$. Since U is invariant with respect to $\Omega(s)$, the integral is $(e_3 x \mathbf{r}, \mathbf{v}) = c$. As another example suggesting the dynamics of the orientation of a satellite, consider $U(x,y,z,u,v,w) = U((x^2+y^2+z^2), (u^2+v^2+w^2))$. U is invariant with respect to rotations of the x,y,z variables as well as of the u,v,w variables, and all six integrals have an expression ready for an orthogonal decomposition of the system velocity (x',y',z',u',v',w').

Proof of Theorem 4. To show that $(d/dt)\langle (d/ds)G(\mathbf{r},s), \mathbf{v}\rangle|_{s=0} = 0$, note that $(d/dt)\langle (d/ds)G(\mathbf{r},s), \mathbf{v}\rangle = \langle (d^2/dtds)G(\mathbf{r},s), \mathbf{v}\rangle + \langle (d/ds)G(\mathbf{r},s), \mathbf{r}''\rangle$. The last term at s=0 is $\langle (d/ds)G(\mathbf{r},s), \nabla U\rangle|_{s=0} = (d/ds)U(G(\mathbf{r},s))|_{s=0}$. According to (b), this equals $(d/ds)U(\mathbf{r}) = 0$. Thus, it remains to show that $\langle (d^2/dtds)G(\mathbf{r},s), \mathbf{v}\rangle|_{s=0} = 0$. Using the smoothness of G, interchange the order of differentiation to obtain $\langle (d/ds)DG(\mathbf{r},s)(\mathbf{v}), \mathbf{v}\rangle|_{s=0}$. By differentiating the isometry condition (c), $0 = (d/ds)\langle \mathbf{v},\mathbf{v}\rangle = (d/ds)\langle DG(\mathbf{r},s)(\mathbf{v}), DG(\mathbf{r},s)(\mathbf{v})\rangle = 2\langle (d/ds)DG(\mathbf{r},s)(\mathbf{v}), DG(\mathbf{r},s)(\mathbf{v})\rangle$. But by (a), $DG(-,0)$ is the identity mapping. Thus $0 = \langle (d/ds)DG(\mathbf{r},s)(\mathbf{v}), DG(\mathbf{r},s)(\mathbf{v})\rangle|_{s=0} = \langle (d/ds)DG(\mathbf{r},s)(\mathbf{v}), \mathbf{v}\rangle|_{s=0}$. This completes the proof.

4. SOME ANALYTIC CONSEQUENCES AND THE PLANE OF MOTION. In this section, I'll continue to exploit the decomposition of **v** based on the angular momentum by improving upon the analytic results described in Section 1. The assumptions are that d = 3, that U admits the angular momentum integrals, and that U is homogeneous. Because U is homogeneous, radial changes play a distinguished role, so decompose $\mathbf{w}_2 = \mathbf{W}_2 + \mathbf{W}_3$ where $\mathbf{W}_2$ is the projection of $\mathbf{w}_2$ in the radial direction $\mathbf{r}/r$, and $\mathbf{W}_3$ is what remains. It is easy to see that $\mathbf{W}_3$ is the velocity component corresponding to configurational changes while $\mathbf{W}_2 = r'\mathbf{r}/r = I'\mathbf{r}/2I$ changes the scale within a Eulerian similarity class. Again, by orthogonality we have

(4.1) $\langle \mathbf{W}_1, \mathbf{W}_1\rangle + (I')^2/2I + \langle \mathbf{W}_3, \mathbf{W}_3\rangle = 2(U + h).$

If $\langle -,- \rangle$ is positive definite, then $\langle \mathbf{W}_1, \mathbf{W}_1\rangle$ is bounded below by $c^2/2I$. Thus, a weaker version of Eq. 4.1 is

(4.2) $c^2 + (I')^2 \le 4IT = 4I(U + h).$

Thus *Sundman's inequality holds for all n-particle systems where the m_i's are positive, and it is sharp only if $d \le 2$.* Moreover, it follows that Sundman's inequality should be viewed as an approximation of the orthogonal decomposition of the system velocity vector.

Inequality 4.2 can be used to improve upon the conclusions in Section 2. To do so, we need the next technical lemma.

Lemma. Suppose that U is homogeneous of degree α. Then,

(4.3) $I'' = (2 + \alpha)U + 2h = (2 + \alpha)T - \alpha h.$

Let $S_\alpha(\mathbf{r}) = (c^2 + (I')^2)/4I^{(2+\alpha)/2} - hI^{-\alpha/2}.$

(4.4) $S_\alpha' = (2+\alpha)I'[4IT - \{c^2 + (I')^2\}]/8I^{(4+\alpha)/2}$. If $m_i > 0\ \forall\ i$, then

(4.5) $T_\alpha(I(t)) \le S_\alpha(t) = S_\alpha(\mathbf{r}(t)) \le C_\alpha(\mathbf{r}(t)).$

Proof. By definition, $2I = \langle \mathbf{r},\mathbf{r}\rangle$, so $I'' = \langle \mathbf{v},\mathbf{v}\rangle + \langle \mathbf{r},\mathbf{r}''\rangle = \langle \mathbf{v},\mathbf{v}\rangle + \alpha U$. The last equality follows after replacing $\mathbf{r}''$ with ∇U and using Euler's theorem for homogeneous functions. Eq. 4.3 follows from the energy integral and Eq. 2.3. Equation 4.5 is a direct consequence of Eq. 4.2 and the definition of $C_\alpha(\mathbf{r})$. Equation 4.4 follows by differentiation and Eq. 4.3.

We now show how the dynamics imposes new restrictions on the admissible configurations and on how the system can expand. Because $m_i > 0$, the expression in the square brackets of Eq. 4.4 is Sundman's inequality, so it is bounded below by $\mathbf{W}_3{}^2$. Indeed, as we've shown, this inequality is zero iff all particles are in the x-y plane and $\mathbf{W}_3 = 0$. Thus, if $\alpha > -2$, the usual situation is for the signs of S_α' and of I' to agree; if $\alpha < -2$, they disagree.

This means that the growth properties of S_α and I are related and that (from Eq. 4.5) the *aggregate growth behavior of the system imposes added restrictions on which configurations can occur.* For instance, if $\alpha>-2$ and if t_o is a local minimum point for I, then until I reaches its next local maximum, the constraints on configurations are $T_\alpha(r(t)) < S(r(t_o)) \le S_\alpha(t) \le C_\alpha(r(t))$; these are more severe than those imposed by T_α in Eq. 2.10.

To illustrate the consequences of this result, recall from Section 2 that there are values of h and c where the Newtonian three body problem admits an equilateral triangle configuration, r_e. Suppose a local minimum value for I, I_m, satisfies $T_{-1}(I_m) > C_{-1}(r_e)$. The results of Section 2 do not disqualify a scenario where, as the system grows, the particles eventually form an equilateral configuration. We now know from the lemma that such a scenario cannot occur; as the system expands to the next local maximum of I, the configurations must satisfy $C_{-1}(r_e)<T_{-1}(I_m)\le C_{-1}(r(t)$. Thus, if this system ever does form an equilateral configuration, it occurs after the next local maximum value for I.

Somewhat surprisingly, these inequalities from the lemma not only relate growth behavior to configurations, but also they *bound possible growth behavior of the system.* This is clear if $U>0$, so that $C_\alpha>0$.

Theorem 5. *Let $U>0$ be homogeneous of degree $\alpha > -2$ and $c \neq 0$. If $I'(t_o)< 0$, then $S_\alpha(t_0)$ and c determine a positive lower bound for the next local minimum value of I. If $-2<\alpha<0$ and $I'(t_o)>0$, then the next local maximum value of I is bounded below by a value determined by $S_\alpha(t_0)$. All local maxima for I are bounded below by $(2+\alpha)c^2/4|\alpha h|$. If $h \ge 0$ or if $\alpha>0$, then $I'(t_o)>0$ implies that $I \to \infty$.*

To include $\alpha \le -2$, replace $c \neq 0$ with $h<0$, and $I'(t_o)\ge 0$; the conclusion is that the first local maximum of I is bounded above.

Proof. Because $I'(t)<0$ for $t>t_o$, we have that $S'\le 0$, so $S(t) \le S(t_o)$. But $T_\alpha \le S_\alpha$ where equality occurs only at critical points of I. Therefore, at least until the first local minimum of I,

$$T_\alpha(r(t)) \le S_\alpha(r(t) \le S_\alpha(r(t_o)). \tag{4.6}$$

From the representation of the graph of $T_a(r)$ in Figure 2 (for $h<0$), it is obvious that $S_a(r(t_o))$ determines a lower bound for the next minimum of $r(t)$. Conversely, if $I'>0$, then the inequality $S_a(t)\geq S_a(t_o)$ holds at least until after the next local maximum point t_1 where $I'(t_1)=0$. At this maximum value, $S_a(t_1) = T_a(t_1)\geq S_a(t_0)$.

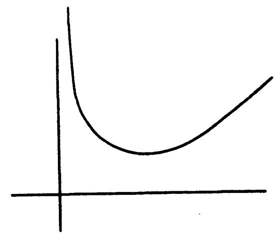


Figure 2

Thus, $I(t_1) \geq x$ where x is the larger of the two solutions to $T_a(x) = S_a(t_o)$. This value of x is bounded below by the minimum point for T_a, which is $(2+a)c^2/4|ah|$. The equation $T_a(x) = S_a(t_o)$ has only one solution if $h \geq 0$ or if $a>0$, so here $I' \neq 0$. Indeed, choose any point where $S_a(r)>T_a(I)$; the difference between these values determines a positive lower bound for $I'/I^{(2+a)/4}$ for all subsequent t. Because I is increasing, this imposes a positive lower bound on I', so $I \to \infty$.

Corollary 5.1. *Let U be a positive, homogeneous function of degree $a > -2$. Let t_o be a local minimum point for I. Until the next local maximum point for I, the configurations satisfy the constraint $T_a(I(t_o)) \leq C_a(r(t))$.*

The next corollary extends the important Weierstrass-Sundman theorem about complete collapse to other n-particle systems. An interesting feature is the simplicity of this new proof.

Corollary 5.2. *If $U > 0$ is homogeneous of degree $a>-2$, and if $c\neq 0$, then $I \not\to 0$.*

This corollary does not hold for $a \leq -2$. For instance, if $a = -2$, then, according to Eq. 4.3, $I'' = 2h$. Thus if $h < 0$, the orbit must satisfy $I = ht^2 + I'(0)t + I(0)$; i.e., if the solution lasts long enough, $I \to 0$.

Proof. Let $E = (2+a)c^2/4|ah|$ for $h \neq 0$ and 1 if $h = 0$. If $I \to 0$, then $\forall$ t after some t_o, $I(t) < E$. Because $I\to 0$, there is $t_1 > t_o$ where $I'(t_1) < 0$. According to Theorem 5, $S_a(t_1)$ determine a positive lower bound for I that serves at least until after I has its next local maximum value. According to Theorem 5, the next local maximum value for I is greater than E. This contradicts the requirement that $I(t)<E$, and the proof is completed.

It turns out that if $c \neq 0$ and if something resembling a complete collapse occurs, then $h<0$ and the motion involves wild oscillations where $\limsup I = \infty$ while $\liminf I = 0$ as $t \to t_o$. (This can't happen for the Newtonian n-body problem; if $c \neq 0$, then once r becomes sufficiently small, the subsequent motion requires $r \to \infty$. See [4].) This assertion is immediate from the following stronger statement that an upper bound for I defines a positive lower bound for I. Namely, if the radius of the universe, r, is bounded above, then r must be bounded away from zero.

Corollary 5.3. *Assume $U > 0$ is homogeneous of degree a, $c \neq 0$, and the motion exists for all time. If $a>0$ or if $-2<a<0$ and $h\geq 0$, then $I(t)$ has a unique, positive minimum and $I \to \infty$ as $|t| \to \infty$. If $h<0$ and $-2<a<0$, then all local maxima values for I are bounded below by $(2+a)c^2/4ah$. If $I < D < \infty$ for all t, then I has a positive lower bound determined by D and c. If $a \leq -2$, $h<0$, and $\liminf I > 0$ as $t \to \infty$, then $\limsup I < \infty$.*

Proof. We only need to show that D determines a lower bound for I. At a local maximum, $I' = 0$, so $S_a = T_a$. Of the two values satisfying $T_a(x) = S_a$, the smaller one serves as a lower bound for the next local minimum of $I(t)$. It follows from the graph of T_a that this lower bound for $I(t)$ is bounded below by the smaller of the two values $T_a(x) = T_a(D)$.

I've required U to be positive, but related statements hold for negative U, such as a "near-neighbor" interaction where $U = -\Sigma_k (\mathbf{r}_k - \mathbf{r}_{k+1} - \mathbf{r}_{k-1})^2$, or for n-body systems where "large" charges are added to the masses. Here $C_a(\mathbf{r}) < 0$, so Eq. 4.5 assumes the form $S_a(t) + |C_a(\mathbf{r}(t))| \leq 0$. The difference in the analysis is that $S_a(t) < 0$, (and $h \geq 0$), so rather than the configurational measure being above the curve $S_a(t)$, it is caught between the curves $y = S(t)$ and $y = 0$. The next theorem suggests the kind of possible results.

Theorem 6. *Let $U < 0$ be homogeneous of degree $a > -2$ and $c \neq 0$. Then $h > 0$ and $I \geq x$ where $T_a(x) = 0$. If $I'(t_o) > 0$, then the configurations formed at time t satisfy the inequality $|S_a(t_o)| \geq |C_a(\mathbf{r}(t))|$. If $I'(t_o) < 0$, then the next minimum value of I is bounded below by the smallest zero of $g(x) = T_a(x) - S_a(t_o)$.*

Complete collapse can occur if c = 0, but what is the motion? For instance, Painleve' [5] wondered whether the collapsing particle for the Newtonian three body problem could have an infinite spin. The answer involves two velocity components. The first and original question is whether $\mathbf{W}_1$ admits an a situation where the particles approach a fixed configuration that is spinning infinitely fast. By using a clever, complicated series of canonical transformations, this issue was resolved for the Newtonian three body problem by C. L. Siegal [12]. With different techniques, it was solved for collapsing n-body systems by Saari and Hulkower [10], and for all collisions by Saari [11]. The second kind of "infinite spin" is where the particles of the collapsing system do *not* approach any fixed configuration, so it involves properties of $\mathbf{W}_3$. (See [10,11] for details concerning the Newtonian n-body problem; here, the solution of the second kind of spin involves the properties of a certain central manifold.) In keeping with my emphasis on symmetry, I'll only consider the first kind of "infinite spin".

Theorem 7. *Let $U>0$ be homogeneous of degree $a > -2$. Suppose that $I \dashrightarrow 0$ as $t \dashrightarrow t_o$. If $\mathbf{r}(t)/I^{1/2}(t)$ approaches a fixed Eulerian similarity class as $t \dashrightarrow t_o$, then $\mathbf{r}/I^{1/2}$ approaches a definite limit as $t \dashrightarrow t_o$.*

Proof. For complete collapse, c = 0, so $\mathbf{W}_1 = \mathbf{0}$. Thus, the system has no rotational velocity terms. This completes the proof.

We now turn to the fourth topic in the introductory section. A *flat solution* is a where at each instant of time, all n particles are on a plane in R^3 passing through $\mathbf{0}$. (All solutions for n=3 are flat solutions.) Call this the "plane of motion." How does the plane of motion moves? To obtain intuition, note that a change of the orientation of the plane just rotates the configuration. Thus we should expect changes in the plane of motion to be governed by the rotational velocity $\mathbf{W}_1 = \mathbf{W}_1(\mathbf{r},\mathbf{c})$. Consequently, the dynamics of the plane should be completely determined by $\mathbf{r}$ and constants of the system.

To make this precise, let $\mathbf{n}(t)$ be the unit vector on the intersection (the "line of nodes") of the plane of motion and the invariable plane where $\mathbf{e}_3 \times \mathbf{n}$ is a vector below the plane of motion. Let $\boldsymbol{\delta}$ be the unit vector in the plane of motion that projects onto $\mathbf{e}_3 \times \mathbf{n}$. ($\boldsymbol{\delta}$ is well defined unless the plane of motion coincides with the invariable plane or the plane of motion passes through $\mathbf{e}_3$.) Express each $\mathbf{r}_i = a_i(t)\mathbf{n} + b_i(t)\boldsymbol{\delta}$. By differentiation, we have
$\mathbf{v}_i = \boldsymbol{v}_i + \{(q(t)\mathbf{e}_3 + p(t)\mathbf{n})x\mathbf{r}_i\} + s(t)((\mathbf{n}x\boldsymbol{\delta})x\mathbf{r}_i)$ where $q(t)$ is given by $\mathbf{n}' = q(t)\mathbf{e}_3 x\mathbf{n}$, $i(t)$ is the inclination of the plane of motion (the angle between the planes), $p(t) = i'$, $s(t)$ is such that $\Sigma m_i \; x \; \{a_i'\mathbf{n} + b_i'\boldsymbol{\delta}\} - s(t)(\mathbf{n}x\boldsymbol{\delta})x\mathbf{r}_i\} = \mathbf{0}$, and $\boldsymbol{v}_i = a_i'\mathbf{n} + b_i'\boldsymbol{\delta} - s(t)(\mathbf{n}x\boldsymbol{\delta})x\mathbf{r}_i$. In other words, the vectors $\boldsymbol{v}_i$ are in the plane of motion and consist of the components of $\mathbf{W}_2$ and $\mathbf{W}_3$ while $\mathbf{v} - (\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n)$ is determined by $\mathbf{W}_1$. Here, $q(t)$ measures the rate of change of the line of nodes, $p(t)$ measures the change in the inclination of the plane of motion, and $s(t)$ measures the rate of rigid body rotation within the plane of motion. Because $\mathbf{W}_1 = \mathbf{W}_1(\mathbf{r},\mathbf{c})$, we should expect q, p, and s to be determined by $\mathbf{r}$ and $\mathbf{c}$. Let $A = (a_1,\ldots,a_n)$, $B = (b_1,\ldots,b_n)$, $\langle A,B\rangle = \Sigma m_i a_i b_i$, and $A^2 = \langle A,A\rangle$.

Theorem 8.[8] *Suppose $c \neq 0$ for a flat solution of Eq. 2.2. Then, if $E = A^2B^2 - \langle A,B\rangle^2$.*

$$q(t) = cB^2/E, \quad p(t) = c\sin(i)\langle A,B\rangle/E, \quad \text{and} \tag{4.7}$$
$$s(t) = -c\cos(i)\{(B^2)^2 + \langle A,B\rangle^2\}/\{A^2 + B^2\}E$$

It is often stated that without a flat solution there isn't a natural frame to define the rotation of the n-particle system. This isn't so; at $t = 0$, let $(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ be an orthonormal basis. A natural rotating coordinate system is

$$\mathbf{k}_i' = Sx\mathbf{k}_i, \tag{4.8}$$

where S is the axis of rotation. Recall, if $\mathbf{r}$ is not collinear, then S is uniquely determined; if $\mathbf{r}$ is collinear, then the one missing component is determined by continuity. Equation 4.8 is a natural extension of the rotation of the plane of motion. In this manner, the same kind of variables can be defined for a non-flat solution with a similar relationship for the movement of the key variables. (See [8] for the equations.)

Some interesting conclusions follow from Theorem 8. For instance, the only way the plane of motion for a flat solution can attain a local maximum or minimum inclination is if the particles form a certain kind of configuration with respect to the line of nodes.

Corollary 8.1. *For a flat solution, assume that $c \neq 0$.*

a. A critical point for the inclination, $p(t) = i'=0$, occurs iff either $i = 0$, or $\langle A,B\rangle = 0$. The inclination is increasing iff $\langle A,B\rangle > 0$.

b. The system has no rotation in the plane of motion ($s = 0$) iff $i = \pi/2$.

c. If the initial configuration is not in the invariable plane, then $i>0$ for all time the solution exists.

Details of the proof are in [8];, however, the conclusions follow fairly directly from Theorem 8. A slight complication comes from the "Cauchy inequality" term, $E = A^2B^2 - \langle A,B\rangle^2$, in the denominators. This term is zero iff **A** is a scalar multiple of **B** iff the configuration is collinear. However, for $\mathbf{c} \neq 0$ and a collinear configuration, the line is in the invariable plane; i.e., the x-y plane. (If **r** is collinear, then for some **d**, each $\mathbf{r}_j = \tau_j\mathbf{d}$. Thus $\mathbf{c} = \Sigma\tau_j\mathbf{d} \times \mathbf{v}_j$ and $(\mathbf{d},\mathbf{c}) = 0$.) Collinear configurations arise naturally for flat solutions.

Theorem 9. *If $c \neq 0$, and if at $t = t_o$ the particles form a collinear configuration, then the particles are in the invariable plane. Let U be a smooth function of the distances between particles. Consider flat solutions that are not in the invariable plane for all time. Whenever all n particles pass through the invariable plane at the same time, they form a collinear configuration.*

It is easy to show that Theorem 9 need not be true for $n \geq 4$ unless the solution is flat. Theorem 9 sheds light on the topological characterization when physical space is R^3. Recall that if $\mathbf{r}$ defines a collinear configuration, then $M_{\mathbf{r}}$ is two dimensional. We see from Theorem 9 that such a configuration occurs only in the invariable plane when all n particles pass through this plane. So, the missing variable caused by the singularity of the rotation group action corresponds to a limiting value of the missing component for $\mathbf{S}$ when the configuration becomes collinear.

Proof. Because U is a function of the mutual distances, $\mathbf{r}'' = \nabla U$ is well defined in the x-y plane. Suppose a flat solution is not always in the x-y plane, but at t_0 all n particles are all in the x-y plane and the configuration is not collinear. This forces the plane of motion to coincide with the x-y plane. The solution is flat, so all components of $\mathbf{W}_3$ are in this plane. As $\mathbf{W}_2 = r'\mathbf{r}/r$, these velocity components also are in the x-y plane. The remaining component is $\mathbf{W}_1$. Clearly, for this configuration and c, there is a choice of $\mathbf{W}_1$ with all of its components in the x-y plane; just let $\mathbf{W}_1$ be the appropriate multiple of $\mathbf{E}_3 x \mathbf{r}$. For this choice, all of the velocity and position vectors are in the x-y plane, so the orbit must be in the x-y plane for all time. But, $\mathbf{W}_1$ and S are uniquely determined by $\mathbf{r}$ and $\mathbf{c}$ if $\mathbf{r}$ is not collinear. Thus, this choice of $\mathbf{W}_1$ is unique. According to uniqueness of solutions, this contradicts the assumption that the solution is not restricted to the x-y plane. (It is the choice of the missing component of $\mathbf{S}$ that determines whether the motion for the starting collinear configuration is in the invariable plane.)

Another singularity situation for the characterization of $M_{\mathbf{h},\mathbf{c}}$ is if $\mathbf{c} = \mathbf{0}$. This motivates the next theorem which generalizes a classical result for the three body problem.

Theorem 10. *Let U be a function of the mutual distances between particles. If a flat solution of the n-body problem has* $\mathbf{c} = \mathbf{0}$, *then the plane of motion is fixed for all time.*

This means that the surface $M_{\mathbf{h},\mathbf{c}}$ is given by the product of this surface for a coplanar problem and the various, possible orientations of this plane. For N>3 and non flat solutions, a similar statement holds. This is because of Eq. 4.8. If c = 0, then S = 0, so the frame remains fixed.

I'll conclude by introducing a result to suggests what happens in those singular situations where $\mathbf{W}_i \equiv 0$; i.e., where there is no rotation, or I is a constant, or the configuration doesn't change similarity classes. I'll consider $\mathbf{W}_3 \equiv 0$; the homographic solution can rotate and change size, but not shape.

Theorem 11. *Let U be homogeneous of degree a and invariant with respect to rigid body rotations of the configuration. If d = 2 and $W_3 \equiv 0$, then the homographic solution is a central configuration.*

Without added restrictions on U, Theorem 11 is false for d = 3. (See [14,pp. 292-295].)

Proof. In the natural fashion determined by the definitions of $\mathbf{W}_j$, decompose velocity space into the component subspaces V_j, j=1,2,3. According to the assumption, $\mathbf{v} = \mathbf{S}x\mathbf{r} + \mathbf{W}_2$, so $\mathbf{v}' = [\mathbf{S}'x\mathbf{r} + (2r'/r)\mathbf{S}x\mathbf{r}] + r''(\mathbf{r}/r) + \mathbf{S}x(\mathbf{S}x\mathbf{r})$. Because d = 2, the last term is $-|\mathbf{S}|^2\mathbf{r}$, so the last two terms are in V_2. The bracketed term is in V_1, so it follows that $\mathbf{r}'' = \mathbf{v}'$ has no components in V_3. The rotation invariance assumption on U forces ∇U to have no components in V_1, so $\nabla U = \xi\mathbf{r} + \tau$, where $\tau \in V_3$. But, $\mathbf{v}' = \nabla U$, so $\tau \equiv \mathbf{0}$ and $\nabla U \equiv \xi\mathbf{r}$ -- $\mathbf{r}$ is a central configuration. Incidentally, the bracketed term also must be zero, so the homographic motion is determined from the solution $\{\mathbf{S}xr^2\mathbf{r}\} = \mathbf{d}$.

References 1. Cabral, H., "On the integral manifolds of the N-body problem," Invent. Math. **20** (1973), 59-72.

2. Easton, R., Some topology of the three body problem, Jour. Diff Eq. **10** (1971), 371-377.

3. Hulkower, N., "The zero energy three body problem," Indiana Univ. Math. J. **27** (1978), 409-448.

4. Marchal,C. and D. G. Saari, Hill regions for the general three abody problem, Celestial Mechanics **19** (1975), 115-129.

5. Painleve, P., "Lecons sur la theorie analytique des equations differentielles," Hermann, Paris, 1897.

6. Pollard, H., "Gravitational systems," Jour Math. Mech. **17** (1967), 601-612.

7. Pollard, H., *Celestial Mechanics*, Carus Math Monograph #18, MAA, 1976.

8. Saari, D. G., "From rotations and inclinations to zero configurational velocity surfaces, I. A natural rotating coordinate system," Celestial Mechanics, **33** (1984), 299-318.

9. Saari, D. G., "From rotations and inclinations to zero configurational velocity surfaces, II. The best possible configurational velocity surfaces," Celestial Mechanics **40** (1987), 197-223.

10. Saari, D. G., and N. Hulkower, "On the manifolds of total collapse orbits and of completely parabolic orbits for the n-body problem," Jour. Diff. Eq. **41** (1981), 27-43.

11. Saari, D. G., "The manifold structure for collision and for hyperbolic-parabolic orbits in the n-body problem," Jour. Diff Eq. **55** (1984), 300-329.

12. Siegal, C. L., "Der Dreierstoss", Ann. Math. **42** (1941), 127-168.

13. Smale, S., "Topology and mechanics, II. The planar n-body problem," Invent. Math. **11** (1970), 45-64.

14. Wintner, A. *The Analytic Foundations of Celestial Mechanics*, Princeton University Press, 1941.

Department of Mathematics
Northwestern University
Evanston, Illinois 60208.

http://dx.doi.org/10.1090/conm/081/986256

Contemporary Mathematics
Volume **81**, 1988

The Charged Isosceles 3-Body Problem

Pau Atela
Department of Mathematics
Boston University
Boston, Mass. 02215

O. Introduction

The goal of this paper is to study the behavior near collisions of the following physical system. Within the setting of the isosceles 3-body problem (3 point masses in the plane move subject to gravitational attraction. One of them, with mass m is given an initial velocity along a vertical axis. The other two, with same mass M, are given initial positions and velocities symmetric with respect to the vertical axis. Due to these initial conditions, the configuration of the three particles is at all times an isosceles triangle) consider the effect of adding charges to the bodies, same charge e to the symmetric ones, producing a repulsion force between them, and a charge f of opposite sign to the third one.

Rescaling time we group parameters into $Q = \frac{e^2 - GM^2}{4(GMm - ef)}$ and we have a two parameter family (Q and m) of Hamiltonian equations describing the movement. Q essentially measures the difference between electrical and gravitational forces acting on the symmetric bodies and m is the mass of the third particle.

For $Q = -1$, the system is the classic isosceles 3-body problem with possible double and triple collisions. For $Q = 0$, when there is no interaction between the symmetric bodies, the equations of the system are exactly those of the anisotropic Kepler problem studied by Gutzwiller and Devaney, this giving a different physical interpretation to it. The system has now only one singularity, corresponding to triple collision. Double collisions need not be regularized, the two symmetric bodies pass through each other naturally.

For $Q > 0$ these two symmetric particles repel each other. The set of singularities corresponding to double collisions is unaccessible for finite energies. Only triple collisions can occur. We focus on this case and we use geometric techniques to gain a picture of the local behavior of solutions near collisions where a remarkable bifurcation occurs within a family of periodic orbits. By the McGehee transformation, we blow up the singularity replacing it with an invariant boundary. This is the collision manifold to which we can naturally extend the flow. Study of this flow gives us information about what happens near collision. For $Q = 0$ (A.K.P.) the collision manifold is a 2-torus, for $Q > 0$ it pinches into 2-spheres (symmetric bodies can no longer pass through each other), this being in a sense the simplest collision manifold. We search for symmetric periodic orbits near collision. As the mass m of the third particle crosses a certain value m_0, an ∞-furcation occurs. For $m > m_0$ there are only a finite number of such periodic orbits. Decreasing m perhaps new ones are born one by one spitted out from the collision-ejection orbit. For $m = m_0$ we reach a threshhold still having only a finite number of them. For $m < m_0$ we suddenly have an infinite number of them coming off the collision-ejection orbit.

Consider the following physical setting. We have three point masses in the plane with positions $q_i \in \mathbb{R}^2$ $i = 1, 2, 3$ and masses m for q_3 and M for q_1 and q_2. Their movement is subject to gravitational interaction and we will suppose that they are given charges as well. Let e be the charge of q_1 and q_2 and f the charge of q_3. We have then, additional forces to consider and we will restrict our study to isosceles configurations. That is, the third particle is given an initial velocity with a vertical direction and the two other symmetric particles are given initial positions and velocities symmetric with respect to this axis (Fig. 1).

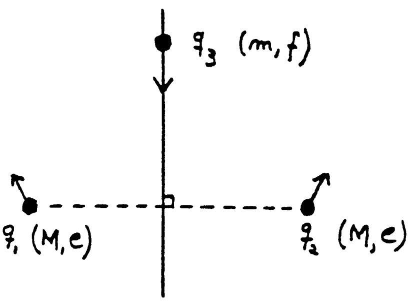


Fig. 1.

With any of these initial conditions, by symmetry, the configuration will remain an isosceles triangle for all time with the third particle moving along the vertical axis. We are interested in what happens near collisions.

From Newton's law we have that the equations of motion on the plane are given by

$$\begin{aligned}
Mq_1'' &= \frac{-GM^2 + e^2}{|q_1 - q_2|^3}(q_1 - q_2) + \frac{-GMm + ef}{|q_1 - q_3|^3}(q_1 - q_3) \\
Mq_2'' &= \frac{-GM^2 + e^2}{|q_2 - q_1|^3}(q_2 - q_1) + \frac{-GMm + ef}{|q_2 - q_3|^3}(q_2 - q_3) \\
mq_3'' &= \frac{-GMm + ef}{|q_3 - q_1|^3}(q_3 - q_1) + \frac{-GMm + ef}{|q_3 - q_2|^3}(q_3 - q_2)
\end{aligned}$$

As the configuration is always an isosceles triangle, we can choose more suitable coordinates (x_1, x_2) to describe relative positions. Let x_1 be half the oriented distance between q_1 and q_2 and x_2 be the oriented distance between q_3 and the middle point of the segment $\overline{q_1 q_2}$ (Jacobi coordinates) (Fig. 2).

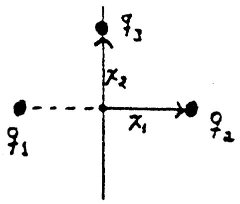


Fig. 2. Jacobi Coordinates

$$x_1 = \frac{q_2 - q_1}{2}$$
$$x_2 = q_3 - \frac{q_1 - q_2}{2}$$

We can think of each of these coordinates as independent, so our configuration space now is $(x_1, x_2) \in \mathbf{R}^2$. In these new coordinates and after rescaling time (set new time $z = (Gm - \frac{ef}{M}t)$ the equations of motion are much simpler

$$\begin{cases} \ddot{x}_1 = -\frac{x_1}{(x_1^2+x_2^2)^{3/2}} + \frac{Q\, x_1}{|x_1|^3} \\ \ddot{x}_2 = (\frac{2+m}{m})\frac{-x_2}{(x_1^2+x_2^2)^{3/2}} \end{cases} \tag{1.1}$$

where the parameter $Q = \frac{e^2 - GM^2}{4(GMm - ef)}$ is measuring the difference between gravitational and "electrical" forces. Remember that e and f have opposite signs, so this denominator is always positive. $Q < 0$ would mean that we have stronger gravitational forces and so particles q_1 and q_2 attract each other and we can have double collisions. Notice that if $Q = -1$ we have the eqs. of the isosceles 3-body problem [D]. If $Q = 0$, gravitational and electrical forces cancel each other and there is no interaction between these two particles although the third one attracts both. We do not have double collisions between q_1 and q_2 in the proper sense. They pass through each other naturally and there is no need of regularization. The equations are exactly the same as those for the anisotropic Kepler problem studied by Gutzwiller and Devaney. In this paper we will be interested in the case $Q > 0$ where there is repulsion between the symmetric bodies. In this case it is possible for triple collisions to occur, but not double collisions. This can be clearly seen from the shape of the Hill's regions that we will discuss in a moment.

Defining the potential $V(x) = \dfrac{-1}{(x_1^2 + x_2^2)^{1/2}} + \dfrac{Q}{|x_1|}$, $M = \begin{pmatrix} 1 & 0 \\ 0 & \frac{m}{2+m} \end{pmatrix}$

as mass matrix and $y = M\dot{x}$, where $x = (x_1, x_2) \in \mathbf{R}^2$, we can write the equations of motion as $M\ddot{x} = -\nabla V(x)$ and in Hamiltonian form:

$$\begin{cases} \dot{x} = M^{-1}y \\ \dot{y} = -\nabla V(x) \end{cases}$$

with Hamiltonian $H(x, y) = \frac{1}{2}y^t M^{-1} y + V(x)$.
H is also called the energy integral and, as is well known, it is constant along any orbit. Consider an orbit with $H = h$. As the first term of H is always positive, along the orbit the inequality $V(x) \leq h$ holds. That is, the orbit is confined to the region in the $x = (x_1, x_2)$ plane where $V(x) \leq h$, the so-called Hill region.

In Fig. 3 we sketch for different values of Q and h the form of the corresponding Hill region.

The shaded regions are the corresponding Hill's regions $V(x) \leq h$ where the movement is confined.

ϕ means that there is no possible movement with such values of h.

§2. The Collision Manifold

We now study the behavior of this system near triple collision. We blow up the singularity at the origin replacing it by a 2-manifold (collision manifold) which is invariant under the natural extension of the original flow. This is done by a change of coordinates due to McGehee [M].

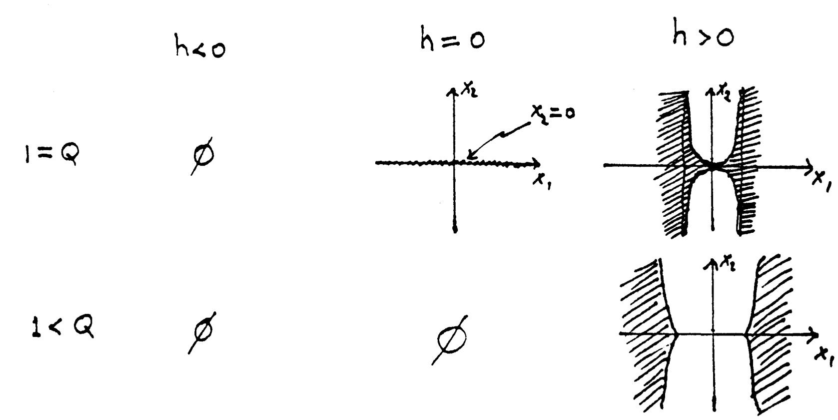


Fig. 3a.

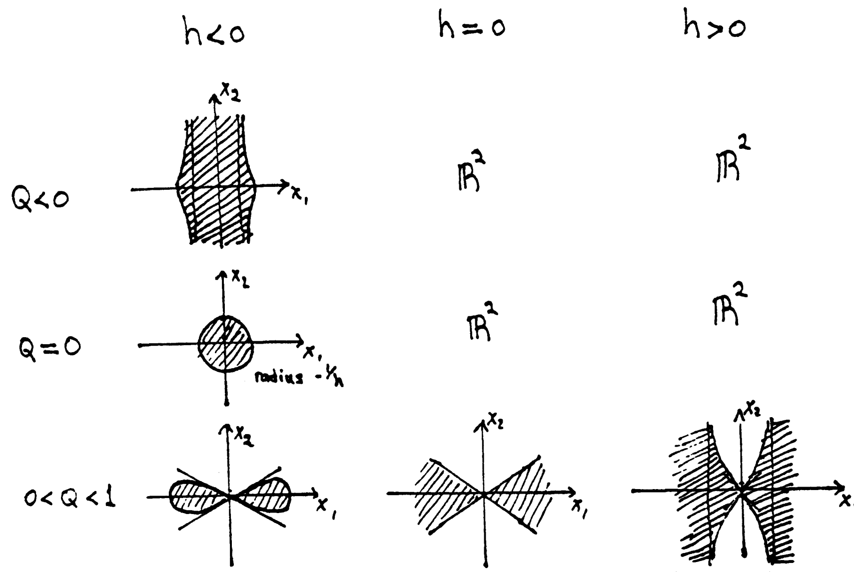


Fig. 3b.

In the case $0 < Q < 1$ that we will consider here, a simple two sphere arises as triple-collision manifold.

McGehee's coordinates are given by

$$\begin{aligned} r &= (x^\tau M x)^{1/2} \\ s &= r^{-1} x \\ v &= r^{1/2} s^t y \\ u &= r^{1/2}(M^{-1} y - y s) \end{aligned}$$

It is easily seen that $s^t M s = 1$, which defines an ellipse. Taking an angular coordinate θ on this ellipse, a new variable $\overline{u}$ such that $\overline{u}^2 = u^t M u$ and rescaling time by a factor $r^{3/2}$, the equations of motion (1.1) become then (see [D] for details)

$$\begin{cases} \dot{r} = rv \\ \dot{v} = u^2 + \frac{1}{2}v^2 + V(\theta) \\ \dot{\theta} = u \\ \dot{u} = -\frac{1}{2}uv - V'(\theta) \end{cases} \tag{2.1}$$

with the energy relation (integral)

$$rh = \frac{1}{2}(u^2 + v^2) + V(\theta) \tag{2.2}$$

where

$$V(\theta) = \frac{-1}{\sqrt{1 + \frac{2}{m}\sin^2\theta}} + \frac{Q}{|\cos\theta|}$$

and $V'(\theta)$ denotes derivative respect to θ.

The geometric meaning of these variables is essentially the following: In the (x_1, x_2)-plane

r measures the distance to the origin,

θ is the angle of the position vector (x_1, x_2),

u is the angular velocity, and

v is the radial component of the velocity.

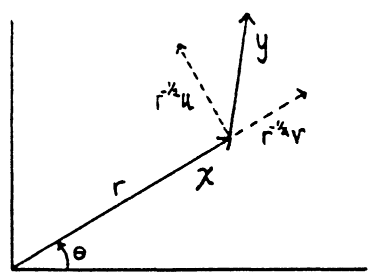


Fig. 4. McGehee coordinates

In this form, the equations no longer have singularities and the flow extends to $r = 0$ where we have the collision manifold. This is the common intersection of all the energy manifolds (2.2) with $r = 0$. We call this manifold Λ.

As we have the energy integral (2.2), we can reduce one dimension by looking only at the eqs. for the $(u, \theta, v) \in \mathbf{R}^3$ variables. We will try to visualize the flow in this space.

So Λ is given by

$$\Lambda : \begin{cases} r = 0 \\ \frac{1}{2}(u^2 + v^2) + V(\theta) = 0 \end{cases} \tag{2.3}$$

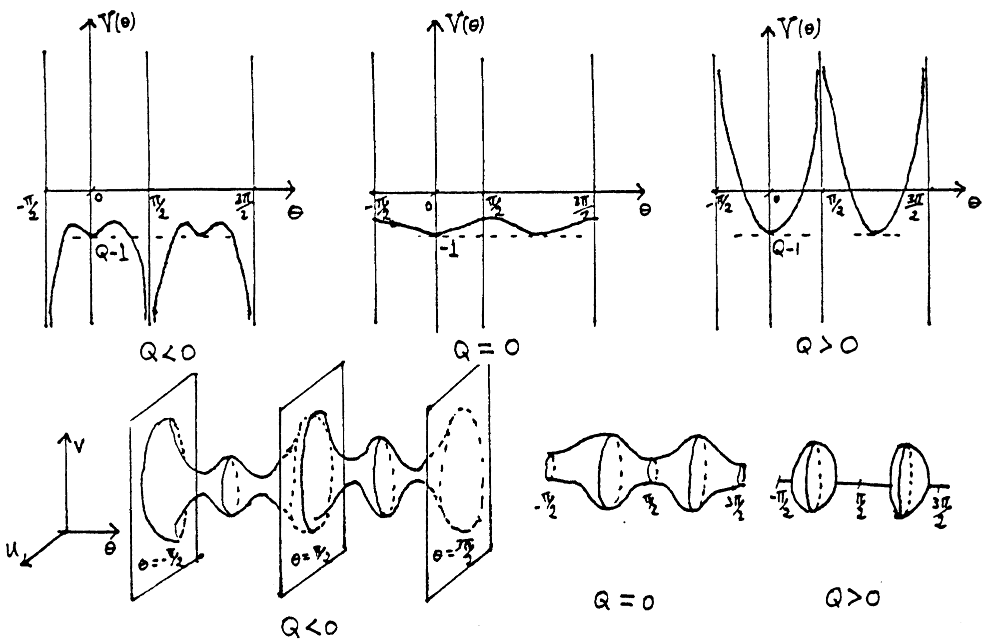
In (u, θ, v) coordinates, Λ is a surface of revolution around the θ-axis.

In Fig. 5 we sketch the graph of $V(\theta)$ for the three cases $Q < 0$, $Q = 0$ and $Q > 0$.

From (2.3) we see that we have collision manifold for those values of θ where $V(\theta) \leq 0$. To visualize Λ in (u, θ, v) coordinates, we rotate the portion of the graph of $V(\theta)$ that lies below the θ-axis and rotate it around this axis creating a surface of revolution. See Fig. 6.

Notice here what we could already see sketching the Hill's region. For $Q = 1$, Λ is a single point and for $Q > 1$ Λ is empty (no collisions).

Fixing $h < 0$, the energy manifold (2.2) in the (θ, u, v) space is the interior of the collision manifold and it is here where the flow takes place. The planes $\theta = -\pi/2$ and $\theta = \pi/2$ correspond to double collisions (bodies q_1 and q_2 are colliding). For $Q = -1$ we see in the figure the collision manifold of the isosceles 3-body problem without regularizing double collisions. The bodies collide with infinite speed. If we increase Q, i.e. increasing



Figs. 5,6. The potential function $V(\theta)$ and the collision manifold Λ.

"electrical" charges, Λ remains with same shape until we reach $Q = 0$ where there's no interaction between bodies q_1 and q_2. They can interchange positions by passing through each other. In (u, θ, v) coordinates, orbits can cross these "collision planes" naturally. Notice that for $Q = 0$, equations (1.1) are exactly those of the so-called Anisotropic Kepler problem [D]. Now, if $Q > 0$, the two bodies (q_1 and q_2) repel each other. Movement is confined to the interior of one of the two spheres. This corresponds to having q_1 to the left and q_2 to the right or vice versa.

For the remainder of this paper we will focus on the case $0 < Q < 1$ where $\Lambda \cong S^2$. We study the flow on Λ and later we see the influence it has over near collision periodic orbits.

Flow of Λ.

The natural extension of the flow to the collision manifold Λ is given by

$$\begin{cases} \dot{v} = \frac{1}{2}u^2 \\ \dot{\theta} = u \\ \dot{u} = -\frac{1}{2}uv - V'(\theta) \end{cases}$$

There are only two fixed points on this "sphere," the North and South poles $u = \theta = 0$ $v_0 = \pm\sqrt{-2(Q-1)}$. As $\dot{\theta} = u$ and $\dot{v} = \frac{1}{2}u^2 \geq 0$, the flow is spiralling counterclockwise from the South to the North poles around the v axis (Fig. 7).

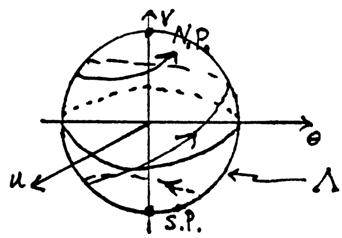
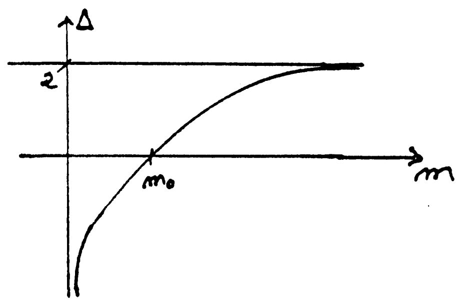


Fig. 7

Let's see what happens at the poles. The linearized flow at the fixed poles has matrix

$$\begin{pmatrix} v_0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -V''(0) & -v_0/2 \end{pmatrix}$$

with eigenvalues $\lambda = \frac{-v_0 \pm \sqrt{\Delta}}{4}$ in the (u, θ) plane where $\Delta = 18(\frac{m-16}{9m} - Q)$. Bifurcation occurs at $\Delta = 0$. The graph of Δ as a function of m is (recall that Q depends on m also):



Let m_0 be the value of m for which $\Delta = 0$.

For these eigenvalues we know that qualitatively we have 3 different local behaviors of the linearized system at the poles (Fig. 8):

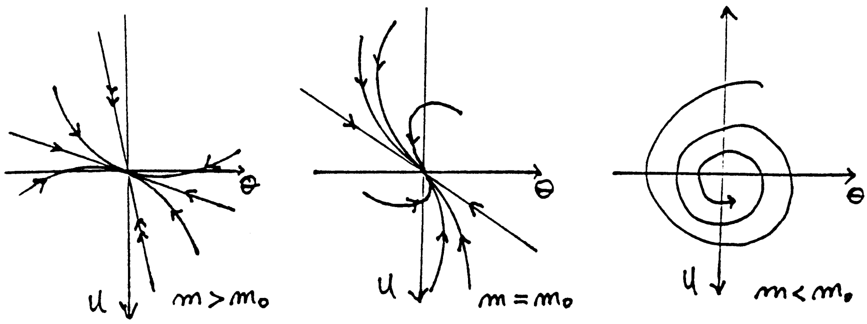


Fig. 8. Local flow at the poles

Similarly for the South pole, reverse arrows and change orientation. We note that for $m = m_0, \lambda_1 = \lambda_2$ but the linear part is not diagonalizable.

We know that the nonlinear flow will have the same qualitative behavior around the origin. We will be interested if there is or there isn't infinite spiralling around the origin.

We summarize this in:

Proposition 1

1. The collision manifold Λ has only two fixed points, the South and North poles.
2. The flow spirals around Λ from South to North.
3. For $m \geq m_0$ the flow spirals finitely many times on its way from South to North.
4. For $m < m_0$ the flow spirals infinitely many times on its way from South to North.

Flow in the energy manifold.

For a fixed energy value $H = h < 0$, we look now at the flow away from the collision manifold Λ. We can visualize the energy manifold $rh = \frac{1}{2}(u^2 + v^2) + V(\theta)$, which contains the flow for $H = h$, as the interior of the "sphere" Λ in the (u, θ, v) coordinates.

The equations of motion are given by the last three in (2.1). (The coordinate r can always be read from the energy relation). Note that we have two important symmetries: if (r, u, θ, v) is a solution, then

(a) $(r(-t), -u(-t), \theta(-t), -v(-t))$

(b) $(r(t), -u(t), -\theta(t), v(t))$ are also solutions.

(b) tells us that the flow is symmetric with respect to the v-axis and (a) says that the system is reversible, i.e., symmetric respect to the θ-axis changing time orientation.

Definition

We will call the θ-axis ($u = v = 0$) the zero velocity line. (Recall that u and v are components of the velocity vector y.)

An immediate consequence of the symmetry (b)(reversibility), is that an orbit crossing twice the zero velocity line is necessarily periodic. In original coordinates (config. space) the zero velocity line is the border of the Hill's region. An orbit touching it must necessary trace back on itself (see [D]) (Fig. 9).

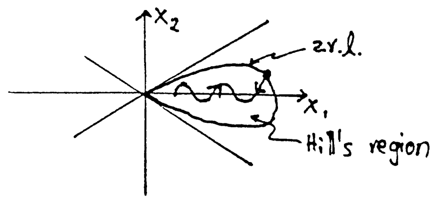
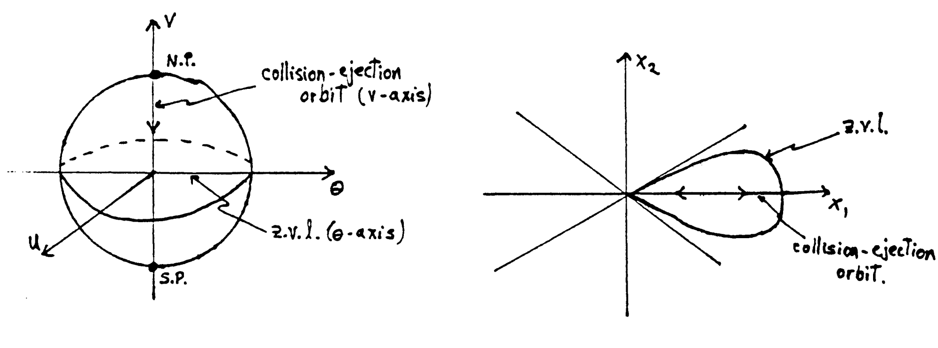


Fig. 9.

So we find symmetric periodic orbits by following the zero velocity line under the flow looking for second crossings.

It is easily seen from the equations (2.1) that there are no fixed points for the flow away from the collision manifold Λ, i.e. in the sphere's interior.

As $V'(0) = 0$, the v-axis ($\theta = u = 0$) is invariant under the flow, being itself an orbit where $\dot{v} = \frac{1}{2}v^2 + Q - 1$. It is a heteroclinic orbit going from the North pole (ejection) into the South pole (collision). Actually, it is the only collision orbit for a fixed value of the energy H. We will refer to it as the collision-ejection orbit which in configuration space runs over the x_1-axis (see Fig. 10).



Phase space Config. Space

Fig. 10

Every other orbit crosses the semi-disc $u = 0$, $\theta > 0$ (and does so infinitely many times), i.e., we have:

Proposition 2

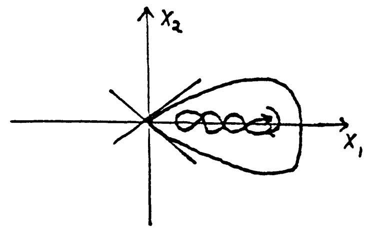
The 2-dimensional "semi-disc" $u = 0,\ \theta > 0$ is a Poincaré Section of the flow.

Proof.

Intersections are transversal because $\dot{u} = -V'(\theta) < 0$ on it. An orbit crosses and continually returns this section because $\dot{\theta} = u$ all over, i.e. the flow is always going in turns around the V-axis.

Remark Any semi-disc that is the intersection with a vertical plane containing v-axis is also a Poincaré-Section.

So the flow near the v-axis (collision-ejection orbit) goes down towards the South pole and then goes spiralling up near the surface (collision manifold Λ) towards the North pole to go down again. A typical orbit near the collision ejection orbit in configuration space:



§3 An "∞-furcation" of periodic orbits

In this section we show the existence of a beautiful "∞-furcation" of periodic orbits coming off the collision-ejection orbit as the parameter m passes the value m_0.

In (u, θ, v)-space, let D be a small disc neighborhood disc of the origin contained in the horizontal plane $v = 0$. $D = \{(u, \theta, v) | u^2 + \theta^2 < \delta^2 \quad v = 0\}$. (Fig. 11.)

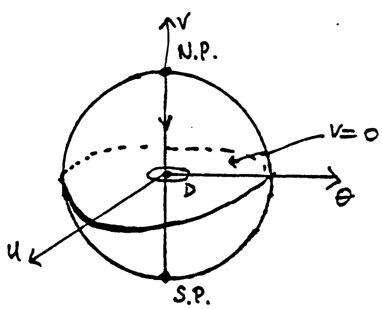


Fig. 11.

D has a segment of the zero-velocity line (θ-axis) as diameter and the collision-ejection orbit crosses it orthogonally at the origin which is the center of D.

Remark

The whole disc $v = 0$ is not a section of the flow. Some orbits touch it tangentially and some may never cross it.

If D is small, it is crossed transversally by a tubular neighborhood of the collision-ejection orbit. We define the Poincaré map F (first return map) from D (or a smaller subdisc if necessary) to D (see fig. 12). That is, take $x \in D$ and follow its orbit. This orbit goes first

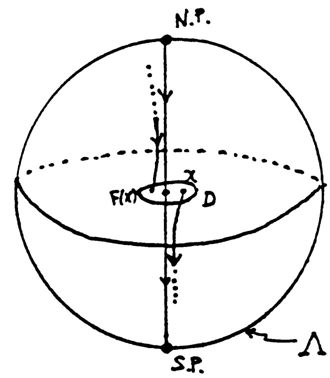


Fig. 12.

near the collision orbit towards the South pole getting close to the collision manifold Λ. It then spirals up towards the North pole to get close again to the collision-ejection orbit and it will cross D on its way down. This new point of intersection is $F(x)$. By continuity we define $F(0) = 0$.

In this way periodic points of F correspond to periodic orbits of the flow.

The map $F : D \to D$ inherits symmetries from the flow. It is symmetric with respect to the origin (central symmetry) and has reversible symmetry with respect to the θ-axis; that is, if R denotes reflexion with respect to the θ-axis, then $F^{-1} = RFR$ i.e. we have

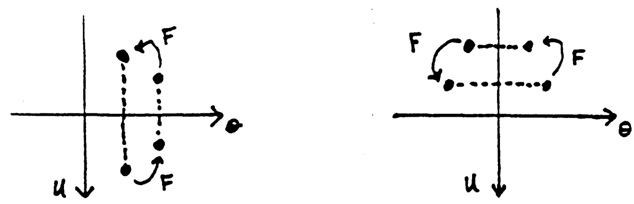
Proposition 3 *The map F has the following properties (see Fig. 13):*

(i) $F = -F(-\theta, -u)$

(ii) $F^{-1} = RFR$

as a consequence, F is also reversible w.r.t. the u-axis

(iii) $F^{-1} = SFS$ *where S is the u-axis reflection.*



reversibility (ii) reversibility (iii)

Fig. 13.

Let Θ and U denote the segments of the θ-axis and the u-axis contained in D.

Corollary

The sets $F(\Theta) \cap \Theta$ and $F(U) \cap U$ are invariant consisting of periodic points. Moreover, $F^2 = Id$ on these sets.

So, as for any reversible map, to find periodic orbits one can search for intersections of iterates of the symmetry axis with itself.

We will search for this kind of periodic orbits in our system. In configuration space, typical periodic orbits of this kind are shown in figure 14:

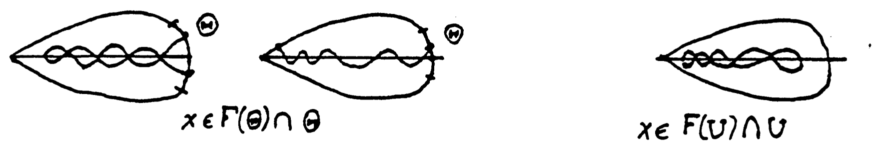


Fig. 14.

We take the segment Θ and we follow it along the flow until it crosses D again to see how $F(\Theta)$ looks like (Fig. 15).

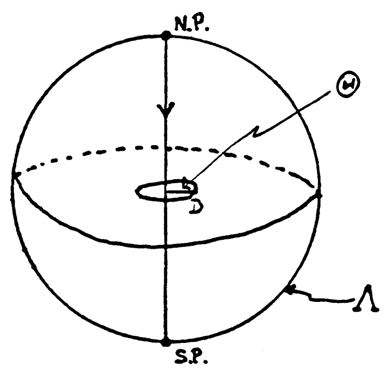


Fig. 15.

As D is small, Θ will remain close to the collision ejection orbit towards the South pole. Then it will spiral around the collision manifold Λ reaching near the North pole to slide down again nearby the collision-ejection orbit reaching D. See that Θ flows inside a tubular neighborhood of both the collision-ejection orbit and the collision manifold Λ. This tubular neighborhood is a solid torus with a sphere (Λ) in its interior. See Fig. 16.

We add another copy of the collision-ejection orbit (See Fig. 17).

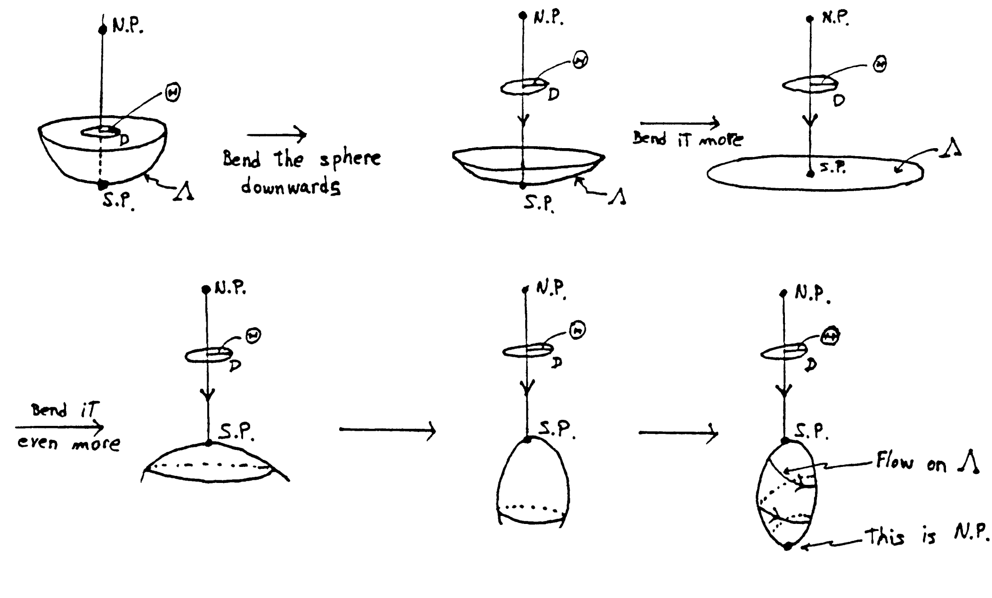


Fig. 16

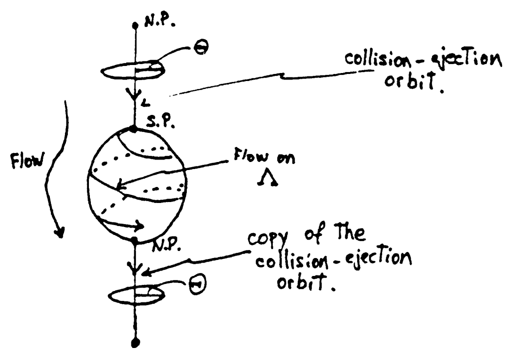


Fig. 17

Now it is easy to see what the image $F(\Theta) \subset D$ looks like. $F(\Theta)$ will have finite or infinite turns around the origin depending on the flow on the collision manifold Λ. More precisely, this depends on the flow around the poles. From Proposition 1, for $m > m_0$ and

$m = m_0$ the flow on Λ does a finite number of turns. Therefore, $F(\Theta)$ crosses Θ at most at finitely many points. The local picture is depicted in Fig. 18.

That is, $F(\Theta) \cap \Theta$ is a finite set of (periodic) points. But when $m < m_0$ we have on Λ infinite spiraling at both poles. In consequence, $F(\Theta)$ will be an infinite spiral around the origin, having ∞-many intersections with Θ accumulating to the origin (See Fig. 19):

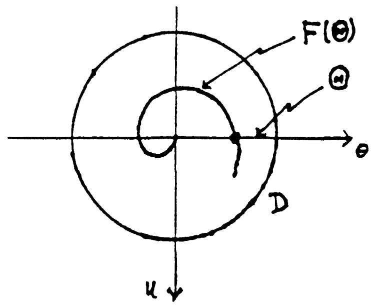


Fig. 18

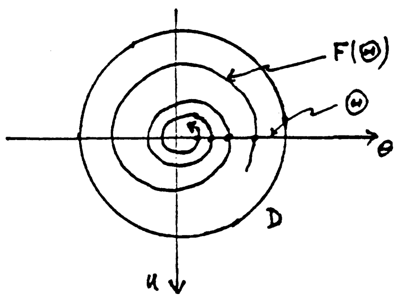


Fig. 19

The points of $\Theta \cap F(\Theta)$ correspond to periodic orbits of the flow. In fact, we have

Proposition

$\Theta \cap F(\Theta)$ *are fixed points of* F

Proof

Orient both Θ and $F(\Theta)$ saying that the origin 0 is a left extreme.

Take $x \in \Theta \cap F(\Theta)$. $F(x)$ also lies on this set (Corollary of Proposition 3). As F preserves the given orientation and $F(F(x)) = x$ (same Corollary), $F(x)$ cannot be to the left of x nor to the right. Therefore $F(x) = x$.

So, for $m > m_0$ there is a finite number (at most) of fixed points of F. As m decreases perhaps the origin spits new ones but one by one as shown in Fig. 20.

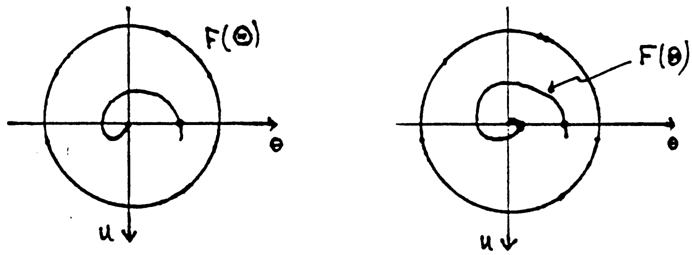


Fig. 20

For $m = m_0$ we still have finitely many of these fixed points. If m crosses the value m_0 to become $m < m_0$ we have a sudden "explosion" (∞-bifurcation) of fixed points of F.

The local bifurcation diagram is in Fig. 21:

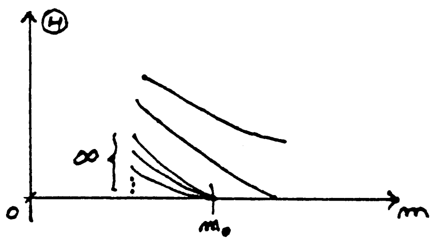


Fig. 21. Bifurcation diagram

Thus, we have proved

Theorem *In the Hamiltonian system (1.1), if $Q > 0$, as m decreases and crosses the value m_0, there is a simultaneous appearance of ∞-many periodic orbits (∞-furcation) of low period coming off the collision-ejection orbit. For $m \geq m_0$ there are only finitely many of them. For $m < m_0$ there are ∞-many.*

Acknowledgment. The author would like to thank R. Devaney for introducing him to the problem and giving continuous encouragement during this work.

References

[D] R. Devaney, Singularities in Classical Mechanical Systems, *Ergodic Theory and Dynamical Systems I*, Birkhäuser, Boston, 1981, 211-333.

[M] R. McGehee, Triple collision in the collinear three body problem, *Invent. Math.* 27 (1974), 191-227.

http://dx.doi.org/10.1090/conm/081/986257

Contemporary Mathematics
Volume 81, 1988

Central Configurations in R^2 and R^3

Dieter S. Schmidt[1]

ABSTRACT. Relative equilibria and central configurations are specialized solutions of the general N-body problem. The mutual distances between the N bodies can be used as coordinates. It will simplify the discussion of known results about planar relative equilibrium solutions of the 4 body problem. The same method can be used to derive new results for non planar central configurations of the 5 body problem.

1. INTRODUCTION. Relative equilibria are stationary solutions of the N-body problem in a rotating plane. Dziobek (1900) used the mutual distances between the bodies as coordinates. He derived a necessary and sufficient condition under which a configuration of four bodies can be a relative equilibrium solution.

MacMillan and Bartky (1932) published an extensive treatment on permanent configurations in the problem of four bodies. It appears that they were not aware of the work of Dziobek as they used their own set of coordinates which makes their paper more difficult than necessary. We will show that, with the help of the 6 mutual distances between the four bodies, their results can be derived more easily.

Another contribution to the problem of relative equilibria for four bodies was made by Palmore (1973). He saw that some relative equilibria are degenerate in a certain sense and that this allows for the bifurcation of new families of relative equilibria. Whereas Palmore used Morse theory and obtained qualitative results, Meyer and Schmidt (1987a) have shown that, here too, the coordinates used by Dziobek will simplify the problem to the point where approximations to the solution can be calculated and the implicit function theorem can be used to guarantee uniqueness. We will outline these results because our method can be extended to deal with the corresponding

1980 Mathematics Subject Classification (1985 Revision). 70F10

[1]This research was supported by a grant from the Applied and Computational Mathematics Program of DARPA.

problem in R^3.

It is impossible to have relative equilibria solutions unless all bodies lie in a plane. Nevertheless the N bodies can maintain a similar configuration in R^3 provided that the size of the configuration increases (or decreases) proportionally to $t^{2/3}$ where t is time. Such a solution to the N-body problem is called a central configuration and necessarily starts (or ends) with a total collapse of the system. Although these solutions are very special they are of interest. It can be shown that whenever several bodies collide they will assume a central configuration in the limit.

It turns out that the determining equations for relative equilibria and central configurations have the same form. Therefore, planar solutions to these equations can be interpreted as relative equilibria or central configurations whereas non planar solutions can only be viewed as central configurations. For simplicity we will refer to planar solutions always as relative equilibria and we will use the term central configurations only in the three dimensional case. At times when the discussion applies to both cases or when it is clear from the context we may simply use the word configuration to refer to either case.

For three bodies all relative equilibria are given by the collinear solutions of Euler and the equilateral triangular solutions of Lagrange. The common theme for this presentation is the use of mutual distances between the bodies as coordinates but they are singular in the collinear case. Therefore, we have to exclude the collinear case from our discussion.

For four bodies it is already impossible to enumerate all relative equilibria solutions. In addition to the papers mentioned above see also Simo (1978) for a detailed numerical study of the problem. Pedersen (1944) studied the problem when one of the bodies is infinitesimally small. The same problem is also the starting point of a work by Arenstorf (1982) where he asks which solutions can be continued to the full four body problem.

Central configurations in R^3 of 4 bodies are mentioned in Brumberg (1958) but the major portion of this paper is on the stability of relative equilibria in the planar problem. References to relative equilibria in the five body problem are Williams (1938), Palmore (1976) and Meyer and Schmidt (1987b).

For more than 5 bodies mutual distances are not so useful as coordinates because the number of geometric constraints increases quadratically. Specialized relative equilibria solutions for any number of bodies can be found in Longley (1907), Lindow (1926), Klemperer (1962) and others. The use of bifurcation methods was pioneered by Palmore (1976) and the corresponding calculations were carried out in Meyer and Schmidt (1987b).

2. EQUATIONS OF MOTION: The Hamiltonian function

$$H = \sum_{j=1}^{N} \frac{|p_j|^2}{2\, m_j} - U(q)$$

describes the N-body problem. The N particles with mass m_j $j=1,\dots,N$ are located at the positions $q_j \in R^2$ (or $q_j \in R^3$). The corresponding momenta are p_j. The vector q has dimension 2N (or 3N) and is made up of the individual position coordinates. U is the negative of the potential function i.e.

$$U(q) = \sum_{i<j} \frac{m_i\, m_j}{|q_i - q_j|} \quad .$$

The differential equations of motion in Newtonian form are

$$(1) \qquad m_j\, \ddot{q}_j = \frac{\partial U}{\partial q_j} \quad .$$

For the study of the motion in the plane it is convenient to think of q_j as complex valued coordinates. If the motion is to be studied in a uniformly rotating coordinate system the transformation to the new complex coordinates z_j is then given by

$$q_j = e^{i\nu t}\, z_j$$

where ν is the angular velocity of the rotating coordinate frame. Relative equilibria are stationary solutions of the N-body problem in this rotating coordinate system. In this case the variables z_j are independent of time and the differential equation (1) reduces to the algebraic system

$$(2) \qquad \frac{\partial U}{\partial z} + \lambda\, m_j\, z_j = 0$$

where we have set $\lambda = \nu^2$.

Central configurations are solutions of (1) of the form

$$q_j = t^{2/3} z_j$$

where the coordinates z in R^2 (or in R^3) are again independent of time. It leads to an equation which is identical to (2) with $\lambda = 2/9$.

The form of (2) suggests that λ can be interpreted as a Lagrange multiplier. In this way the solution of (2) are seen to be the extrema of U under the restriction that the moment of inertia

$$I = \frac{1}{2} \sum m_j \, |z_j|^2$$

remains at a fixed value I_0. Thus relative equilibria and central configurations are extrema of the function

(3) $\quad U + \lambda (I - I_0) .$

In the search for these extrema there exists an additional condition which says that λ is positive or even more stringently that it has to take on a predetermined value. It turns out that this poses no constraint. Since U is homogeneous of degree -1 it follows from Euler's theorem on homogeneous functions and from the form of (2) that $\lambda = \frac{U}{2\,I}$. This shows that $\lambda > 0$ at the extrema of (3). By scaling the position vectors q_j uniformly one achieves also that λ is equal to 2/9.

By summing (2) over all bodies it follows from the form of the potential function that $\sum m_j z_j = 0$. This implies that the center of mass has to be at the origin. It also follows from (2) that other configurations can be obtained from any given configuration by rotating it by a finite angle around the origin. This fact causes difficulties. When we discuss bifurcations we will look for degenerate configurations. In this context a configuration is called degenerate if it is not isolated but, in fact, no configuration is isolated because near each configuration there is another one which is obtained from the first by a simple rotation.

For this reason one has to introduce equivalence classes of configurations. Two configurations are called equivalent if they can be transformed into each other by a rotation around the center of mass. Degeneracy means then degeneracy among equivalence classes, i.e. mod SO(2) (or mod SO(3)).

In many cases difficulties of this nature can be avoided by choosing the proper coordinate system. For the N-body problem the mutual distances between the bodies appear to be the appropriate choice. These coordinates were already used by Dziobek (1900).

Let r_{ij} denote the distance between the i^{th} and the j^{th} body. Then the (negative) potential function is

$$U = \sum \frac{m_i \, m_j}{r_{ij}}$$

and the moment of inertia transforms to $I = \frac{1}{2} \sum m_i m_j r_{ij}^2$.
To be correct I should be divided by the total mass $M = \Sigma m_i$ but we absorb this factor into the Lagrange multiplier by setting $\delta = \lambda/M$. The problem consists then in studying the extrema of $U + \delta (I - I_0)$.

3. **RELATIVE EQUILIBRIA FOR THE FOUR BODY PROBLEM:** For three bodies the three distances r_{12}, r_{23} and r_{13} determine the configuration up to a reflection. This agrees with the fact that there are two non linear equivalence classes of relative equilibria for the three body problem.

Between four bodies there are six mutual distances. In general six distances define a tetrahedron provided that some obvious geometric inequalities are met. If the configuration is to lie in a plane the volume of this tetrahedron has to be zero.

The following determinant is proportional to the square of this volume, i.e. $288 V^2 = F$ with

$$F = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12}^2 & r_{13}^2 & r_{14}^2 \\ 1 & r_{12}^2 & 0 & r_{23}^2 & r_{24}^2 \\ 1 & r_{13}^2 & r_{23}^2 & 0 & r_{34}^2 \\ 1 & r_{14}^2 & r_{24}^2 & r_{34}^2 & 0 \end{vmatrix} .$$

The derivatives of this determinant with respect to its entries have the following remarkable property

$$\frac{\partial F}{\partial r_{ij}^2} = 32 \, \Delta_i \, \Delta_j$$

where Δ_i and Δ_j are the areas of the triangles opposite to the points P_i and P_j respectively. These areas have to be taken with the appropriate orientation, that is relative to an outward normal of the tetrahedron. For Fig. 1 the individual triangles are

$$\Delta_1 = \Delta\, P_2P_3P_4 \quad > 0$$
$$\Delta_2 = \Delta\, P_1P_4P_3 \quad < 0$$
$$\Delta_3 = \Delta\, P_1P_2P_4 \quad > 0$$
$$\Delta_4 = \Delta\, P_1P_3P_4 \quad < 0$$

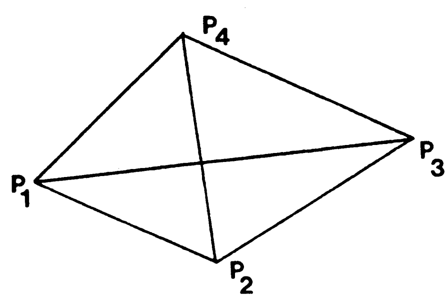


Figure 1: Convex configuration

In order for the configuration to lie in a plane $\Delta_1+\Delta_2+\Delta_3+\Delta_4=0$ has to hold always.

Let σ be the Lagrange multiplier which insures that the volume of the tetrahedron remains zero. Then the relative equilibria of the four body problem are the extrema of

$$U + \delta\ (I - I_0) + \sigma\ F/32 \ .$$

After differentiating with respect to r_{ij}, δ and σ these extrema are found from the following set of eight algebraic equations

$$m_i\ m_j\ (r_{ij}^{-3} - \delta) = \sigma\ \Delta_i\ \Delta_j \qquad 1 \le i < j \le 4 \tag{4a}$$

$$I - I_0 = 0 \tag{4b}$$

$$F = 0 \ . \tag{4c}$$

We would like to remark briefly that for the three body problem no geometric constraint of the form (4c) exists. The right hand sides of the equations in (4a) are therefore 0 and j goes only up to 3. The solution to this set of equations is then $r_{12}= r_{23}= r_{13}= \delta^{-1/3}$. Thus, Lagrange's equilateral triangular solution is obtained without any effort. On the other hand the existence of the collinear solutions is not even indicated by these equations due to the fact that these coordinates are singular there.

Returning to the four body problem the six equations of (4a) are

$$m_1\ m_2\ (r_{12}^{-3} - \delta) = \sigma\ \Delta_1\ \Delta_2 \ , \qquad m_3\ m_4\ (r_{34}^{-3} - \delta) = \sigma\ \Delta_3\ \Delta_4$$

$$m_1\ m_3\ (r_{13}^{-3} - \delta) = \sigma\ \Delta_1\ \Delta_3 \ , \qquad m_2\ m_4\ (r_{24}^{-3} - \delta) = \sigma\ \Delta_2\ \Delta_4$$

$$m_1\ m_4\ (r_{14}^{-3} - \delta) = \sigma\ \Delta_1\ \Delta_4 \ , \qquad m_2\ m_3\ (r_{23}^{-3} - \delta) = \sigma\ \Delta_2\ \Delta_3 \ .$$

The equations have been grouped so that when they are multiplied together pair wise, their right hand sides are identical. This leads to the well known condition

$$(r_{12}^{-3} - \delta)(r_{34}^{-3} - \delta) = (r_{13}^{-3} - \delta)(r_{24}^{-3} - \delta) = (r_{14}^{-3} - \delta)(r_{23}^{-3} - \delta) \ . \tag{5}$$

Besides the geometric constraint this is a necessary and sufficient condition for solving equations (4) for some m_i when the distances are given. It is sufficient because it guarantees that one can solve for δ and with it the ratios for the masses can be found.

If the condition (5) is satisfied than δ can be found from 3 expressions which have to give the same answer, that is

$$\delta = \frac{s_{12}s_{34} - s_{13}s_{24}}{s_{12} + s_{34} - s_{13} - s_{24}} = \frac{s_{13}s_{24} - s_{14}s_{23}}{s_{13} + s_{24} - s_{14} - s_{23}} = \frac{s_{14}s_{23} - s_{12}s_{34}}{s_{14} + s_{23} - s_{12} - s_{34}}$$

We have introduced the notation $s_{ij} = r_{ij}^{-3}$ and we will continue to use it from now on.

From the different expressions that are possible for $s_{ij} - \delta$ and by finding all possible expressions for the ratios of two masses from (4a) one obtains the following list

$$\text{(6a)} \qquad \frac{m_1 \Delta_2}{m_2 \Delta_1} = \frac{s_{23} - \delta}{s_{13} - \delta} = \frac{s_{24} - \delta}{s_{14} - \delta} = \frac{s_{23} - s_{24}}{s_{13} - s_{14}}$$

$$\text{(6b)} \qquad \frac{m_1 \Delta_3}{m_3 \Delta_1} = \frac{s_{23} - \delta}{s_{12} - \delta} = \frac{s_{34} - \delta}{s_{14} - \delta} = \frac{s_{23} - s_{34}}{s_{12} - s_{14}}$$

$$\text{(6c)} \qquad \frac{m_1 \Delta_4}{m_4 \Delta_1} = \frac{s_{24} - \delta}{s_{12} - \delta} = \frac{s_{34} - \delta}{s_{13} - \delta} = \frac{s_{24} - s_{34}}{s_{12} - s_{13}}$$

$$\text{(6d)} \qquad \frac{m_2 \Delta_3}{m_3 \Delta_2} = \frac{s_{13} - \delta}{s_{12} - \delta} = \frac{s_{34} - \delta}{s_{24} - \delta} = \frac{s_{13} - s_{34}}{s_{12} - s_{24}}$$

$$\text{(6e)} \qquad \frac{m_2 \Delta_4}{m_4 \Delta_2} = \frac{s_{14} - \delta}{s_{12} - \delta} = \frac{s_{34} - \delta}{s_{23} - \delta} = \frac{s_{14} - s_{34}}{s_{12} - s_{23}}$$

$$\text{(6f)} \qquad \frac{m_3 \Delta_4}{m_4 \Delta_3} = \frac{s_{14} - \delta}{s_{13} - \delta} = \frac{s_{24} - \delta}{s_{23} - \delta} = \frac{s_{14} - s_{24}}{s_{13} - s_{23}} .$$

Condition (5) only guarantees that the ratios of masses can be computed. It does not say that the masses will be positive. In order to find out under which conditions the masses will be positive requires an analysis of the above equations. This kind of analysis is the contribution of the paper by MacMillan and Bartky. We claim that their results follow more easily when (6a-f) are used.

We will treat symmetric relative equilibria before we go on to the general case. We assume that the masses m_1 and m_2 are equal and the masses m_3 and m_4 are equidistant from the first two. Referring to Fig. 2 we set α the oriented distance of m_4 from the line joining m_1 and m_2. With 2β the distance between the later two masses we have

$r_{12} = 2\beta$

$r_{13} = r_{23} = \sqrt{(\alpha+1)^2 + \beta^2}$

$r_{14} = r_{24} = \sqrt{\alpha^2 + \beta^2}$

$r_{34} = 1$

$\Delta_1 = \Delta_2 = \beta/2$

$\Delta_3 = \alpha\beta$

$\Delta_4 = -(1+\alpha)\beta$

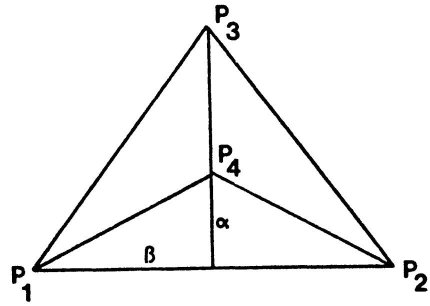


Figure 2) Concave symmetric configuration $\alpha > 0$

Actually, we should work with ratios of distances but for simplicity we have chosen to fix the value of r_{34} to be 1. Note that with this assumption the configuration will be concave for $\alpha > 0$ and also for $\alpha < -1$. In the first case m_4 is the interior point in the second case it is m_3. For $-1 < \alpha < 0$ the configuration will be convex.

With the above values for the distances and areas one finds from (6b,c)

$$\frac{m_3}{m_1} = 2\alpha \frac{(2\beta)^{-3} - (\alpha^2+\beta^2)^{-3/2}}{((\alpha+1)^2+\beta^2)^{-3/2} - 1}$$

$$\frac{m_4}{m_1} = 2(\alpha+1) \frac{(2\beta)^{-3} - ((\alpha+1)^2+\beta^2)^{-3/2}}{1 - (\alpha^2+\beta^2)^{-3/2}} .$$

By drawing the curves where the masses are 0 and ∞ it is easy to determine where both masses are positive. This is indicated by the shaded region in Fig. 3.

Some features of Fig. 3) deserve to be mentioned. For $\alpha = -1/2$ the configuration has an additional symmetry with respect to the line joining m_1 and m_2. At a) and also at b) the configuration is made up of two equilateral triangles butted against each other (Fig. 4). From these two extremes it follows that for a convex symmetric configuration $\sqrt{3}/3 < 2\beta < \sqrt{3}$ or in terms of the ratios of diagonals

$$\sqrt{3}/3 < \frac{r_{12}}{r_{34}} < \sqrt{3}$$

which is a theorem of Longley (1907).

The other vertices of the shaded regions also correspond to configurations where 3 bodies assume a Lagrangian configuration but the remaining mass has then to be zero or infinite. Only at c) in Fig. 3) can the

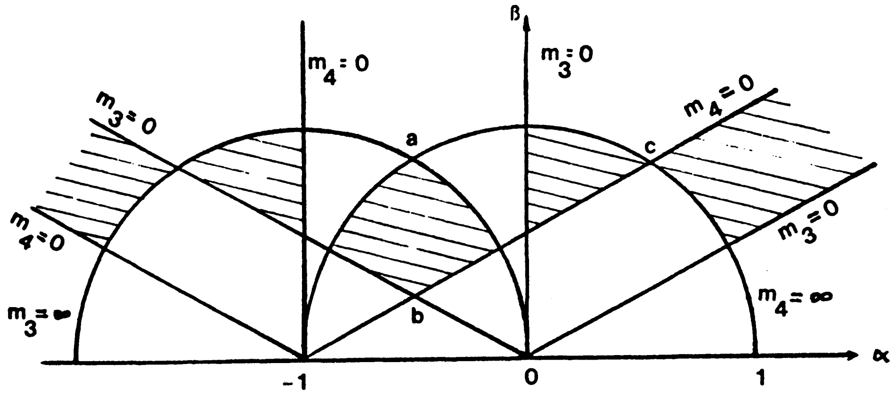
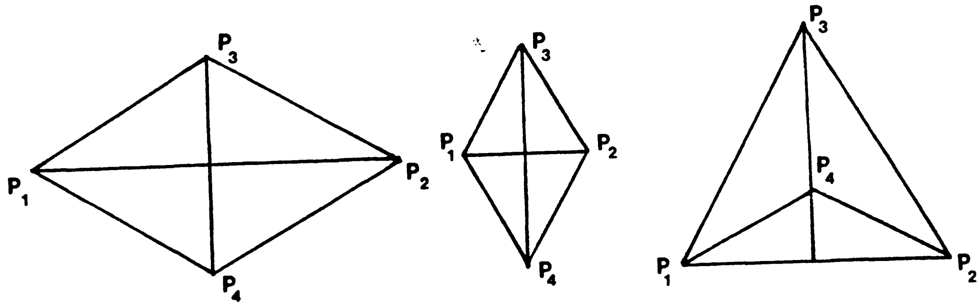


Fig. 3) Regions (shaded) with positive masses for symmetric configurations.



a) $\alpha=-1/2$, $\beta=\sqrt{3}/2$
$m_1=m_2$, $m_3=m_4=\infty$

b) $\alpha=-1/2$, $\beta=\sqrt{3}/6$
$m_1=m_2$, $m_3=m_4=0$

c) $\alpha=1/2$, $\beta=\sqrt{3}/2$
$m_1=m_2=m_3$, m_4=arbitrary

Fig 4) Three degenerate relative equilibria of Fig 3).

mass m_4 be arbitrary. It is the situation which Palmore has exploited to look for the bifurcation of new families of relative equilibria.

For the discussion of the general convex case we assume that the labeling of the vertices is as in Fig 1). The signs of the area of the triangles are then $\Delta_1>0$, $\Delta_2<0$, $\Delta_3>0$ and $\Delta_4<0$. We continue to use the notation $s_{ij}=r_{ij}^{-3}$ and we also set $\delta=\rho^{-3}$. If one starts with $s_{12}-\delta>0$ and asks for positive masses, equation (6) leads to the following inequalities

(7) $$r_{12}, r_{23}, r_{34}, r_{14} \le \rho \le r_{13}, r_{24}$$

i.e. all exterior sides are smaller than the diagonals. The last expression for the mass ratios in (6b,e) gives even more detailed information. If r_{12} is the smallest side than $r_{12}< r_{14}< r_{34}$ and $r_{12}< r_{23}< r_{34}$ which means that the shortest and longest side have to face each other. This result can be found already in Dziobek (1900).

If we had assumed initially that $s_{12} - \delta < 0$ then all inequalities would reverse but this can not be realized geometrically.

MacMillan and Bartky also introduced the variable ρ but in a more complicated way. Once they have accomplished this the rest of their paper consists in exploiting equations (5) and (7). The proof to the following theorems can be found in their paper.

THEOREM: *For any convex relative equilibrium the condition that the masses are positive imposes the following restriction on the ratios of the diagonals*

$$\sqrt{3}/3 < \frac{r_{12}}{r_{34}} < \sqrt{3}$$

THEOREM: *For any given ratio of (positive) masses* $m_1:m_2:m_3:m_4$ *there exists at least one convex relative equilibrium solution.*

For concave configurations fewer results are known. If m_4 is the interior mass (Fig. 5) then (6a-f) provides the following inequalities

(8) $$r_{12}, r_{23}, r_{13} > \rho > r_{14}, r_{24}, r_{34}$$

i.e. all exterior sides are longer than the interior ones. Furthermore, if one assumes an ordering of the exterior sides, for example

$$r_{12} > r_{23} > r_{13} \quad \text{then it follows that} \quad r_{34} > r_{14} > r_{24} .$$

It means that the longest exterior side lies opposite the longest interior side.

By considering each exterior edge one at a time, the interior body can then only lie in a half plane which is defined by its perpendicular bisector and which does not contain the remaining vertex. Thus the interior body has to lie in the intersection of three half planes. Since the three perpendicular bisectors intersect in one point this region will not be empty but it may lie outside of the given triangle, unless the center of its circumscribing circle

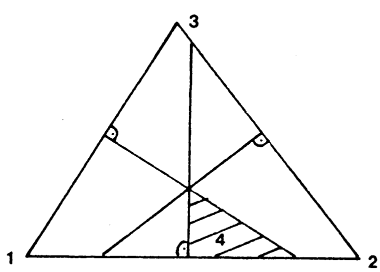


Fig 5) Concave configuration with region in which P_4 can lie when exterior triangle is fixed.

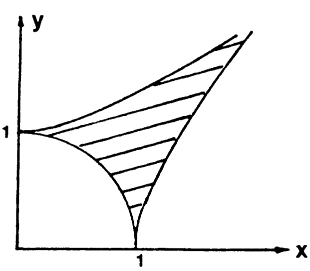


Fig 6) Restriction on the shape of exterior triangle in Fig 5) $x=r_{13}/r_{12}$, $y=r_{23}/r_{12}$

lies inside of it. This restriction can be expressed algebraically by the following inequalities $a^2 < b^2 + c^2$ where a, b and c are any permutations of the three distances r_{12}, r_{23} and r_{13}. A graphic representation of this restriction is given in Fig. 6.

A very special configuration is the one with three equal masses at the vertices of an equilateral triangle and a fourth body of arbitrary mass at the center. Palmore (1973) has shown that from this family of relative equilibria other families bifurcate at a particular value of the fourth mass.

Meyer and Schmidt (1987a) have shown that by using mutual distances the implicit function theorem provides a way for computing these families. This demonstrates that these families are unique and have to be symmetric. We outline this approach here. Fixed are the three masses $m_1 = m_2 = m_3 = 1$. The critical value of the fourth mass is denoted by m_c and has yet to be determined. With ϵ as the perturbation parameter we set

$$m_4 = m_c + \epsilon$$

$$r_{12} = \sqrt{3} + \epsilon\, a_{12} + \dots \qquad r_{14} = 1 + \epsilon\, a_{14} + \dots$$

$$r_{23} = \sqrt{3} + \epsilon\, a_{23} + \dots \qquad r_{24} = 1 + \epsilon\, a_{24} + \dots$$

$$r_{13} = \sqrt{3} + \epsilon\, a_{13} + \dots \qquad r_{34} = 1 + \epsilon\, a_{34} + \dots$$

As before we have to find the extrema of the function

$$V = U + \delta\,(I - I_0) + \sigma F .$$

V depends on eight variables which we combine into the vector

$$z = (\delta, \sigma, r_{12}, r_{23}, r_{13}, r_{14}, r_{24}, r_{34}).$$

For $\epsilon = 0$ V has an extremum if $\delta = \dfrac{3m_c + \sqrt{3}}{3m_c + 9}$ and $\sigma = \dfrac{(\sqrt{3} - 9)m_c}{27m_c + 81}$.

The Hessian with respect to z is

$$\frac{(532\sqrt{3} - 720)}{6889}\, m_c^2\, (249\, m_c - 64\sqrt{3} - 81)^2$$

For $m_4 = m_c$ the Hessian has a 6x6 nonzero subdeterminant which can be obtained by deleting the last two rows and columns. Thus by the implicit function theorem one can solve $\partial V/\partial z = 0$ for $\delta, \sigma, r_{12}, r_{23}, r_{13}, r_{14}$ in terms of a_{24} and a_{34} when ϵ is near zero. In terms of the Liapunov Schmidt reduction process this means that at first order in ϵ all variables can be expressed in terms of the two parameters a_{24} and a_{34}, that is

$$r_{12} = \sqrt{3} + \epsilon\,\alpha\,a_{34} + \ldots \qquad r_{14} = 1 - \epsilon\,(a_{24} + a_{34}) + \ldots$$
$$r_{23} = \sqrt{3} + \epsilon\,\alpha\,a_{24} + \ldots \qquad r_{24} = 1 + \epsilon\,a_{24} + \ldots$$
$$r_{13} = \sqrt{3} - \epsilon\,\alpha\,(a_{24} + a_{34}) + \ldots \qquad r_{34} = 1 + \epsilon\,a_{34} + \ldots$$

where $\alpha=(64\sqrt{3}+81)/83$.

Continuing with the Liapunov-Schmidt reduction process we find from the second order terms the following bifurcation equations

$$(a_{34} - \beta)(2\,a_{24} + a_{34}) = 0$$
$$(a_{24} - \beta)(a_{24} + 2\,a_{34}) = 0$$

where $\beta=(3089347\sqrt{3} - 4531167)/18889832$. This pair of equations has three nontrivial solutions

$$a_{24} = \beta, \qquad a_{34} = \beta$$
$$a_{24} = \beta, \qquad a_{34} = -2\beta$$
$$a_{24} = -2\beta, \qquad a_{34} = \beta$$

Each solution preserves one of the symmetries of the original equilateral triangle. For example from the second solution we obtain

$$r_{12} = \sqrt{3} - 2\,\epsilon\,\alpha\beta + \ldots$$
$$r_{13} = r_{23} = \sqrt{3} + \epsilon\,\alpha\beta + \ldots$$
$$r_{14} = r_{24} = 1 + \epsilon\,\beta + \ldots$$
$$r_{34} = 1 - 2\,\epsilon\,\beta + \ldots$$

Because we used the implicit function theorem to find these solutions, they are unique in the sense that no other solution can bifurcate from the family of equilateral configurations. In particular no scalene solution can exist nearby.

4. CENTRAL CONFIGURATIONS IN R^3:

The use of mutual distances as coordinates allows us to present new results for central configurations in the three dimensional space. Our methods are similar to those presented in the previous section and, therefore, the results will exhibit some similarity. Again we would like to point out that the coordinates defined by the mutual distances of the bodies in R^3 have a singularity when 4 bodies lie in a plane. Therefore, the results for planar configurations can not be obtained from what follows.

Six distances define a tetrahedron in R^3 provided that the obvious geometric inequalities are met. The tetrahedron is defined up to a reflection. For the four body problem central configurations are then the extrema of

$$U + \delta\,(I - I_0)\ .$$

From the partial derivatives one obtains six equations of the form

$$m_i\, m_j\, (r_{ij}^{-3} - \delta) = 0 \qquad 1\le i<j\le 4$$

Their solution is $r_{ij} = \delta^{-1/3}$. It means that the only central configuration for the four body problem are those that form a regular tetrahedron.

If the 10 mutual distances between 5 bodies allow for a geometric realization in R^3 then the following determinant has to be zero

$$F = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12}^2 & r_{13}^2 & r_{14}^2 & r_{15}^2 \\ 1 & r_{12}^2 & 0 & r_{23}^2 & r_{24}^2 & r_{25}^2 \\ 1 & r_{13}^2 & r_{23}^2 & 0 & r_{34}^2 & r_{35}^2 \\ 1 & r_{14}^2 & r_{24}^2 & r_{34}^2 & 0 & r_{45}^2 \\ 1 & r_{15}^2 & r_{25}^2 & r_{35}^2 & r_{45}^2 & 0 \end{vmatrix} .$$

Again we have a relationship of the form $\dfrac{\partial F}{\partial r_{ij}^2} = 576\, \Delta_i \Delta_j$ where Δ_i is the oriented tetrahedron opposite the point P_i.

We will give first our results of the bifurcation analysis. Place 4 bodies of mass 1 at the vertices of a regular tetrahedron and a body of arbitrary mass m_5 at its center. This is a central configuration for the five body problem and therefore an extremum of the function

$$V = U + \delta\,(I - I_0) + \sigma F.$$

The function V depends on 12 variables δ, σ, r_{ij} $1\le i<j\le 5$. The Hessian of V with respect to these variables is nonsingular for positive m_5 except for

$$m_5 = m_c = \frac{10368 + 1701\sqrt{6}}{54952}\ .$$

At this value the corresponding Jacobian reduces its rank by 3. If we set

$$m_5 = m_c + \epsilon$$

and use a first order bifurcation analysis we find that the mutual distances depend on three parameters, i.e.

$$r_{12} = \tfrac{2}{3}\sqrt{6} + \epsilon\,\gamma\,(a_{35} + a_{45}) + \dots$$

$$r_{13} = \tfrac{2}{3}\sqrt{6} + \epsilon\,\gamma\,(a_{25} + a_{45}) + \dots$$

$$r_{14} = \tfrac{2}{3}\sqrt{6} + \epsilon\,\gamma\,(a_{25} + a_{35}) + \dots$$

$$r_{23} = \tfrac{2}{3}\sqrt{6} - \epsilon\,\gamma\,(a_{25} + a_{35}) + \dots$$

$$r_{24} = \tfrac{2}{3}\sqrt{6} - \epsilon\,\gamma\,(a_{25} + a_{45}) + \dots$$

$$r_{34} = \tfrac{2}{3}\sqrt{6} - \epsilon\,\gamma\,(a_{35} + a_{45}) + \dots$$

$$r_{15} = 1 - \epsilon\,(a_{25} + a_{35} + a_{45}) + \dots$$

$$r_{25} = 1 + \epsilon\, a_{25} + \dots$$

$$r_{35} = 1 + \epsilon\, a_{35} + \dots$$

$$r_{45} = 1 + \epsilon\, a_{45} + \dots$$

where γ is a constant with the value $\gamma = (6912 + 1134\sqrt{6})/20607$.

At second order we find three quadratic bifurcation equations in a_{25}, a_{35}, and a_{45}. The quadratic equations factor easily into linear terms which deliver exactly four nontrivial solutions. They correspond to the four axis of symmetry of the regular tetrahedron. The solutions are

$$a_{25} = a_{35} = a_{45} = \alpha \quad , \qquad a_{25} = -a_{35} = a_{45} = \alpha$$

$$-a_{25} = a_{35} = a_{45} = \alpha \quad , \qquad a_{25} = a_{35} = -a_{45} = \alpha$$

with $\alpha = (151593992204\sqrt{6} - 217016509824) / 1339339327695$. For example, the first solution gives the following new central configuration to first order

$$r_{12} = r_{13} = r_{14} = \tfrac{2}{3}\sqrt{6} + 2\,\epsilon\,\alpha\,\gamma + \dots$$

$$r_{23} = r_{24} = r_{34} = \tfrac{2}{3}\sqrt{6} - 2\,\epsilon\,\alpha\,\gamma + \dots$$

$$r_{25} = r_{35} = r_{45} = 1 + \epsilon\,\alpha + \dots$$

$$r_{15} = 1 - 3\,\epsilon\,\alpha + \dots$$

The preservation of the symmetry is apparent and the shape of the central configuration which bifurcates from the regular tetrahedron is easily visualized.

Results about admissible central configurations for the five body problem, that is, those with positive masses can be found by adopting the methods which were described for the planar four body problem. The algebraic equations which have to be solved are in appearance identical to those in (4), except that we have to use the new definitions of Δ_i and F. The equations are

(9a) $m_i m_j (r_{ij}^{-3} - \delta) = \sigma \Delta_i \Delta_j \qquad 1 \le i < j \le 5$

(9b) $I - I_0 = 0$

(9c) $F = 0$

Instead of the single condition (5) there are now five distinct conditions of the same form

$$(r_{ij}^{-3} - \delta)(r_{kl}^{-3} - \delta) = (r_{ik}^{-3} - \delta)(r_{jl}^{-3} - \delta) = (r_{il}^{-3} - \delta)(r_{jk}^{-3} - \delta) \tag{10}$$

where the indices (i,j,k,l) are chosen from the set (1,2,3,4,5).

The conditions (10) are necessary and sufficient to solve (9a) for δ. There are 15 different ways of expressing δ, i.e.

$$\delta = \frac{s_{ij}\, s_{kl} - s_{ik}\, s_{jl}}{s_{ij} + s_{kl} - s_{ik} - s_{jl}}$$

From this and (9a) ten ratios for the different masses can be found.

$$\frac{m_i \Delta_j}{m_i \Delta_j} = \frac{s_{jk} - \delta}{s_{ik} - \delta} = \frac{s_{jk} - s_{jl}}{s_{ik} - s_{il}} \tag{11}$$

The analysis of (11) is now the same as for the four body problem.

The convex case is characterized by the following geometric fact. The signs of two oriented tetrahedra are different from those of the other three. In order to be specific let us say that

$$\Delta_1, \Delta_2, \Delta_3 > 0 \quad \text{and} \quad \Delta_4, \Delta_5 < 0 .$$

Furthermore, there exists exactly one interior diagonal and in our case it is r_{45} (see Fig 7). This diagonal intersects the interior of a triangle which here is composed of r_{12}, r_{23}, and r_{13}.

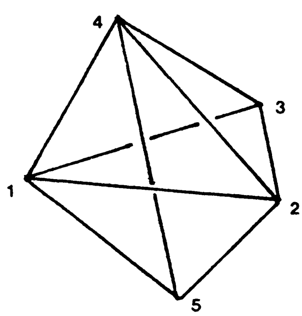


Fig. 7) Convex case in R^3

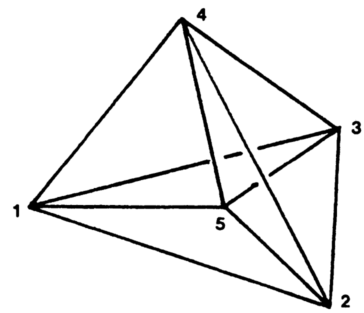


Fig. 8) Concave case in R^3

THEOREM: $r_{45}, r_{12}, r_{23}, r_{13} > \rho > r_{14}, r_{24}, r_{34}, r_{15}, r_{25}, r_{35}$.
In addition assume an ordering of the sides of the triangle surrounding the interior diagonal, say $r_{12} > r_{23} > r_{13}$, *then*

$$r_{34} > r_{14} > r_{24} \quad \textit{and} \quad r_{35} > r_{15} > r_{25} .$$

The distances are pair wise always longer or always smaller, that is,

$$r_{j4} > r_{j5} \quad \textit{for } j = 1,2,3 \quad \textit{or} \quad r_{j4} < r_{j5} \quad \textit{for } j = 1,2,3.$$

The theorem gives an ordering of the edges which is reminiscent to that of the concave case in the four body problem. The longest edge emanating from P_4 (and from P_5) has to be opposite the longest side of the triangle which is cut by the interior diagonal. Although the theorem does not give a more restrictive condition on the length of the interior diagonal besides $r_{45}>\rho$ the last part of the theorem gives it implicitly. If one of the points P_4 or P_5 is given it specifies a region in which the other point can lie.

The concave case is characterized by the fact that one mass lies inside the tetrahedron which is formed by the other four masses. Another way of expressing this is to say that the volume of just one oriented tetrahedron has a sign which is different from the signs of the other four.

In accordance with Fig 8) we assume that $\Delta_1,\Delta_2,\Delta_3,\Delta_4 > 0$ and $\Delta_5 > 0$.

THEOREM: *All interior edges* $\le \rho \le$ *all exterior edges.*
Furthermore, if we assume an ordering for the interior edges say

$$r_{15} < r_{25} < r_{35} < r_{45}$$

then

$$r_{14} < r_{13} < r_{12} \quad , \quad r_{34} < r_{23} < r_{13}$$
$$r_{24} < r_{23} < r_{12} \quad , \quad r_{34} < r_{24} < r_{14} .$$

This set of inequalities provides only a partial ordering for the exterior edges because the relationship between $r_{14}$and r_{23} is missing. This partial ordering is $r_{15} < r_{25} < \{r_{35},r_{15}\} < r_{25} < r_{35}$ but it is easier to interpret the inequalities as they are stated in the theorem.

Each inequality gives the ordering of the three exterior edges that start at an exterior vertex. It relates them to the length of the interior edges which end at the other three exterior vertices. It is an inverse relationship among the two sets of three edges. The longest exterior edge connects to the shortest interior edge, etc. From this it follows that not all tetrahedra can be completed to form a central configuration by finding a location for the fifth mass in its interior. The inequalities of the theorem require that the center of the circumscribing sphere of the tetrahedron has to lie inside this tetrahedron.

We conclude with the presentation of a symmetric central configuration. Three equal masses are placed at the vertices of an equilateral triangle. The other two masses are equidistant to these vertices. We model the situation after the four body problem and set

$$r_{12} = r_{23} = r_{13} = \sqrt{3}\,\beta$$
$$r_{14} = r_{24} = r_{34} = \sqrt{(\alpha+1)^2 + \beta^2}$$
$$r_{15} = r_{25} = r_{35} = \sqrt{\alpha^2 + \beta^2}$$
$$r_{45} = 1$$
$$\Delta_1 = \Delta_2 = \Delta_3 = \sqrt{3}\,\beta^2/12$$
$$\Delta_4 = \sqrt{3}\;\alpha\,\beta^2/4$$
$$\Delta_5 = -\sqrt{3}(1 + \alpha)\,\beta^2/4$$

The ratios of the masses are then

$$\frac{m_4}{m_1} = 3\,\alpha\,\frac{(3\beta^2)^{-3/2} - (\alpha^2+\beta^2)^{-3/2}}{((\alpha+1)^2+\beta^2)^{-3/2} - 1}$$

$$\frac{m_5}{m_1} = 3\,(\alpha+1)\,\frac{(3\beta^2)^{-3/2} - ((\alpha+1)^2+\beta^2)^{-3/2}}{1 - (\alpha^2+\beta^2)^{-3/2}}$$

The region in the α - β plane where both masses are positive can be displayed in a diagram which will look very similar to Fig 3). The only difference is that the slopes of the lines where m_4 and m_5 are zero changes from $\sqrt{3}/3$ to $\sqrt{2}/2$.

Point a) in Fig 3) remains at $\alpha = -1/2$, $\beta = \sqrt{3}/2$. It represents two tetrahedra of height 1/2 and length of the base triangle 3/2 butted against each other. The ratio of the masses becomes infinite if this configuration is to be realized. Point b) is now at $\alpha = -1/2$, $\beta = \sqrt{2}/4$ due to the different slope. It corresponds to two regular tetrahedra with a common face, but the value of the masses has to be zero. Finally point c) is at $\alpha = 1/3$ and $\beta = 2\sqrt{2}/3$. It is a concave configuration with all outside edges of length $2\sqrt{6}/3$. The interior mass is completely arbitrary. It is the case which leads to the bifurcation analysis given earlier.

Bibliography

Arenstorf, R.F., 1982: "Central Configurations of Four Bodies with one Inferior Mass", Celest. Mech. 28, 9-15.

Brumberg, V.A., 1958: "Permanent Configurations in the Problem of Four Bodies and their Stability", Soviet Astron. 34, 57-79

Dziobek, O., 1900: "Über einen merkwürdigen Fall des Vielkörperproblems", Astron. Nach. 152, 32-46.

Klemperer, W.B., 1962: "Some Properties of Rosette Configurations of Gravitating Bodies in Homographic Equilibrium", Astron. J. 67, 162-167

Lindow, M., 1926: "Der Kreisfall im Problem der n+1 Körper", Astron. Nachr. 228, 233-248.

Longley, W.R., 1907: "Some Particular Solutions in the Problem of n Bodies", Bull. Am. Math. Soc,, 13, 324-335

MacMillan, W.D. and Bartky, W. 1932: "Permanent Configurations in the Problem of Four Bodies", Trans. Am. Math. Soc., 34, 838-875.

Meyer, K.R. and Schmidt, D.S.,
1987a: "Bifurcations of Relative Equilibria in the 4 and 5 Body Problem", to appear
1987b: "Bifurcations of Relative Equilibria in the N-Body Problem and Kirchhoff problems", to appear.

Palmore, J.I.,
1973: "Relative Equilibria of the n-Body Problem", Thesis, Univ. Calif., Berkley.
1976: "Measure of Degenerate Relative Equilibria I", Ann. Math. 140, 421-29.

Pedersen, P., 1944: "Librationspunkte im restringierten Vierkörperproblem", Dan. Mat. Fys. Medd. 21, 6, 1-80.

Simó, C., 1977: "Relative Equilibrium Solutions in the Four Body Problem", Celest. Mech. 18, 165-84.

Williams, W.L., 1938: "Permanent Configurations in the Problem of Five Bodies", rans. Am. Soc. 44, 563-579.

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45221

http://dx.doi.org/10.1090/conm/081/986258

Contemporary Mathematics
Volume 81, 1988

STABLE MANIFOLDS IN HAMILTONIAN SYSTEMS

Clark Robinson[1]

ABSTRACT. In this paper several applications of stable manifold theory are given to Hamiltonian systems. The examples are of such a nature so that there is an invariant submanifold which is normally hyperbolic. The goal is to explain what assumptions are necessary in this situation for there to exist stable manifolds of the whole submanifold and of orbits in the submanifold. We also state some general theorems for the existence and differentiability of stable manifolds of such a submanifold. Sometimes these theorems can be used directly, but other times they merely form the model which motivates the type of result which is to be expected. In the latter type of problem, the proof of the stable manifold theory is used and not the actual theorems themselves. In each of the examples, a preliminary choice of coordinates is necessary before the theorems can be applied. In one case this is done by using so called McGehee coordinates; in another, the method of higher order averaging removes a troublesome angular dependence. In any case, some preliminary work is usually needed before the stable manifold theorem or theory can be applied. By understanding the underlying ideas of this theory for invariant submanifolds the reader should be able to determine other situations where it applies.

1. INTRODUCTION. Stable manifolds have been used on a variety of problems in Hamiltonian systems and for a variety of purposes. These submanifolds organize the behavior of nearby orbits to cause the existence of horseshoes (chaos), oscillation, or stability of various sorts. An example of this use is the existence of oscillatory orbits for the three body problem as shown by Sitnikov, [32]. Also see [1]. The treatment of Moser emphasizes the connection with stable manifolds, [22]. This work uses the paper by McGehee, [19], which proves the necessary stable manifold result for a degenerately hyperbolic fixed point. In Section 4 below, we discuss the corresponding stable manifold result for partially parabolic orbits for the planar three body problem where one particle goes to infinity and the other two remain bounded. This latter problem has more degrees of freedom so we need the stable manifold of a degenerately hyperbolic invariant submanifold, a three sphere.

1980 *Mathematics Subject Classification.* (1985 *Revision*). 58F05, 58F15, 70F07.
[1] Supported by the National Science Foundation.

A second problem where stable manifolds have been used to organize nearby orbits is the passage near triple collision. McGehee studied these orbits in [20] for the collinear 3-body problem, and later together with Mather proved the existence of escape and capture orbits, [18]. Also see the paper by Moeckel in these proceedings, [21], for a description of many of this type of problem. In section 3 we discuss completely parabolic orbits for the three body problem, which is closely connected with this set up and is mathematically very similar to total collapse. Using the existence of a stable manifold of a two-sphere together with a linearization result, infinite rotation about the limiting axis of motion is shown to be possible for completely parabolic motion. This contrasts with the impossibility for infinite rotation for total collapse.

A second use of stable manifolds is to separate the behavior of orbits on the two sides, hence the name separatrix for stable manifolds in the plane. In both the types of parabolic orbits mentioned above, they separate the orbits where the motion remains bounded from the orbits where the particles reach infinity with positive velocity, hyperbolic orbits. Another higher dimensional example, where the stable manifolds separate phase space and confine other orbits, are certain problems involving capture in resonance. Here the manifolds separate the orbits which are captured in resonance from those which pass through resonance. We do not discuss this type of example in this paper. See [24] or [25]. Also see [10] or [30].

A third use of stable manifold theory, and the last one discussed here, is to prove the structure of the orbits which are asymptotic to different limiting orbits as time goes to infinity. For example, in the planar partially parabolic case discussed above and in Section 4, the map, which assigns to a parabolic orbit the limiting motion of the binary as t goes to infinity, can be shown to be a smooth map on the set of all parabolic orbits. To control the orbits which are nearly parabolic, this natural foliation of the parabolic orbits needs to be extended to nearly parabolic orbits to be a continuous foliation that is invariant. This extension is not intrinsically defined, but does show that the change in eccentricity of the binary is a negligible amount no matter how close the motion come to parabolic motion. See [26, Theorem E and Section 4] for the precise statement and use of this foliation.

Such foliations are also used to study attractors. Williams has studied hyperbolic attractors by collapsing stable manifolds to points and forming a branched manifold. See [33], [34], or [35]. To be able to form these equivalence classes and proceed with the investigation, it is necessary to show that the stable manifolds of points in the attractor vary differentiably as the point varies. See [14, Theorem 6.5].

The two examples to be treated in this paper are given in sections 3 and 4: completely parabolic orbits and partially parabolic orbits. Section 2 introduces the ideas and some more specific results from stable manifold theory. The particular emphasis is on the existence and differentiability of a stable manifold of an invariant submanifold.

I would also like to state that this paper emphasizes the use of the stable manifold theory and makes no attempt to give the definitive result. Also, some references have been given to make a connection with the existing literature but they are not exhaustive by any means. I have also often referred to my papers on the particular problem where others have made more fundamental contributions.

2. STABLE MANIFOLD THEORY FOR INVARIANT SUBMANIFOLDS.

The simplest case of stable manifold theory is for a fixed point with eigenvalues with nonzero real part:

$$\dot{x} = ax + O(x^2 + y^2)$$
$$\dot{y} = -by + O(x^2 + y^2)$$

with $a, b > 0$. Here and throughout the paper $O(r)$ means terms of order r and higher. These remainder terms are assumed to be differentiable. For this type of system there is an invariant curve for the nonlinear equations, $W^s(0)$, tangent to the y-axis of orbits which go to the origin as $t \to \infty$. Similarly $W^u(0)$ is a curve tangent to the x-axis of orbits which go to 0 as $t \to -\infty$.

In this paper we are interested in examples where there is an invariant set, S, which is more complicated than a single fixed point: often an invariant manifold. We are interested in showing that the set of orbits which are asymptotic to the S as $t \to \infty$, $W^s(S)$, is a smooth manifold. Also we are interested in the subsets of points which are asymptotic to the orbit of a particular point, $W^s(p) = \{q : d(\varphi^t(p), \varphi^t(q)) \to 0$ as $t \to \infty\}$. Here $d(\cdot, \cdot)$ is the distance between points. When the set is hyperbolic in the correct sense it is possible to show that the orbits which are asymptotic to S are in phase with some orbit within S, so $W^s(S) = \cup\{W^s(p) : p \in S\}$. Also under the correct assumptions the map $\pi : W^s(S) \to S$, which satisfies $\pi W^s(p) = p$, is a differentiable map.

We will proceed to make these results more precise. We denote the norm of a linear map L by $||L|| = \sup\{|Lv| : |v| = 1\}$. The minimum norm of a linear map L is defined as $m(L) = \inf\{|Lv| : |v| = 1\}$, and when L is invertible this equals $||L^{-1}||^{-1}$. The minimum norm measures the least amount any vector is stretched by the linear map in the same way that the norm measures the maximum stretch.

Assume that the phase space is a manifold M and $\varphi^t : M \to M$ is a C^r flow. Assume that S is a compact submanifold without boundary that is invariant, $\varphi^t(S) = S$ for all t. For the examples, S is often a circle, torus, or sphere. The submanifold is said to be *normally hyperbolic* if the tangent bundle of M, restricted to S, splits into three continuous subbundles

$$TM|S = N^u \oplus TS \oplus N^s$$

each of which is invariant by the derivative of φ^t (first variation of the flow), $D\varphi^t$, and such that

(a) $D\varphi^T$ expands N^u more sharply than it expands anything in TS, i.e., for each p in S there is a $T = T(p)$ such that $m(D\varphi^T(p)|N^u_p) > \max\{||D\varphi^T(p)|T_pS||, 1\}$,

(b) $D\varphi^T$ contracts N^s more sharply than it contracts anything in TS, i.e., for each p in S there is a $T = T(p)$ such that $||D\varphi^T(p)|N^s_p|| < \min\{m(D\varphi^T(p)|T_pS), 1\}$. Note that by a change of metric on M to a so called adapted metric, the time $T = T(p)$ can be taken uniformly to be one in the definition. See [15], [8], or [31].

The submanifold S is said to be *r-normally hyperbolic* if φ^t is C^r and for all p in S and $0 \leq k \leq r$

(a) $$m(D\varphi^T(p)|N^u_p) > ||D\varphi^T(p)|T_pS||^k$$

(b) $$||D\varphi^T(p)|N^s_p|| < m(D\varphi^T(p)|T_pS)^k.$$

Notice that the assumptions of normal hyperbolicity are not pointwise conditions on the differential equations but depend on the derivative of the differential equation along the orbit which determines $D\varphi^T$ by the first variation equation. Therefore they can not be verified by eigenvalue conditions. ($D\varphi^T(p)$ is usually a linear map between different subspaces.) See the example of a closed orbit in [11,p. 121].

We can now state the theorem.

THEOREM 2.1. Assume S is a compact r-normally hyperbolic submanifold of M without boundary for a C^r flow φ^t on M, with $r \geq 1$. Then

(i) there exists a C^r locally φ^t-invariant submanifolds $W^u(S)$ and $W^s(S)$ tangent at S to $N^u \oplus TS$ and $TS \oplus N^s$ respectively.

(ii) $W^s(S)$ consists of all points whose forward φ^t orbit stay near S for all $t \geq 0$. Similarly, $W^u(S)$ consist of all points whose backward φ^t orbit stays near S for all $t \leq 0$.

(iii) $W^s(S) = \cup\{W^s(p) : p \in S\}$ where each $W^s(p)$ is a C^r manifold and the map, $\pi : W^s(S) \to S$ given by $\pi(W^s(p)) = p$, is C^r. Points q of $W^s(p)$ are characterized by the fact that the distance from $\varphi^t(q)$ to $\varphi^t(p)$ goes to zero as $t \to \infty$ at an asymptotic rate of at least that of $\|D\varphi^t(p)|N_p^s\|$.

(iv) Similarly, $W^u(S) = \cup\{W^u(p) : p \in S\}$, and $\pi : W^u(S) \to S$ is C^r where $\pi(W^u(p)) = p$. Also for q in $W^u(p)$, the distance from $\varphi^t(q)$ to $\varphi^t(p)$ goes to zero as $t \to -\infty$ at an asymptotic rate of at least that of $\|D\varphi^t(p)|N_p^u\|$.

This theorem has appeared many times in various generalizations and with different conclusions as to the differentiability. For a proof see [15, Theorem 4.1]. For a slight variation see [8, Theorem 6], [7], and [29]. Also see Hale, [12].

There are various approaches to proving stable manifold theorems. Basically, the proof either uses the graph transform idea of Hadamard, the variation of parameters with boundary conditions of Peron, or the method of partial differential equations of Sacker. We follow the first of these methods. Also see [15] and [14]. For the second method see [17], [16], [2]. For the third method see [29]. Also see [13] for other references.

In the approach of [15], the key step to prove the differentiability is a C^r-section theorem. The idea is that once the manifold has been shown to exist as a Lipschitz manifold then the derivatives of the manifold (foliation, or other object studied) can be viewed as elements of a bundle over the manifold. Each fiber of the bundle is a space of linear maps of possible derivatives, or trial derivatives. By taking a possible derivative and taking it forward by the derivative of the flow we get an induced flow on the bundle of trial derivatives. The C^r-section theorem gives conditions for there to be an invariant section of such a bundle and for this section to be differentiable. With some technical work and care, this section can be shown to be the derivatives of the invariant manifold. For a proof see [14], [15], or [31].

THEOREM 2.2 (C^r-section Theorem). Let $\pi : E \to X$ be a vector bundle over a metric space X with a norm on the fibers. Let $X_0 \subset X$ be a subset and D be a disk bundle of radius C in E, where C is a finite constant. Let $D_0 = D \cap \pi^{-1}(X_0)$ be the restriction of D to X_0. Let $F : D_0 \to D$ be a $C^{r+\alpha}$ map for $r \geq 0$ and $0 \leq \alpha < 1$ which covers $h : X_0 \to X$, i.e., the C^r derivative is α-Holder (if $\alpha = 0$ then disregard this aspect). Assume h is an overflowing $C^{r+\alpha}$ homeomorphism, $h(X_0) \supset X_0$. We assume the derivatives of F and h are uniformly bounded for all degrees $\leq r$, and the Holder constant is uniform. Suppose that F contracts on each fiber, i.e., there exists a constant k, $0 < k < 1$, such that for each x in X_0, the restriction of F to the fiber over x is Lipschitz with constant at most k, $\mathrm{Lip}(F|D_x) \leq k$. (If F is differentiable, then this can be expressed as a bound on the derivative on each fiber, $\|DF(p_x)|D_x\| \leq k$.) Assume $\mathrm{Lip}(h^{-1}) \leq \mu$, $k\mu^{r+\alpha} < 1$, and $k\mu^j < 1$ for $0 \leq j \leq r$. Then there is a unique section $\sigma : X_0 \to D_0$ such that

$$F(\text{image } \sigma) \cap D_0 = \text{ image } \sigma$$

and σ is $C^{r+\alpha}$.

In other words, if there is a fiber contraction, then there exists a continuous invariant section. In fact, if σ is any section, then $\sigma_1 = \Gamma(\sigma) \equiv F \circ \sigma \circ h^{-1}$ is its graph transform, and $\Gamma^n(\sigma)$ converges to the invariant section as n goes to infinity. To get the existence of derivatives the possible negative effect of h^{-1} must be overcome by a larger fiber contraction. The logic behind this is that if σ is a differentiable section then the derivative of its graph transform is given by

$$D\sigma_1(x) = DF(\sigma \circ h^{-1}(x))D\sigma(h^{-1}(x))Dh^{-1}(x).$$

Thus the term $Dh^{-1}(x)$ has a possible negative effect on the derivative of the transformed section. Similarly

$$D^r\sigma_1(x) = DF(\sigma \circ h^{-1}(x))D^r\sigma(h^{-1}(x))[Dh^{-1}(x)]^r + \cdots$$

where the unspecified terms involve lower order derivatives of σ. Thus the contracting effect of DF, k, must overcome the possible negative effect of $[Dh^{-1}(x)]^r$, μ^r. With the assumptions of the theorem, if σ is a differentiable section then $\Gamma(\sigma)$ and all its derivatives converge, so they converge to the invariant section and its derivatives.

Another way to understand the negative effect of h^{-1} is that if h brings close together two points which were farther apart then this makes the derivative of σ_1 larger. The constant $\mathrm{Lip}(h^{-1})$ measures the extent that h brings together points which are farther apart.

In the next three sections, we use these ideas to examine three specific examples of differential equations.

3. COMPLETELY PARABOLIC ORBITS The first example is completely parabolic motion for three bodies where the mutual distances between all the particles go to infinity and the velocities all go to zero. This example is easier than the others because the invariant manifold is made up of all fixed points for the equations, so the normal hyperbolicity can be calculated by means of eigenvalues. (Similar type analysis applies to triple collision, see [3] or [4] for a similar treatment.)

Let $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ where $\mathbf{q}_j$ is the location of the jth particle in R^3. Thus $\mathbf{q}$ is an element of R^9. We assume the center of mass of the system is fixed at the origin which reduces the dimension of the configuration space to six and phase space, which includes the velocities, to twelve dimensions. Let $\mathbf{M} =$ diagonal$(m_1 I_3, m_2 I_3, m_3 I_3)$ be the 9×9 matrix of the masses, where I_3 is the 3×3 identity matrix, and m_j is the mass of $\mathbf{q}_j$. Let $V(\mathbf{q})$ be the Newtonian potential energy which is negative and has terms proportional to the inverse of the mutual distance between each pair of particles. The equations of motion are

$$\dot{\mathbf{q}} = \mathbf{M}^{-1}\mathbf{p}, \qquad \dot{\mathbf{p}} = -\partial V/\partial \mathbf{q}.$$

To study parabolic motion we introduce McGehee coordinates which divides the distance by the square root of the moment of inertia. See [3], [4], or [20]. Let $\rho = (\mathbf{q}^t\mathbf{M}\mathbf{q})^{-1/2}$. The scaled configuration is $\mathbf{s} = \rho\mathbf{q}$, where $\mathbf{s}$ lies on the ellipsoid $\mathbf{s}^t\mathbf{M}\mathbf{s} = 1$. The velocity is decomposed into the radial component along $\mathbf{s}$ and the component tangent to the ellipsoid, and then each component is scaled by $\rho^{-1/2}$:

$$v = \rho^{-1/2}\mathbf{s}^t\mathbf{p}, \qquad \mathbf{u} = \rho^{-1/2}\mathbf{M}^{-1}\mathbf{p} - v\mathbf{s}.$$

With this scaling, there is a common factor of $\rho^{-3/2}$ in each term of the differential equation, so making the change of time scale $d\tau/dt = \rho^{3/2}$ and using $(')$ for $d/d\tau$, the equations become

$$\begin{aligned}
\rho' &= -v\rho \\
v' &= \mathbf{u}^t\mathbf{M}\mathbf{u} + (1/2)v^2 + V(\mathbf{s}) \\
\mathbf{s}' &= \mathbf{u} \\
\mathbf{u}' &= -(1/2)v\mathbf{u} - (\mathbf{u}^t\mathbf{M}\mathbf{u})\mathbf{s} - V(\mathbf{s})\mathbf{s} - \mathbf{M}^{-1}\nabla V(\mathbf{s}).
\end{aligned} \tag{3.1}$$

The energy relation is $h/\rho = (1/2)\mathbf{u}^t\mathbf{M}\mathbf{u} + (1/2)v^2 + V(\mathbf{s})$, where $h = 0$ for parabolic orbits. The equations extend naturally to $\rho = 0$ to give the motion at infinity. See [4] for more details.

For parabolic orbits, we get that $v' = (1/2)\mathbf{u}^t\mathbf{M}\mathbf{u}$ on the energy surface $h = 0$. Thus $v(\tau)$ is nondecreasing and the flow on the manifold at infinity is gradient-like. If $v(\tau)$ is unbounded, then $V(\mathbf{s})$ must go to minus infinity by the energy relation, so the upper bound on the distance between one binary pair must go to zero in the scaled variables. In his classification of motion [28], Saari showed that this could not occur for completely parabolic orbits, even in the scaled variables. In fact, if $V(\mathbf{s})$ goes to zero, this corresponds to the binary staying a bounded distance apart and the distance to the third body growing at the rate corresponding to hyperbolic motion. For the planar isosceles problem discussed in [3], it is easily seen that the binary would have to undergo repeated binary collisions and so could not correspond to parabolic motion. Thus, v must remain bounded and so $\mathbf{u}(\tau)$ must go to zero.

Since v is monotone and stays bounded, its limit set is contained in a single level set of v. Also $\mathbf{u}$ must be identically equal to zero in the limit set and so $\mathbf{u}'$ must also equal zero. Looking at the $\mathbf{u}'$ equation, we need $-V(\mathbf{s})\mathbf{s} - \mathbf{M}^{-1}\nabla V(\mathbf{s})$ to go to zero. Solutions, $\mathbf{s}_0$, of this equation are the central configurations which are either collinear or an equilateral triangle. The value of v at the fixed point is determined by the energy relation, $v^2 = -2V(\mathbf{s})$, or for $v > 0$, $v_0 = (-2V(\mathbf{s}_0))^{1/2}$. Thus the trajectory must approach the set where $\rho = 0$, $\mathbf{u} = 0$, $\mathbf{s}_0$ where $\mathbf{s}_0$ is parallel to $\mathbf{M}^{-1}V(\mathbf{s}_0)$, and $v_0 = \pm(-2V(\mathbf{s}_0))^{1/2}$. These are just the set of fixed points.

Because the invariant manifold we are considering in this problem is made up completely of fixed points the normal hyperbolicity conditions can be determined by eigenvalues. Let D be the linearization of equations (3.1) at the fixed point, and B be the submatrix from the linearization of $\mathbf{M}^{-1}V(\mathbf{s}) + V(\mathbf{s})\mathbf{s}$. A direct calculation shows that if μ is an eigenvalue of B and $\{\lambda^+, \lambda^-\}$ are the corresponding eigenvalues of D, then

$$\lambda^+, \lambda^- = -(1/4)v_0 \pm (1/4)v_0(1 - 16\mu v_0^{-2})^{1/2}.$$

See [4, p. 238].

Notice that the original equations are invariant by an action of $SO(3)$, so each fixed point lies on a 2-sphere of fixed points. Thus B has zero as an eigenvalue with multiplicity two, $\mu_{-1} = \mu_0 = 0$. The corresponding eigenvalues for D are $\lambda_{-1}^+ = \lambda_0^+ = 0$ and $\lambda_{-1}^- = \lambda_0^- = -v_0/2$.

For the rest of the section we will only consider the orbits which are asymptotic to the collinear central configuration and ignore the equilateral case. The two eigenvalues of B, μ_{-2}, μ_{-3}, corresponding to perturbations along the axis are negative for all choices of the masses. The two eigenvalues of B, associated with perturbations off the axis and orthogonal to the $SO(3)$ action, are positive. Thus the corresponding eigenvalues of D satisfy the following:

$\lambda_j^+ > 0$,	$\lambda_j^- < -v_0/2$	$j = -3, -2$	along the axis
$\lambda_j^+ = 0$,	$\lambda_j^- = -v_0/2$	$j = -1, 0$	$SO(3)$ directions
$-v_0/2 <$	$\mathrm{Re}(\lambda_j^{+-}) < 0$	$j = 1, 2$	off the axis.

Applying the theorems of section 2 yields the following Theorem. Notice that the stable manifold of particular points on the sphere differ by an action of $SO(3)$ because the equations are invariant by this action.

THEOREM 3.2. For the three body problem, the set of completely parabolic orbits which are asymptotic to a collinear configuration form 10 dimensional C^∞ manifold of extended phase space (as given by equations (3.1).) The set of solutions of (3.1) which are asymptotic to a particular point on the sphere of such configurations also forms a C^∞ manifold and the assignment of the point on the sphere (the limiting axis of the collinear configuration) is a C^∞ map. In fact, the stable manifolds of particular points differ by the action of $SO(3)$, $AW^s(p) = W^s(Ap)$. The rates of convergence are determined by the eigenvalues. In fact if the motion is not collinear, then the distance of the unscaled coordinates to the limiting axis does not go to zero but grows as t or τ goes to infinity.

For parabolic orbits in the stable manifold, the scaled variables which measure directions off the axis approaches the equilibrium at a rate which is at most $\exp(-\tau v_0/2)$. However, $\rho \sim \exp(-\tau v(0))$, so these unscaled directions grow at a rate which is at least $\exp(\tau v(0)/2)$.

The eigenvalues can be explicitly calculated for the isosceles 3-body problem. This restricted phase space is obtained by assuming $m_1 = m_3$, and looking at the invariant set of configurations which is an isosceles triangle with axis of the triangle along the x-axis. For this system, the eigenvalues at the collinear central configuration are

$$\zeta^{+-} = -2^{-3/2}\{1 \pm [(4m_3 - 55m_1)/(m_1 + 4m_3)]^{1/2}\}$$

each with multiplicity two, since we are still working with the spatial problem. See [3, pp. 259–260] where he determines these eigenvalues on the collision manifold. Also compare [27]. Note that these eigenvalues are complex whenever $m_3/m_2 > 4/55$. If the motion is not coplanar then it has two oscillatory modes in perpendicular planes. Thus the motion of the parabolic orbit will have an infinite rotation about the axis as τ or t goes to infinity. The rate of change of the angle about this axis with respect to τ has a nonzero limit determined by the imaginary part of the eigenvalue. A calculation shows that this rate of change with respect to t goes to zero as it must because of the conservation of angular momentum since the distance from the axis is unbounded. This infinite rotation in the completely parabolic motion contrasts with the case of total collapse where such infinite rotation is not possible. See [27] for further discussion.

4. PARTIALLY PARABOLIC ORBITS In this section, we consider the motion of three particles is the plane, with positions $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$, and with masses m_1, m_2, and m_3. The distance between the first two particles is to remain bounded and the distance to the third particle is to become unbounded as t goes to infinity (or minus infinity). We want to consider the situation where the third particle has just enough energy to reach infinity, so we assume that its velocity goes to zero. An orbit is called *ω-partially parabolic* or just *ω-parabolic* (resp. α-parabolic) if as t goes to infinity (resp. minus infinity) $\mathbf{r}_2 - \mathbf{r}_1$ remain bounded, $\mathbf{r}_3 - \mathbf{r}_1$ goes to infinity, and the velocity of $\mathbf{r}_3$ goes to zero. These parabolic orbits separate the orbits of the system in which the third particle remains bounded and those in which it reaches infinity with non-zero velocity.

There are several change of coordinates to get the equations into the form in which to apply the stable manifold theory. See [5] for more details on these change of coordinates. After using the angular momentum integral to determine one variable, and dropping the angular variable which measures the direction of m_3 from the center of mass, we are left with 6 real variables, or 2 real and 2 complex variables. The first real variable, x, is defined so that x^{-2} is proportional to the distance from the center of mass to m_3. The real variable y is proportional to radial component of the momentum of m_3. Using a complex variable to denote the position in the plane and complex multiplication, z is the position of the binary measured relative to the axis formed by the third mass:

$$\mathbf{r}_2 - \mathbf{r}_1 = z(\mathbf{r}_3/|\mathbf{r}_3|).$$

The final complex variable, w, is determined by the momentum vector of the binary relative to the unit vector formed from $\mathbf{r}_3$. Given these variables, the Levi-Cevita regularization removes the singularity caused by collisions of the binary:

$$z = 2\mu\xi^2, \qquad w = \mu\gamma^{-1}\eta\bar{\xi}^{-1}$$

where $\mu = m_1 + m_2$ and $\gamma^2 = \xi\bar{\xi} + \eta\bar{\eta}$. Then letting $K = 4\gamma\mu\xi\bar{\xi}$ and rescaling time by multiplying the vector field by K the equations become

$$\begin{aligned} x' &= -Kx^3y \\ y' &= -K[x^4 + O(x^6)] \\ \xi' &= \eta + O(x^4) \\ \eta' &= -\xi + O(x^4) \end{aligned}$$

with energy integral $h = H = (1/2)k_1^{-1}\beta^2y^2 - k_2(\xi\bar{\xi} + \eta\bar{\eta})^{-1} + O(x^2)$, where k_1, k_2, and β are constants determined by the masses. These equations extend naturally to $x = 0$, where, for $h < 0$, $H = h$ defines a three sphere for each y (and in particular for $y = 0$). These are the same as equations [5, 1.8 on p. 123]. The next change of coordinates in [5] should have been unitary, so there is a square root missing. There are also several typographical errors in the energy formulas on the top of [5, p. 124], which were corrected by the author of that paper making them compatible with the above.

The motion of the system for m_3 at infinity is given by the motion on the subset with $x = 0$. Those orbits which correspond to parabolic orbits at infinity would have $y = 0$ in addition. Fixing $x = 0$, $y = 0$, and a total negative energy $h < 0$, the motion that results is the Hopf flow on the three sphere: $\xi' = \eta$, $\eta' = -\xi$ with $|\xi^2| + |\eta^2| = 1$. The ω-parabolic orbits are those which are asymptotic to this S^3, i.e., to $W^s(S^3)$, and the set of α-parabolic orbits is $W^u(S^3)$. The motion normal to S^3 is to lowest order terms

$$\begin{pmatrix} x \\ y \end{pmatrix}' = Kx^3 \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y. \end{pmatrix}$$

This motion normal to the invariant manifold is attracting and repelling but the rates are not determined by the linear terms but only occur at the fourth order terms in x and y. Therefore we can not apply the usual stable manifold theorems are discussed in section 2. However the time T-map of motion on the S^3 is an isometry so these rates of hyperbolicity are stronger to any order than any rates within the invariant submanifold, so it is reasonable to hope that the stable and unstable manifolds still exist.

Indeed this is the case, the manifolds $W^s(S^3)$ and $W^u(S^3)$ are C^∞ manifolds. Moreover, they are smoothly foliated by the orbits asymptotic to a single orbits in the three sphere. Let $\pi : S^3 \to S^2$ be the map which assigns to an orbit of the Hopf flow in the three sphere a point in the two sphere; let $\alpha(p)$ be the set of α-limit points of p and $\omega(p)$ be the set of ω-limit points; and finally let

$$\omega^* = \pi\omega : W^s(S^3) \to S^2 \qquad \text{and}$$
$$\alpha^* = \pi\alpha : W^u(S^3) \to S^2$$

be the maps which assign the limiting motion of the binary as t goes to $\pm\infty$. (Giving the point in S^2 is the same as specifying the motion of the binary by the eccentricity and the axis of the ellipse of motion.)

THEOREM 4.1. The manifolds $W^s(S^3)$ and $W^u(S^3)$ are C^∞ submanifolds. Moreover the maps

$$\omega^* = \pi\omega : W^s(S^3) \to S^2 \qquad \text{and}$$
$$\alpha^* = \pi\alpha : W^u(S^3) \to S^2$$

are C^∞.

This theorem is proved in [26], and [5] contains a very similar result. Earlier, McGehee had proved a similar result where S^3 is replaced by a point [19]. As discussed in section 1, this latter theorem is used by Moser to prove the Sitnikov result on the existence of oscillatory, escape, and capture orbits, [22].

There are several ways to approach the difficulty caused by the lower order hyperbolicity. One way is to divide through in the equations by a factor of x^3 (actually a change of time scale) which results in the equations

$$x' = -Ky$$
$$y' = -K[x + O(x^3)]$$
$$\xi' = \eta x^{-3} + O(x)$$
$$\eta' = -\xi x^{-3} + O(x).$$

The first two equations are now hyperbolic. (Although K does vanish for $\xi = 0$, the integral over a positive time interval gives a positive value.) However the last two equations have a singularity at $x = 0$. Although S^3 is not a product of S^2 and S^1, the analytical difficulty can be understood by writing the equations as if they were a product with z in S^2 and θ in S^1. The equations then would be of the form

$$\theta' = x^{-3} + O(x)$$
$$z' = O(x).$$

These equations have a sheer in the θ direction. Points which are far apart on the trial unstable manifold can be brought close together by the graph transform. However the equations are well behaved in the z-direction. The difficulty can be overcome by averaging out the dependence in the equations on the θ variable by a C^∞ change of coordinates. Then the graph transform can be used for points with $x \neq 0$ and extended to $x = 0$ in the obvious manner. This process works to show that both the whole manifold of ω-parabolic orbits and the foliation by orbits asymptotic to a particular limiting motion of the binary are C^∞.

In an appendix to this section we give the details for the averaging argument since the treatment in the original paper is somewhat confused. Also a slightly more general set of equations than those above are given for which the result is true.

Another aspect of this example, which is different than the previous one of completely parabolic motion, is that the calculation of the hyperbolicity is not a matter of calculating eigenvalues at fixed points. In fact the calculations are along closed orbits of the flow in S^3. However, the fact that the flow is almost a constant one in the normal direction to S^3 minimizes the difficulty of verifying the assumptions in this case.

The original motivation for this problem came from the attempt by Easton and McGehee, [6] to show the existence of oscillatory motion for the planar three body problem. Also see [26] and [5]. In addition to showing that these manifolds are C^∞, there are several transversality conditions which need to be verified. So far no one has shown that the necessary map has zero as a regular value. However, Quillen has verified numerically that two of the necessary coordinates behave correctly, [23]. He also realized that the best way to set up the problem is to perturb way from $m_1 = m_2 = 0$ and not from $m_3 = 0$. This keeps the problem as a perturbation away from a known solution but the integrals of motion do not change as much with the correct choice. As observed in [26] and verified by this numerical work, the correct bi-parabolic orbit to consider is the one where at one instant all three particles pass through a common line (the axis of the binary) in a perpendicular direction, and the third particle has just the right velocity to be a ω-parabolic orbit. By the symmetry of the equation, this particular orbit will also be α-parabolic. The work of Quillen partially verifies the transversality conditions necessary to show that the symbolic dynamics can be used to prove the existence of the oscillatory orbits.

4A. APPENDIX: AVERAGING OF THE PARABOLIC EQUATIONS The type of equations treated are of the form

$$\dot{\theta} = 1 + p(x)g(\theta, z, x)$$
$$\dot{z} = p(x)h(\theta, z, x)$$
$$\dot{x} = p(x)f(\theta, z, x).$$

In these equations, x is a vector variable. Later in the section we split x into the contracting and expanding coordinates. Also, θ is an angular variable, z is a variable in a compact manifold so that the set $x = 0$ is a compact manifold, $p(x)$ is a homogeneous polynomial of degree $p > 0$, and f, g, and h are of degree at least one in x (not necessarily homogeneous.) It is important in the proof that the same function $p(x)$ occur in all the equations in the lowest order terms (except for the 1 in the $\dot{\theta}$ equation). The following result says that we can average out the dependence on θ to all orders, leaving dependence on θ only in terms which are C^∞ flat. Note that the form of the equations is much simpler than given in [26].

PROPOSITION 4A.1. There is a C^∞ change of coordinates such that (keeping the same letters for the new coordinates) the equations become

$$\dot{\theta} = 1 + p(x)\{g^{av}(z, x) + G(\theta, z, x)\}$$
$$\dot{z} = p(x)\{h^{av}(z, x) + H(\theta, z, x)\}$$
$$\dot{x} = p(x)\{f^{av}(z, x) + F(\theta, z, x)\}$$

where G, H, and F are C^∞-flat as a function of θ, i.e., they are all $O(|x|^k)$ for all k.

PROOF. Averaging, by means of a change of coordinates which is of a homogeneous degree in x, can only remove the θ dependence from the terms of this single degree in x. Therefore, we proceed by induction on the degree in x. Assume that the averaging has been carried out to terms of order j in terms of x, and the resulting equations are of the form

$$\dot{\theta} = 1 + p(x)g^{av}_{j-1}(z, x) + p(x)g_j(\theta, z, x)$$
$$\dot{z} = p(x)h^{av}_{j-1}(z, x) + p(x)h_j(\theta, z, x)$$
$$\dot{x} = p(x)f^{av}_{j-1}(z, x) + p(x)f_j(\theta, z, x)$$

where f^{av}_{j-1}, g^{av}_{j-1}, and h^{av}_{j-1} are all made up of terms of degree less than or equal $j-1$ in x and are independent of θ, and f_j, g_j, and h_j are all made up of terms of degree greater than or equal to j.

To remove the dependence on θ in the terms of degree j, we make the change of variables

$$\theta = \psi + p(\xi)u(\psi,\zeta,\xi)$$
$$z = \zeta + p(\xi)v(\psi,\zeta,\xi)$$
$$x = \xi + p(\xi)w(\psi,\zeta,\xi)$$

where u, v, and w are of degree j in ξ. By differentiating these equations and setting them equal to the original differential equations with the substitutions made, we get the following equality

$$\begin{pmatrix} 1 + p(\xi + pw)g(\psi + pu, \zeta + pv, \xi + pw) \\ p(\xi + pw)h(\psi + pu, \zeta + pv, \xi + pw) \\ p(\xi + pw)f(\psi + pu, \zeta + pv, \xi + pw) \end{pmatrix}$$
$$= \begin{pmatrix} 1 + pu_\psi & pu_\zeta & U_\xi \\ pv_\psi & I + pv_\zeta & V_\xi \\ pw_\psi & pw_\zeta & I + W_\xi \end{pmatrix} \begin{pmatrix} \psi' \\ \zeta' \\ \xi' \end{pmatrix}$$

where everything is evaluated at (ψ,ζ,ξ), $U = pu$, $V = pv$, and $W = pw$, and $g = g_{j-1} + g_j$, $h = h_{j-1} + h_j$, and $f = f_{j-1} + f_j$.

Note that in first side of this equality we have the term $p(\xi + p(\xi)w(\psi,\zeta,\xi))$, but this equals

$$p(\xi) + \int_0^1 p'(\xi + sp(\xi)w(\psi,\zeta,\xi))p(\xi)w(\psi,\zeta,\xi)ds$$
$$= p(\xi)\{1 + fn(\psi,\zeta,\xi)\},$$

where fn is of degree at least j in ξ. Therefore

$$p(\xi + p(\xi)w(\psi,\zeta,\xi))g(\psi + pu, \zeta + pv, \xi + pw)$$
$$= p(\xi)g_{j-1}^{av}(\zeta,\xi) + p(\xi)g_j(\psi,\zeta,\xi) + p(\xi)O(\xi^{j+1}).$$

Note that this means that the same function p is a factor of the equation. This is the point which was not done correctly in [26]. Similar expressions are true for the terms ph and pf. We use these in the first side of the above equality.

To solve for the derivatives ψ', ζ', and ξ' we need to invert its coefficient matrix using the formula that $(I + A)^{-1} = I - A + \Sigma_{n=2}^{\infty}(-1)^n A^n$:

$$\begin{pmatrix} 1 + pu_\psi & pu_\zeta & U_\xi \\ pv_\psi & I + pv_\zeta & V\xi \\ pw_\psi & pw_\zeta & I + W_\xi \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} 1 - pu_\psi & -pu_\zeta & -U_\xi \\ -pv_\psi & I - pv_\zeta & -V_\xi \\ -pw_\psi & -pw_\zeta & I - W_\xi \end{pmatrix} + O(\xi^{2(p+j-1)})).$$

(Remember that $p(\xi)$ is of degree p and u, v, w are of degree j in ξ.)

Combining the above and multiplying the two matrices we get

$$\begin{pmatrix} \dot{\psi} \\ \dot{\zeta} \\ \dot{\xi} \end{pmatrix} = \begin{pmatrix} 1pg_{j-1}^{av} + p[g_j - u_\psi] \\ ph_{j-1}^{av} + p[h_j - v_\psi] \\ pf_{j-1}^{av} + [f_j - w_\psi] \end{pmatrix} + p(\xi)O(\xi^{j+1}).$$

Then we can solve $g_j - u_\psi$, $h_j - v_\psi$, and $f_j - w_\psi$ for u, v, and w so the difference are terms of degree at least $j+1$ in ξ. This completes the induction step of the proof of the proposition.

Next by a change of time scale, we can divide the differential equations by $p(x)$ and take the time one map of the flow to get a map which is C^∞ flat in θ and hyperbolic in x. At this point we need to split the x coordinates (or really the x and y coordinates of Section 4) into (X, Y) where the X coordinates are expanded and the Y coordinates are contracted by the lowest order terms. In terms of the equations of Section 4, $x = 2^{-1/2}(X+Y)$ and $y = 2^{-1/2}(-X+Y)$. Then the time one map (of the rescaled equations) become

$$\begin{aligned} \theta(1) &= \theta + q_1(z, X, Y) + R_1(\theta, z, X, Y) \\ z(1) &= z + q_2(z, X, Y) + R_2(\theta, z, X, Y) \\ X(1) &= A(z)X + q_3(z, X, Y) + R_3(\theta, z, X, Y) \\ Y(1) &= B(z)Y + q_4(z, X, Y) + R_4(\theta, z, X, Y) \end{aligned}$$

where $(A(z)X, B(z)Y)$ is the linear term as a function of X and Y resulting from the linear terms of f^{av} with z $m(A) > 1$ and $\|B\| < 1$, the R_j terms are C^∞-flat in X and Y, and the derivatives of the functions q_j satisfying the following conditions as $|(X, Y)|$ goes to zero:

$$\left| \frac{\partial^{i+j+k} q_1}{\partial z^i \partial X^j \partial Y^k}(z, X, Y) \right| \leq C(i,j,k)|(X, Y)|^{-p-j-k}$$

$Dq_2(z, X, Y)$ has a bounded limit $Dq_2(z, 0, 0)$ with

$$\begin{aligned} &\partial q_2/\partial z(z, 0, 0) = 0 \\ &|Dq_i(z, X, Y)| = O(|(X, Y)|) \qquad \text{for } i = 3 \text{ and } 4. \end{aligned}$$

As explained in [26, p. 368], it is now possible to show that these equations have a Lipschitz unstable manifold. This argument also shows that it is real analytic for $X \neq 0$ by showing that the graph transform preserves functions with specified radii of convergence (which depends on X). The manifold is not always real analytic at $X = 0$, see [19]. This much of the result is shown in [5] without averaging the equations. To show that it is C^∞ at $X = 0$, we need to show that the graph transform preserves sections sufficiently flat in θ, as given by the following lemma, [26, Claim 3.6].

CLAIM 4A.2. For all $j \geq p+3$, the graph transform, Γ, preserves sections with

$$
\begin{aligned}
&|w_4(\theta,z,X)| \leq |X|, \\
&\left|\frac{\partial w_4(\theta,z,X)}{\partial\theta}\right| \leq C_j|X|^j, \\
&\left|\frac{\partial w_4(\theta,z,X)}{\partial z}\right| \leq C_z|X|, \qquad \left|\frac{\partial w_r(\theta,z,X)}{\partial X}\right| \leq C_X|X|
\end{aligned}
$$

for constants C_j, C_z, and C_X chosen sufficiently large.

PROOF. The following calculations corrects the typographical errors contained in [26]. The notation $O(k)$ is used to mean $O(|(X,Y)|^k)$, and $O(\infty)$ to mean $O(k)$ for all k. We also write the equations as if X and Y were scalars, $A(z) = \lambda$, and $B(z) = \mu$. Actually, similar estimates hold in higher dimensions where $m(A(z)) \geq \lambda$ and $||B(z)|| \leq \mu$.

Assume the estimates hold for w and let $v = \Gamma(w)$. The proof that v satisfies the first condition is standard and we will skip. Then differentiating the graph transform gives

$$Dv_4(u) = DF_4(m_{-1})Dw(u_{-1})[DF_0(m_{-1})Dw(u_{-1})]^{-1}$$

where $u = (\theta,z,X)$, $u_{-1} = (F_0 \circ w)^{-1}(u)$, and $m_{-1} = w(u_{-1})$. Then evaluating the terms at the correct points as indicated above and using bounds which are independent of w unless indicated,

$$Dw(u_{-1}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ C_j\lambda^{-j}|X|^j & C_z\lambda^{-1}|X| & C_X\lambda^{-1}|X| \end{pmatrix}$$

$$DF_0Dw = \begin{pmatrix} 1+O(\infty) & O(-p) & O(-p-1) & O(-p-1) \\ 0 & 1+O(1) & O(0) & O(0) \\ 0 & 0(1) & \lambda+O(1) & O(1) \end{pmatrix} Dw$$

$$= \begin{pmatrix} 1+O(2) & O(-p) & O(-p-1) \\ O(j) & 1+O(1) & O(0) \\ O(j+1) & O(1) & \lambda+O(1) \end{pmatrix},$$

$$(DF_0Dw)^{-1} = \begin{pmatrix} 1+O(2) & O(-p) & O(-p-1) \\ O(j) & 1+O(1) & O(0) \\ O(j+1) & O(1) & \lambda^{-1}+O(1) \end{pmatrix}$$

$$Dw[DF_0Dw]^{-1} =$$

$$\begin{pmatrix} 1+O(2) & O(-p) & O(-p-1) \\ O(j) & 1+O(1) & O(0) \\ O(j+1) & O(1) & \lambda^{-1}+O(1) \\ C_j\lambda^{-j}|X|^j+O(j+1) & C_z\lambda^{-1}|X|+O(2) & C_X\lambda^{-2}|X|+C_zO(1) \end{pmatrix}$$

$$DF_4 = (O(\infty), \quad O(1), \quad O(1), \quad \mu+O(1),)$$

$$Dv_4 = \big(C_j\lambda^{-j}\mu|X|^j+O(j+1), \quad C_z\lambda^{-1}\mu|X|+O(1), \\ C_X\lambda^{-2}\mu|X|+C_zO(1)+O(1)\big).$$

Therefore, by taking C_j, C_z, and C_X large enough the bound on the derivatives of v_4 are satisfied. (Remember that terms like $O(1) = O(|(X,Y)|)$ in the second component have bounds which are independent of the bounds on the sections, and since $|v_4| \le |X|$ this term is also $O(|X|)$.)

The next claim is used to show that Γ preserves sections with uniform bounds on the derivatives. The exponent on X is assumed to be higher on partial derivatives involving θ because these terms get multiplied by terms of order $O(-p)$ or $O(-p-1)$ which appear in $[DF_0Dw]^{-1}$.

CLAIM 4A.3 ([26, 3.7]). Γ preserves C^{i+j+k} sections with derivatives bounded by

$$\left|\frac{\partial^{i+j+k}w_4}{\partial\theta^i\partial z^j\partial X^k}\right| \le C(J,i,j,k)|X|^{iJ}$$

for $J \ge \max(n, p+1+i+j+k)$ and for suitably chosen $C(n,i,j,k)$, as long as $|\partial w/\partial\theta| = O(J)$, $|\partial w/\partial z| = O(1)$, and $|\partial w/\partial X| = O(1)$.

In the proof the in [26] it is noted that

$$\begin{aligned} D^k v_4 &= DF_4 D^k w[DF_0 Dw\}^{-k} \\ &\quad - DF_4 Dw[DF_0 Dw]^{-1} DF_0 D^k w[DF_0 Dw]^{-k} \\ &\quad + \text{ terms in } D^j w \text{ with } j < k. \end{aligned}$$

The first term is a contraction, the second term is shown to be smaller than the first term, and the terms in the lower derivatives of w are already bounded. To be more careful, the term that involve the derivatives of F_1 should be checked, since they have some terms which are unbounded. However, they are acted on by terms of the lower derivatives of v_4, so are of order $O(1)$ and are $O(\infty)$ in the terms which involve a derivative of F_1 with respect to θ. Then, this derivative acts on terms involving some $D^m w[DF_0 Dw]^{-m}$ which gives term which are at least $O(1)$ and $O(iJ)$ if the term involves i derivatives with respect to θ. The result follows as discussed in [26].

Using the above two claims, it follows that if w^0 is a section that is C^∞ flat in θ, then so is $w^k = \Gamma(w^{k-1}) = \Gamma^k(w^0)$. Therefore this forms an equicontinuous family. Since w^k converges C^0 uniformly to some w^*, it follows that w^* is C^∞. This completes the proof of the result for equations in the form given at the beginning of this section. For the modifications for the equations given in Section 4 see [26].

REFERENCES

1. V. M. Alekseev, "Quasirandom dynamical systems I, II, III," Math. USSR-Sbor., 5 (1968), pp. 73–128; 6 (1968), pp. 505–560; 7 (1969) pp. 1–43.
2. E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
3. R. Devaney, "Triple collision in the planar isosceles three body problem," Invent. Math. 60 (1980), pp. 249–267.
4. R. Devaney, "Singularities in classical mechanical systems," in Ergodic Theory and Dynamical Systems I, ed. A. Katok, Progress in Math. 10 (1981), Birkhauser, Boston/Basel/Stuttgart, pp. 211–333.
5. R. Easton, "Parabolic orbits for the planar three body problem," J. Diff. Equat. 52 (1984), pp. 116–134.
6. R. Easton and R. McGehee, "Homoclinic phenomena for orbits doubley asymptotic to an invariant three-sphere," Indiana Univ. Math. J. 28 (1979), pp. 211–240.
7. N. Fenichel, "Persistence and smoothness of invariant manifolds for flows," Indiana Univ. Math. J. 21 (1971), pp. 193–226.
8. N. Fenichel, "Asymptotic stability with rate conditions," Indiana Univ. Math. J. 23 (1974), pp. 1109–1137.
9. N. Fenichel, "Asymptotic stability with rate conditions II," Indian Univ. Math. J. 26 (1977), pp. 81–93.
10. R. Haberman, "Energy bounds for the slow capture by a center in sustained resonance," SIAM J. Appl. Math. 43 (1983), pp. 244–256.
11. J. Hale, *Ordinary Differential Equations*, Wiley, 1969.
12. J. Hale, "Integral manifolds of perturbed differential systems," Annals of Math. 73 (1961), pp. 496–531.
13. P. Hartman, *Ordinary Differential Equations*, Wiley, 1964.
14. M. Hirsch and C. Pugh, "Stable manifolds and hyperbolic sets," in Proc. Symposium in Pure Math. No. 14 (1970), Amer. Math. Soc., Providence R. I., pp. 133–164.
15. M. Hirsch, C. Pugh, and M. Shub, *Invariant Manifolds*, Lecture Notes in Math. No. 583 (1977), Springer-Verlag, New York/Berlin/Heidelberg.
16. M. Irwin, "On the stable manifold theorem," J. London Math. Soc. 2 (1970), pp. 196–198.
17. A. Kelly, "The stable, ceter stable, center, center unstable and unstable manifolds," J. Diff. Equat. 3 (1967), pp. 546–570.
18. J. Mather and R. McGehee, "Solutions of the collinear four-body problem which become unbounded in finite time," in *Dynamical Systems: Theory and Applications* (ed. J. Moser), Lecture Notes in Physics 38 (1975), Springer-Verlag, New York/Berlin/ Heidelberg, pp. 573–597.
19. R. McGehee, " A stable manifold theorem for degenerate fixed points with applications to celestial mechanics," J. Diff. Equat. 14 (1973), pp. 70–88.
20. McGehee, "Triple collision in the collinear three body problem", Invent. Math. 27 (1874), pp. 191–227.
21. R. Moeckel, "Survey on the N-body problem," these Proceedings.
22. J. Moser, *Stable and random motions in dynamical systems*, in Annals of Math. Studies No 77 (1973), Princeton Univ. Press, Princeton, N.J.
23. J. Quillen, "An application of the Melnikov method to the planar problem," thesis Univ. Colorado, 1986.

24. C. Robinson, "Sustained resonance for a nonlinear system with slowly varying coefficients," SIAM J. Math. Anal. 14 (1983), pp. 847–860.
25. C. Robinson, "Capture in resonance: opening a homoclinic orbit through slowly varying coefficients," in *Geometric Dynamics* (ed. J. Palis), Lecture Notes in Math 1007 (1983), Springer-Verlag, New York/Berlin/Heidelberg, pp. 651–662.
26. C. Robinson, "Homoclinic orbits and oscillation for the planar three-body problem," J. Diff. Equat. 52 (1984), pp. 356–377.
27. C. Robinson and D. Saari "N-body spatial parabolic orbits asymptotic to collinear central configurations," J. Diff. Equat. 48 (1983), pp. 434–459.
28. D. Saari, "Expanding gravitational systems," Trans. Amer. Math. Soc. 156 (1971), pp. 219–240.
29. R. Sacker, "A perturbation theorem for invariant Riemannian manifolds," Proc. Symp. at Univ. of Puerto Rico, edited by J. Hale & J. LaSalle, Academic Press, 1967, pp. 43–54.
30. J. Sanders, "On the passage through resonance," SIAM J. Math. Anal. 10 (1979), pp. 1220–1243.
31. M. Shub, *Global Stability of Dynamical Systems*, Springer-Verlag, New York/Berlin/Heidelberg, 1987. (Original French version, Asterisque 56 (1978).)
32. K. Sitnikov, "Existence of oscillating motions for the three body problem," Dokl. Akad. Nauk. USSR 133 (1960), pp. 303–306.
33. R. Williams, "One-dimensional nonwandering sets," Topology 6 (1967), pp. 473–487.
34. R. Williams, " Expanding attractors," Publ. Math. I.H.E.S. 43 (1974), pp. 169–203.
35. R. Williams, "The structure of Lorenz attractors," Publ. Math. I.H.E.S. 50 (1979), pp. 321–347.

DEPARTMENT OF MATHEMATICS
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986259

Contemporary Mathematics
Volume **81**, 1988

REDUCTION OF HAMILTONIAN SYSTEMS FOR SINGULAR VALUES OF MOMENTUM

Judith M. Arms

Abstract. Reduction for singular values of an equivariant momentum for a compact group action has been defined and studied by Arms, Gotay, and Jennings. The present paper builds on that work in the case of torus actions. Results are obtained on the lifting the reduced flow, on the equivalence of regularity and weak regularity, and on the variation of the dimension of the reduced space at singular values. An example shows that the first and last of these results fail to extend to the nonabelian case.

1. INTRODUCTION

The basic idea of reduction is to factor out the symmetries of a system, leaving a smaller system containing the essential information of the original. Specifically, suppose a Hamiltonian system is invariant under some group action, and there is an equivariant momentum mapping for the action. (A momentum mapping generalizes such ideas as integrals in involution and angular momentum. See §2 for details.) For any regular value of momentum, the orbit space of the level set for that value is called the reduced space. Any invariant Hamiltonian on the original phase space induces a Hamiltonian on the reduced space. There is a computable algorithm for lifting the reduced Hamiltonian flow back up to the flow of the original Hamiltonian. One can compute how the canonical structure on the reduced space varies with the value of the momentum; this provides a tool for studying perturbation questions.

The present paper generalizes some of these ideas to the case of singular values of the momentum for a torus action. Section 2 briefly reviews reduction in the regular (and weakly regular) case and §3 discusses the lifting process for the same. Reduction is defined (following [AGJ]) for the singular case in §4. Theorem 1 in §5 describes reduction and lifting of the reduced flow in terms of global invariants of the action. This description applies to both the regular and singular cases, and shows that the flow in the orbit directions is

1980 Mathematics Subject Classification (1985 Revision). 58F05

Partially supported by the National Science Foundation.

determined by how the Hamiltonian depends on the momentum. A discussion of Theorem 1 in §6 leads naturally to the question of how the reduced space varies with the value of the momentum. The latter has been studied in the regular case by Duistermaat and Heckman [DH]. By Theorem 2 in §7, if the momentum is weakly regular at all smooth points in a level set, then the action factors through the action of a smaller torus for which the momentum is regular at all smooth points. As a corollary, the Duistermaat and Heckman results generalize to the weakly regular case. For singular values that interpolate between (weakly) regular values, the reduced space is singular but, by Theorem 3 in §8, has the same dimension as the reduced spaces for nearby regular or weakly regular values. Finally, §9 gives an example to show that Theorems 1 and 3 do not generalize to the nonabelian case.

Acknowledgements. The present paper grew out of joint work of the author with Mark Gotay and George Jennings [AGJ], and insights garnered from conversations with these two coworkers are gratefully acknowledged, as are helpful conversations with Tom Duchamp.

2. REDUCTION IN THE REGULAR CASE.

Let us first review how the reduction process works in the case of regular (or weakly regular) values. The case where the group is abelian is classical; the generalization to the nonabelian case was done independently by Meyer [Me] and Marsden and Weinstein [MaW].

We begin by establishing some notation. Let (P, Ω) be a connected symplectic manifold. For any $f \in C^\infty(P)$, let X_f represent its Hamiltonian vector field. The Poisson bracket of two functions f and $g \in C^\infty(P)$ is given by $\{f, g\} := \Omega(X_f, X_g)$. Suppose G is a compact Lie group acting (on the left) on P by symplectomorphism. Let $\mathfrak{g}$ be the Lie algebra of G. For any $\xi \in \mathfrak{g}$, let ξ_P represent the generator of the action on P of the one parameter subgroup $\{e^{t\xi}\}$. We assume throughout that there exists an equivariant momentum map $\mu : P \to \mathfrak{g}^*$, where $\mathfrak{g}^*$ is the dual Lie algebra. That is, if μ is evaluated on ξ, the resulting function (μ, ξ) is a Hamiltonian for ξ_P; and for every $q \in P$ and $g \in G$, $\mu(g \cdot q) = g \cdot \mu(q)$. (Here the dot on the left hand side represents the given action of G on P and that on the right hand side represents the coadjoint action of G on $\mathfrak{g}$.) Let $V_\nu = \mu^{-1}(\nu)$; when $\nu = 0$, the subscript often will be omitted. We will call V_ν a constraint set because $\mu =$ constant often appears as a constraint on initial data for a Hamiltonian system. Let G_ν represent the isotropy group of ν under the coadjoint action. (For additional background information, see [AbM] or [AGJ].)

THEOREM. [Me, MaW; see also AbM, pp. 299–300, 304] *Let P, Ω, G, G_ν, μ, and V_ν be as described above. If ν is a (weakly) regular value of μ, then the reduced space $\widehat{V}_\nu := V_\nu / G_\nu$ has a unique symplectic form $\widehat{\omega}$ such that $\pi^*(\widehat{\omega}) = i^*(\Omega)$, where π is the canonical projection of V_ν onto $\widehat{V}_\nu$ and i is the inclusion of V_ν into P. Furthermore if*

$H \in C^\infty(P)$ is G-invariant, then the Hamiltonian flow of H induces a flow on $\widehat{V}_\nu$; and this flow is also Hamiltonian with Hamiltonian function $\widehat{H}$ defined by $\widehat{H} \circ \pi = H \circ i$.

Remark. The reduced space $\widehat{V}_\nu$ may fail to be a manifold because of finite isotropy groups for the action of G_ν on V_ν. However the singularities are not serious: it will still be a V-manifold, and will carry all the usual differentiable structures such as forms, vector fields, etc. (See also the discussion of this matter in [DH].) Another way around this difficulty is to work locally, as discussed in the "important remark" on p. 300 of [AbM].

For simplicity, let us assume that $\nu = 0$; then $G_\nu = G$. This is no real loss of generality because there is a standard proceedure for identifying $\widehat{V}_\nu$ as the reduced space for the zero value of a momentum map on a larger space. (See [AGJ]); a similar construction appears in [GS, §26].) The results in [AGJ], which we will be using heavily below, assume $\nu = 0$; but most can be extended to $\nu \neq 0$ as mentioned above. Also when $\nu \neq 0$ is important below G will be abelian. Then the value of ν is completely irrelevant, for $\widetilde{\mu} = \mu - \nu$ is also an equivariant momentum for the G action.

3. LIFTING THE DYNAMICS.

Suppose the reduced Hamiltonian flow is known (i.e. that of $\widehat{H}$ on $\widehat{V}$). One wants an algorithm for lifting the dynamics back to V; that is, for reconstructing the flow of H on V. Such an algorithm is described on pp. 304–5 of [AbM]. Briefly, it goes as follows. Given a trajectory $\widehat{c}(t)$ on $\widehat{V}$, lift to any smooth $\widetilde{c}(t)$ such that $\pi \circ \widetilde{c} = \widehat{c}$. Then the trajectory of H is $c(t) = g(t) \cdot \widetilde{c}(t)$, and the problem is to determine $g(t)$. In the reference it is shown that dg/dt is determined entirely by data at $\widetilde{c}(t)$ (not at the as-yet-to-be determined $c(t)$).

By choosing a particularly nice coordinate system, a sort of "partial action-angle" system, we may describe the lifting process even more explicitly. Essentially the idea is to pull back canonical coordinates on the reduced space and supplement them by the momentum μ itself and coordinates along the orbits. Let $\mathcal{U}$ be a neighborhood of V. There exists a smooth projection $\psi : \mathcal{U} \to V$ that projects each V_ν on V. (There are various ways to construct this projection. For instance, one may use the gradient flow of the norm squared of the momentum, as in Kirwan [K], or the construction in §1 of Duistermaat and Heckman [DH]. The latter deals only with torus actions, but the construction generalizes.) Choose canonical local coordinates $(\widehat{x}, \widehat{y})$ on $\widehat{V}$ and define $x = \widehat{x} \circ \pi \circ \psi$ and $y = \widehat{y} \circ \pi \circ \psi$, where π is the projection onto the orbit space. Choose a coordinate system $\widetilde{\alpha}$ on G so that $\widetilde{\alpha}(\text{identity}) = 0$ and the k-th coordinate axis is a one parameter subgroup with generator ξ_k. Let φ be a cross section for π. Define coordinates α on $G \cdot u$ so that $\alpha(q) = \widetilde{\alpha}(g)$ if $g \cdot \varphi(\pi(q)) = q$. Now (x, y, μ, α) are local coordinates on $G \cdot u$. Furthermore $\Omega|_V$ contains terms $dx_j \wedge dy_j$ and $d\alpha_j$ enters into only in terms of the form $f \cdot d\alpha_j \wedge d\mu_j$. In fact if G is abelian φ and $\widetilde{\alpha}$ can be chosen so that $\Omega = dx_j \wedge dy_j + d\alpha_j \wedge d\mu_j$.

Now suppose $\widehat{c}(t) = (\widehat{x}(t), \widehat{y}(t))$. The lifted trajectory $c(t) = (x(t), y(t), \mu(t), \alpha(t))$, where of course $\mu(t) \equiv$ constant . For the other coordinates one easily computes

$$\frac{dx_j}{dt} = \frac{d\widehat{x}_j}{dt} = \frac{\partial H}{\partial y_j}, \quad \frac{dy_j}{dt} = \frac{d\widehat{y}_j}{dt} = -\frac{\partial H}{\partial x_j}, \quad \text{and}$$

$$\frac{d\alpha_j}{dt} = \sum_k [\{\alpha_j, x_k\}\frac{\partial H}{\partial x_k} + \{\alpha_j, y_k\}\frac{\partial H}{\partial y_k} + \{\alpha_j, \mu_k\}\frac{\partial H}{\partial \mu_k}]. \tag{1}$$

If G is abelian, the last equation simplifies to

$$\frac{d\alpha_j}{dt} = \frac{\partial H}{\partial \mu_k}. \tag{2}$$

In any case (1) shows that $d\alpha_j/dt$ is determined by the dependence of H on the coordinates on the reduced space and the momentum μ.

4. REDUCTION IN THE SINGULAR CASE.

Now consider the case when zero is a singular value of the momentum. It follows from the definition of the momentum that at singular points there is a nontrivial isotropy subgroup. The flow of any invariant Hamiltonian commutes with the group action, so every point in one trajectory of the Hamiltonian will have the same isotropy subgroup. The set of all point invariant under that subgroup is itself a symplectic submanifold (cf. [GS, Thm. 27.2]). Furthermore the points with exactly that subgroup as isotropy subgroup are weakly regular points for the momentum restricted to the submanifold (cf. Lemma 7 in [AMM]). Thus one way to deal with singular values is to restrict a priori to points invariant under a subgroup. This approach will suffice if one is only interested in the trajectory of one particular set of initial conditions. In most cases, however, this restricted view will not give the entire reduced space for a singular value (because the preimage of such a value includes both (weakly) regular and singular points, and/or singular points with varying isotropy groups), let alone any idea of how the reduced spaces for varying values of momentum fit together. It is natural to ask for a larger view, i.e. one without the a priori restriction suggested above. Besides the intrinsic interest of such a view, it should be helpful in various applications (e.g. studying perturbation or quantization questions).

A general definition of reduction for singular constraint sets V is given in [AGJ], using ideas of Sniatycki [Sn]. This reduction uses the intrinsic geometry of V, in particular the restriction of the symplectic form Ω to the (generalized) tangent space of V; in the reference it is called the "geometric reduction". When V is the zero set of the momentum for a compact group action, the description of the reduced space and its Poisson algebra may be simplified, as follows.

The reduced space is still $\widehat{V} = V/G$. (Note: this may not be true if G is not compact. An example is the skew action $(x, y) \mapsto (x + ty, y)$, [AGJ], Ex. 7.1]) The smooth functions on $\widehat{V}$ are the G-invariant Whitney smooth functions on V; these will be denoted by $\widehat{W^\infty}(\widehat{V})$. (Whitney smooth functions are the restrictions of smooth functions on the ambient space.)

It remains to define the Poisson bracket on $\widehat{W}^\infty(\widehat{V})$. Given $\widehat{f}$ and $\widehat{g} \in \widehat{W}^\infty(\widehat{V})$, let $f|_V$ and $g|_V$ be the corresponding functions in $W^\infty(V)$ = the Whitney smooth functions on V. Extend to G-invariant functions f and g in $C^\infty(P)$. (This is possible because G is compact, so we can average over the orbits.) Define

$$(3) \qquad [\widehat{f}, \widehat{g}] := \{f, g\}.$$

In Proposition 5.7 of [AGJ] it is shown that the bracket [,] is a well defined Poisson bracket on $\widehat{W}^\infty(\widehat{V})$ and that it agrees with the more generally defined reduction process mentioned above. Given any invariant Hamiltonian H on P, it induces a function $\widehat{H} \in \widehat{W}^\infty(\widehat{V})$. Then the Poisson bracket [,] in (3) is used in the usual way to compute the flow on $\widehat{V}$. That is, $d\widehat{f}/dt = [\widehat{f}, \widehat{H}]$ for any function $\widehat{f} \in \widehat{W}^\infty(\widehat{V})$.) As there are enough invariants to separate orbits in V (Prop. 5.5 of [AG], or Hilbert's invariant theory), this defines the reduced flow on $\widehat{W}$.

5. LIFTING AND INVARIANTS

To lift a single trajectory of the reduced flow, it suffices to restrict a priori to points of one symmetry type and apply the regular case as discussed at the beginning of §4. One would like also to have analogs of equations (1) and (2); that is, to know how the derivatives of the Hamiltonian affect the lifting. We can not use the components of the momentum as corrdinates now; instead, the role of the coordinates in equations (1-2) is played by invariants of the action. With mild restrictions on P, the essence of equation (2) holds for torus actions even when μ is singular; that is, the flow in the orbit directions is determined by the dependence of H on μ. (This need not be the case when G is nonabelian; see §9 below.)

THEOREM 1. *Suppose G is a k-dimensional torus and P has finitely generated homology. Then there exists a (finite) basis for the G- invariants $[C^\infty(P)]^G$ of the form*

$$\{\mu_1, \ldots, \mu_k, \lambda_1, \ldots, \lambda_\ell\}$$

so that for some $m \leq \ell$, $\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m\}$ is a basis for the reduced algebra $\widehat{W}^\infty(\widehat{V})$, where $V = \mu^{-1}(0)$. Thus any invariant Hamiltonian H is $H(\mu, \lambda)$, and the corresponding reduced Hamiltonian $\widehat{H}$ depends only on $\widehat{\lambda}_j, j = 1, \ldots, m$. If a trajectory $\widehat{c}(t)$ of $\widehat{H}$ is lifted to a trajectory $c(t) = g(t) \cdot \widetilde{c}(t)$ as described earlier, the λ_j may be chosen so that dg/dt is determined (at least on some neighborhood) by the partials of H with respect to μ.. (Thus if $m < \ell$, the dependence of H on $\lambda_j, j = m+1, \ldots, \ell$ is irrelevant to the dymanics on V.)

Proof. The fact that P has finitely generated homology implies that there are only finitely many orbit types for the action of G on P (cf. [B].) The following argument is found in Schwarz [Sch]. A theorem of Mostow-Palais [Mo, P] implies that P may be equivariantly embedded in an orthogonal representation space, say X, for G. Then the Hilbert invariant theory (see e.g. [Wy]) says that the invariant polynomials on X are finitely generated. Let $\sigma_1, \ldots, \sigma_s$ be the restriction of these generators to $P \hookrightarrow X$ (discarding any that are redundant after restriction). Then by Theorem 1 of Schwarz [Sch],

$$\sigma^*(C^\infty(\mathbb{R}^s)) = [C^\infty(P)]^G,$$

where $\sigma = (\sigma_1, \ldots, \sigma_s)$. That is, the σ_j are a basis for $[C^\infty(P)]^G$, not just for the invariant polynomials.

Furthermore we may choose $\sigma_j = \mu_j, j = 1, \ldots, k$, as follows. Without loss of generality we may assume that the action of G does not factor through the action of a smaller torus, that is, there is no $\xi \in \mathfrak{g}^*$ such that (μ, ξ) is constant. (If not, replace G by the smaller torus.) The μ_j are certainly invariant because G is abelian, and furthermore they are quadratic polynomials on X because G acts linearly on X. On the other hand they are not combinations of other invariants, as is immediately obvious in the local normal coordinates given in Theorem 4.1 of [AGJ]. (Briefly, in normal coordinates the isotropy group action looks like a subgroup of $U(n)$ acting on $\mathbb{C}^n$, and the "other" components of the momentum are linear.) Relabel the remaining σ_j as $\sigma_{k+j} = \lambda_j, j = 1, \ldots, \ell = s - k$.

Let $I(V)$ be the ideal of smooth functions vanishing on V, and $[I(V)]^G$ the ideal of G-invariants in $I(V)$. After possible recombination and reindexing, $[I(V)]^G$ is generated by the μ_j and possibly some of the λ_j, suppose the last few. Then

$$\widehat{W}^\infty(\widehat{V}) = [W^\infty(V)]^G = [C^\infty(P)/I(V)]^G = [C^\infty(P)]^G/[I(V)]^G \tag{4}$$

is generated by the remaining λ_j, say $j = 1, \ldots, m$. (The last equality above depends on the compactness of G; the proof is similar to the proof of Proposition 5.12 in [AGJ].)

Thus we have any invariant $H = H(\mu, \lambda)$, with the induced $\widehat{H} = \widehat{H}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m)$, as required. It remains to discuss the lifting, and for this we work locally. Choose q in the image of the lift $\widetilde{c}$ and let G_q be the identity component of the isotropy group of q. By Theorem 4.1 and the first paragraph of the proof of Theorem 6.8 in [AGJ], there are local canonical coordinates (x_j, y_j), $j = 1, \ldots, n$ with the following properties. For convenience, impose an almost complex structure locally by letting $z_j = x_j + iy_j$. (i) For some $p \leq n$, G_q acts on the local neighborhood identified as an open subset of $\mathbb{C}^n \approx \mathbb{C}^p \times \mathbb{C}^{n-p}$ by diagonal matrices on the first factor and trivially on the second. (ii) The remaining components of μ (those for G/G_q) are the $y_j, j = p+1, \ldots, n$. (iii) The $x_j, j = p+1, \ldots, n$. parameterize the G/G_q orbits. (Thus for points with isotropy group G_q, like those in the image of $\widetilde{c}$, these x_j parameterize the whole orbit.)

By (iii) the λ_j must be independent of the $x_j, j > p$. By (ii) the λ_j may be modified (by adding multiples of the μ_j and the other λ_j) to eliminate dependence on $y_j, j > p$. (This last may be done so as to preserve the separation of the λ_j which do and do not vanish on V.) It follows that we may choose the λ_j to depend only on $(x_j, y_j), j = 1, \ldots p$.

Thus if $h > p$,

$$\begin{aligned}\frac{dx_h}{dt} &= \{x_n, H\} \\ &= \sum_{j=1}^{k} \{x_h, \mu_j\} \frac{\partial H}{\partial \mu_j} + \sum_{j=1}^{\ell} \{x_h, \lambda_j\} \frac{\partial H}{\partial \lambda_j} \\ &= \sum_{j=1}^{k} \delta_{jh} \frac{\partial H}{\partial \mu_j},\end{aligned}$$

where $\delta_{jh} = 1$ if $j = h$ and μ_j is a component of the momentum for G/G_q, and $\delta_{jh} = 0$ otherwise.

Now $x = (x_{p+1}, \ldots, x_n)$ may be identified with as a point in G/G_q. The required $g(t)$ satisfies

$$\frac{dg}{dt} = \frac{d}{dt}[x(c(t)) - x(\widetilde{c}(t))] = \frac{dx}{dt} - \frac{d}{dt}[x(\widetilde{c}(t))]$$

By the proceeding equation, this shows that dg/dt is determined by the $\partial H/\partial \mu_j$ and is independent of the $\partial H/\partial \lambda_j$. ∎

6. DISCUSSION

The fact that dg/dt is independent of $\partial H/\partial \lambda_j$, $j = 1, \ldots, m$, seems reasonable because the $\widehat{\lambda}_j$'s parameterize the reduced space. The irrelevance of the $\partial H/\partial \lambda_j$, $j > m$ is somewhat harder to understand; in the proof, the reasons are buried in the choice of local normal coordinates.

In fact one can show the Hamiltonian vector fields $X_{\lambda_s}, s > m$, vanish on V. Let $I(\mu)$ be the ideal of smooth functions generated by the components of μ, and $[I(\mu)]^G$ the corresponding ideal of invariants. If $m < \ell$, then

$$[I(\mu)]^G \neq [I(V)]^G \tag{5}$$

(because $\lambda_s, s > m$, is the ideal on the right but not in that on left.) By Theorem 6.8 of [AGJ], inequality (5) may occur only if μ has a semi-definite component, e.g. $\sum_{j=1}^{h} \alpha_j(x_j^2 + y_j^2)$ for some h, $\alpha_j > 0$, in some local normal coordinate system. This implies that invariants must be either constant or quadratic or higher order with respect to the (x_j, y_j), $j \leq h$. It follows that we may restrict to the set $\widetilde{P} = \{x_j = 0 = y_j, j \leq h\}$, ignore the problem component of μ, and obtain the same dynamics. Repeating as necessary to eliminate all semi-definite components of μ, so (5) no longer holds. Thus the λ_s, $s > m$, have been eliminated. This implies they were quadratic in the (x_j, y_j), $j \leq h$, from which it follows that the X_{λ_s}, $s > m$, vanish on $V \subset \widetilde{P}$.

Sniatycki and Weinstein [SnW] have proposed an algebraic reduction which yields a reduced Poisson algebra $[C^\infty(P)/I(\mu)]^G$ but not neccessarily a reduced space. If G is compact,

$$[C^\infty(P)/I(\mu)]^G = [C^\infty(P)]^G/[I(\mu)]^G. \tag{6}$$

Comparing (4) and (6) shows that (5) holds if and only if the algebraic reduction does not give the same algebra as the (geometric) reduction we are using in the present paper.

Thus we could summerize the first part of §6 as follows. It seems reasonable that dg/dt is unaffected by the dependence of H on the λ_j when the geometric algebraic and reductions agree. When they do not agree, one might expect $dH/d\lambda_j$ to affect dg/dt for the λ_j which vanish on V. This does not happen in the torus case, however, because the Hamiltonian vector fields of such λ_j vanish on V.

Theorem 1 shows that in the torus case (5) has no effect on lifting the action from $\widehat{V}$ to V. Condition (5) may be important for other questions in dynamics, however. Consider, for instance, perturbation of initial conditions. Let $V_\nu = \mu^{-1}(\nu)$ and $\widehat{V}_\nu$ the corresponding reduced space, and consider to what extent the dynamics on $\widehat{V}_0$ determines the dynamics on nearby $\widehat{V}_\nu$. Suppose two invariant Hamiltonians differ by ΔH, $\Delta H = 0$ on $V = V_0$. When $I(\mu) = I(\nu)$, $\Delta H \in I(v)$ must be parameterized by the μ, i.e. by the invariant intergrals at the group action. When (5) holds, ΔH may not be so parametrized.

Investigation of perturbation questions requires a good picture of how the $\widehat{V}_\nu$ fit together. When ν varies through regular values, the V_ν and therefore the $\widehat{V}_\nu$ are all differmorphic. For P compact (and G a torus) Duistermaat and Heckman [DH] have obtained a formula for the variation of the cohomology of the symplectic form on the $\widehat{V}_\nu$ with respect to (regular) ν.

7. WEAK REGULARITY FOR TORUS ACTIONS.

In fact Duistermaat and Heckman's result may be extended to the weakly regular case. Recall that μ is said to be weakly regular at $q \in P$ if $\mu^{-1}(\mu(q))$ is a manifold at q whose tangent space is the kernel of $d\mu$.

THEOREM 2. *Suppose G is a torus and μ is weakly regular at each smooth point of a connected component C of one of its level sets. Then μ is regular at each smooth point of C if and only if the action of G (on any invariant neighborhood of C) does not factor throught the action of a smaller torus $\widetilde{G}$.*

Proof. It suffices to prove the proposition on a neighborhood of each point q of C, because C is connected. It further suffices to consider only the action and momentum of the isotropy group G_q of q, because the momentum for G/G_q is regular. Thus without loss of generality we work on a neighborhood of a fixed point of G.

There are canonical coordinates (x_j, y_j) centered at the fixed point in which G acts like a subgroup of $U(n)$ acting on $\mathbb{C}^n$ with coordinates $z_j = x_j + iy_j$. Essentially this follows from the compactness of G via the equivariant Darboux Lemma. See step 1 of the proof of Theorem 4.1 in [AGJ] for more details. (All theorems, etc., quoted in this proof are from [AGJ].)

A further canonical change of coordinates will diagonalize the action of G. Thus if $G = T^d = (S^1)^d$, the jth S^1 factor acts by a diagonal matrix with entries say $\exp(i\alpha_{jk}\theta_j)$, $k =$

$1, \ldots, n$. Let r = rank at the matrix $[\alpha_{jk}]$. If $r < d$, a reparameterization of G can be made so that $\alpha_{jk} = 0$ if $j > r$. (The algebra involved is essentially row reduction of the matrix $[\alpha_{jk}]$. See also the proof of Theorem 6.8.) Let $\widetilde{G} = T^r$ acting as dictated by the row–reduced $[\alpha_{jk}]$. Then the action of G factors through the action of $\widetilde{G}$. That is, there is a projection $\pi : G \to \widetilde{G}$ so that for each $q \in \mathbb{C}^n$, $\pi(g) \cdot q = g \cdot q$.

By Theorem 6.8, weak regularity at all smooth points implies that $I(\mu)$ is a real ideal, i.e. is equal to its own real radical. Let $I_p(\mu)$ be the ideal of polynomials (in the coordinates) generated by the components of μ. By Theorem 6.3, $I(\mu)$ real implies $I_p(\mu)$ real.

Now complexify the real coordinates x_j and y_j and analytically extend μ to the resulting $\mathbb{C}^{2n}$. Let $V_{\mathbb{C}} \subset \mathbb{C}^{2n}$ be the zero set of the extended M. By Theorem 6.5, the fact that $I_p(\mu)$ is real implies that

$$\dim_{\mathbb{R}}(V) = \dim_{\mathbb{C}}(V_{\mathbb{C}}), \tag{7}$$

where of course dimension is computed at smooth points.

From the form of μ given in equation (6.6) in [AGJ], it is easily computed that the (complex) codimension of $V_{\mathbb{C}}$ (at its smooth points) is r. By (7), the real codimension of V is also r. But because μ is weakly regular at smooth points of V, $\operatorname{codim}_{\mathbb{R}} V =$ rank $d\mu$. It follows immediately that μ is regular at the smooth points if and only if $r = d$, that is, $\widetilde{G} = G$. ■

COROLLARY. *In Duistermaat and Heckman [DH] the hypothesis of regularity may be replaced by weak regularity.*

Proof. If μ is only weakly regular, replace G by the $\widetilde{G}$ as above of minimum possible dimension. The momentum $\widetilde{\mu}$ for the $\widetilde{G}$ action will be regular, and the level sets, orbits, and reduced spaces will be the same as those for μ. ■

8. THE DIMENSION OF SINGULAR REDUCED SPACES.

For singular values which are not weakly regular, there is currently no general analog of the results above. Not only are the reduced spaces of singular values not diffeomorphic to those for nearby values (regular or singular), they may not even have the same dimension as those nearby! The most important singular values are those which interpolate between regular (or weakly regular) values, i.e. not those on the boundary of the image of μ. For P compact and G a torus, we can say something at least about the dimension of the singular reduced spaces for these interpolating values.

The convexity results of Guillemin and Sternberg and, independently, Atiyah say that if P is compact and connected and G is a torus, $\mu(P)$ is a convex polytope. (See [§32 of GS], and references therein.) As in the proof of the Corollary, project onto a $\widetilde{G}$ of minimal dimension. This corresponds to projecting $\mathfrak{g}^*$ down to a subspace which contains $\mu(P)$

and such that $\mu(P)$ has nonempty interior in the subspace topology. Call $\nu \in \mu(P)$ interior value of μ if it is in this (nonempty) interior.

THEOREM 3. *Suppose P is compact and connected and G is a torus. The reduced spaces for the interior values of μ all have the same dimension.*

Proof. As before, project G onto a minimal dimensional $\widetilde{G}$ with the same action. By Lemma 32.1 of [GS], each component μ^ξ of μ has a unique local maximum. But P is compact so this unique local maximum occurs on the boundary of $\mu(P)$. Thus singular points in the preimage of interior values for μ must be saddle points for all components μ^ξ; i.e. the nonpositivity condition (6.5) of [AGJ] holds for the level set V_ν of an interior value ν. By Theorem 6.8 of [AGJ] this implies that V_ν is weakly regular at its smooth points. By Theorem 2 and our initial assumption in this proof, V_ν is regular at its smooth points, and codim $V_\nu = \dim \widetilde{G}$. Furthermore the orbit dimension at such regular points is also $\dim \widetilde{G}$. Thus for each interior value ν, $\dim \widehat{V}_\nu = \dim P - 2\dim \widetilde{G}$. ■

Remark. The reduced spaces for boundary values will have smaller dimension than those for interior values. For instances, vertices of the polytope are global extrema for all components of μ, so their reduced spaces are points.

Theorem 3 may be combined with the arguments of Schwarz [Sch] to obtain the following nice picture. Because P is compact, it has finitely generated homology and may be equivariantly embedded in an orthogonal representation space for G, as in the proof of Theorem 1 above. Also as above, we may choose a basis for the G-invariants on P to include the components of μ, say $\{\mu_1, \dots, \mu_k, \lambda_1, \dots, \lambda_\ell\}$. Now by Schwarz [Sch], if $\sigma = (\mu, \lambda) : P \to \mathbb{R}^{k+\ell}$, $\sigma(P)$ is the orbit space of P. Define a projection $\pi : \mathbb{R}^{k+\ell} \to \mathbb{R}^k$. Then $\mu = \pi \circ \sigma$, and $\pi(\sigma(P))$ is the convex polytope $\mu(P)$. For each $\nu \in \mu(P)$, $\pi^{-1}(\nu) = \sigma(\mu^{-1}(\nu)) \approx \widehat{V}_\nu$, so the fibers of π are the reduced spaces. Some of these fibers may be singular. (Even for regular values there may be mild V-manifold singularities resulting from discrete isotropy groups; cf. discussion in [DH].) But by theorem 3, the fibers over interior values all have the same dimension. Thus one may hope to be able to say something about the cohomology of the interior singular fibers [Ar].

9. THE NONABELIAN CASE.

Now suppose G is nonabelian (but still compact.) As in the discussion §6, it seems reasonable to expect an analog of Theorem 1 when the components of μ and the coordinates on $\widehat{V}$ give a basis for the invariants; i.e., when

$$[I(\mu)]^G = [I(V)]^G. \tag{8}$$

It appears to be difficult to prove, however. The problem is to show that, in the notation of Theorem 1, the λ_j commute with the orbit variables.

If (8) fails, then we have the following counterexample. That is, it is not true in general that the lifting of the dynamics from the reduced space is determined by how the Hamiltonian depends on the reduced space coordinates and the momentum! The conclusions of Theorem 3 also fail for this example.

EXAMPLE. Let $SU(2)$ act on $\mathbb{C}^2$ in the canonical way. Regard $\mathbb{C}^2$ as $\mathbb{R}^4$ with the canonical metric. (Thus we have the $SU(2)$ action embedded in the canonical action of $SO(4)$.) Now lift the action to $T^*\mathbb{R}^4 = P$. The zero set V for the momentum $\mu = (\mu_1, \mu_2, \mu_3)$ for the $G = SU(2)$ action is $V = \{(X, Y) \in \mathbb{R}^4 \times \mathbb{R}^4 \approx T^*\mathbb{R}^4$: X and Y are proportional $\}$. (See Example 7.13 of [AGJ] for computation of μ, V, and the invariants discussed below.)

The momentum for the $SO(4)$ action has six components, the three μ's and three others, say f_1, f_2, and f_3. These f_j also vanish on V; but their Hamiltonian vector fields do not. Furthermore they are $SU(2)$ invariant. (Thus (8) fails for this example.) The f_j are invariant Hamiltonians which are independent of μ_j and give rise to the zero Hamiltonian on the reduced space, but nonetheless affect the lifted dynamics because they have nontrivial Hamiltonian vector fields on V. This example shows that the description in Theorem 1 of lifting the dynamics fails here.

The failure of (8) and the lifting result are particularly surprising in this example because V, though singular, is reasonable nice. At smooth points it is regular and (therefore) coisotropic. It has only one singular point, the origin, and the span of the tangent cone at the origin is also coisotropic.

The regularity at smooth points implies that zero is an interpolating value. But for $\nu \neq 0$, $G_\nu = S^1$, not G. Thus $\dim \widehat{V}_\nu = \dim[\mu^{-1}(\nu)/G_\nu] = \dim P - \dim G - \dim G_\nu = 4$; while $\dim \widehat{V}_0 = 2$. Thus Theorem 3 can not be generalized to the nonabelian case.

BIBLIOGRAPHY

[AbM] R. Abraham and J. E. Marsden, Foundations of Mechanics, *2nd ed.*, Benjamin Cummings, Reading, MA (1978).

[AGJ] J. M. Arms, M. J. Gotay, and G. Jennings, "Geometric and Algebraic Reduction for Singular Momentum Maps," to appear.

[AMM] J. M. Arms, J. E. Marsden, and V. Moncrief, "Symmetry and Bifurcations of Momentum Mappings," *Commun. Math. Phys.* **78**, 455–478 (1981).

[Ar] J. M. Arms, in preparation.

[B] G. E. Bredon, Introduction to Compact Transformation Groups, Academic Press, New York, p. 221 (1972).

[DH] J. J. Duistermaat and G. J. Heckman, "On the Variation in the Cohomology of the Symplectic Form of the Reduced Phase Space," *Invent. Math.* **69**, 259–268 (1982).

[GS] V. Guillemin and S. Sternberg, Symplectic Techniques in Physics, Cambridge University Press, Cambridge (1984).

[K] F. Kirwan, Cohomology of Quotients in Symplectic and Algebraic Geometry, Princeton University Press, Princeton (1984).

[MaW] J. E. Marsden and A. Weinstein, "Reduction of Symplectic Manifolds with Symmetry," *Rep. Math. Phys.* **5**, 121–130 (1974).

[Me] K. R. Meyer, "Symmetrics and Integrals in Mechanics," in Dynamical Systems (Proc. Sympos. Univ. Bahia, Salvador, 1971), ed. M. M. Pexioto, Academic Press, New York, 259–272 (1973).

[Mo] G. D. Mostow, "Equivariant Embeddings in Euclidean Space," *Ann. Math.* **65**, 432–446 (1957).

[P] R. S. Palais, "Imbedding of Compact Differentiable Transformation Groups in Orthogonal Representations," *J. Math. Mech.* **6**, 673–678 (1957).

[Sch] G. W. Schwarz, "Smooth Functions Invariant under the Action of a Compact Lie Group," *Topology* **14**, 63–68 (1975).

[Sn] J. Śniatycki, "Constraints and Quantization," Springer Lecture Notes in Mathematics **1037**, (1983).

[SnW] J. Śniatycki and A. Weinstein, " Reduction and Quantization for Singular Momentum Mappings," *Lett. Math. Phys.* **7**, 155–161 (1983).

[Wy] H. Weyl, The Classical Groups (2nd ed.), Princeton University Press, p. 274 (1946).

Department of Mathematics
University of Washington
Seattle, WA 98103

http://dx.doi.org/10.1090/conm/081/986260

Contemporary Mathematics
Volume **81**, 1988

A Variation on the Poincaré-Birkhoff Theorem

BY JOHN FRANKS

Abstract

In this article we give an exposition of the work of C. Conley on chain recurrence and complete Lyapounov functions and use it to prove the following special case of a theorem of P. Carter.
THEOREM. If $f : A \to A$ is a homeomorphism of the annulus which is homotopic to the identity and satisfies a boundary twist condition, then either f has at least one fixed point or there is a smoothly embedded essential curve C in A with $f(C) \cap C = \emptyset$.

The well known fixed point theorem of Poincaré and Birkhoff (see [B] or [BN]) asserts that an area preserving homeomorphism of an annulus which satisfies a boundary twist condition possesses at least two fixed points. In an attempt to replace the area preserving hypothesis with a more topological one, Birkhoff [B2] showed that if $f : A \to A$ satisfies a boundary twist condition and has at most one fixed point, then there is a "ring" $S \subset A$ whose boundary is one component of the boundary of A together with a continuum C (which separates A) such that $f(S)$ is a proper subset of S. This result was improved by P. Carter [Ca] who proved that under the same hypothesis one can find an essential simple closed curve C such that $f(C) \cap C$ consists of at most a single point and if it does consist of a single point, then that point is the unique fixed point of the homeomorphism. She also gives an example of such an f with exactly one fixed point and $f(C) \cap C$ consisting of a single point. Thus one cannot hope to prove that either f has at least 2 fixed points or there is a curve C with $f(C) \cap C = \emptyset$.

It follows immediately from Carter's theorem that either f has at least one fixed point or there is an essential simple closed curve C with $f(C) \cap C = \emptyset$. We give a new proof of this result in Theorem (2.4) below. The proof is elementary and relatively easy compared to the work of Carter. It makes use of the theory of chain recurrent sets and complete Lyapounov functions developed by C. Conley [C]. Since Conley's results are not widely known and his presentation is for flows, we give a brief exposition in §1 of the basic results of this theory in the setting of homeomorphisms.

§1. CHAIN RECURRENCE AND COMPLETE LYAPOUNOV FUNCTIONS

In this section we briefly review the elementary theory of attractor-repeller pairs and chain recurrence developed by Charles Conley in [C]. In the following $f : X \to X$ will denote a homeomorphism of a compact metric space X.

(1.1) **Definition.** An *ε-chain* for f is a sequence $x_1, x_2, \ldots, x_n$ of points in X such that

$$d(f(x_i), x_{i+1}) < \varepsilon \qquad \text{for } 1 \leq i \leq n-1.$$

A point $x \in X$ is called *chain recurrent* if for every $\varepsilon > 0$ there is an n (depending on ε) and an ε-chain $x_1, x_2, \dots, x_n$ with $x_1 = x_n = x$. The set $\mathbf{R}$ of chain recurrent points is called the *chain recurrent set* of f.

It is easily seen that $\mathbf{R}$ is compact and invariant under f.

If $A \subset X$ is a compact subset and there is an open neighborhood U of A such that $f(cl(U)) \subset U$ and $\bigcap_{n \geq 0} f^n(cl(U)) = A$, then A is called an *attractor* and U is its isolating neighborhood. It is easy to see that if $V = X - cl(U)$ and $A^* = \bigcap_{n \geq 0} f^{-n}(cl(V))$, then A^* is an attractor for f^{-1} with isolating neighborhood V. The set A^* is called the *repeller* dual to A. It is clear that A^* is independent of the choice of isolating neighborhood U for A. Obviously $f(A) = A$ and $f(A^*) = A^*$.

(1.2) LEMMA. *The set of attractors for f is countable.*

PROOF: Choose a countable basis $B = \{V_n\}_{n=1}^{\infty}$ for the topology of X. If A is an attractor with open isolating neighborhood U, then U is a union of sets in B. Hence, since A is compact, there are $V_{i_1}, \dots, V_{i_k}$ such that $A \subset V_{i_1} \cup \dots \cup V_{i_k} \subset U$. Clearly $A = \bigcap_{n \geq 0} f^n(U) = \bigcap_{n \geq 0} f^n(V_{i_1} \cup \dots \cup V_{i_k})$. Consequently there are at most as many attractors as finite subsets of B, i.e., the set of attractors is countable. q.e.d.

(1.3) LEMMA. *If $\{A_n\}_{n=1}^{\infty}$ are the attractors of f and $\{A_n^*\}$ their dual repellers, then the chain recurrent set $\mathbf{R}(f) = \bigcap_{n=1}^{\infty}(A_n \cup A_n^*)$.*

PROOF: We first show $\mathbf{R} \subset \cap(A_n \cup A_n^*)$. This is equivalent to showing that if $x \notin A \cup A^*$ for some attractor A, then $x \notin \mathbf{R}(f)$. If U is an open isolating neighborhood of A and $x \notin A \cup A^*$, then $x \in f^{-n}(U)$ for some n. Let m be the smallest such n. Replacing U with $f^{-m}(U)$ we can assume $x \in U - f(U)$. Now choose $\varepsilon_0 > 0$ so that any ε_0-chain $x = x_1, x_2, x_3$ must have $x_3 \in f^2(U)$. If $\varepsilon_1 = d(X - f(U), cl(f^2(U)))$ and $\varepsilon = \frac{1}{2}\min\{\varepsilon_0, \varepsilon_1\}$, then no ε-chain can start and end at x, since no ε-chain from a point of $f^2(U)$ can reach a point of $X - f(U)$. Thus $x \notin \mathbf{R}(f)$. We have shown $\mathbf{R}(f) \subset \cap(A_n \cup A_n^*)$.

We next show the reverse inclusion. Suppose $x \in \bigcap_{n=1}^{\infty}(A_n \cup A_n^*)$. If x is not in $\mathbf{R}(f)$, there is an $\varepsilon_0 > 0$ such that no ε_0-chain from x to itself exists. Let $\Omega(x, \varepsilon)$ denote the set of $y \in X$ such that there is an ε-chain from x to y. By definition, the set $V = \Omega(x, \varepsilon_0)$ is open. Moreover, $f(cl(V)) \subset V$, because if $z \in cl(V)$, there is $z_0 \in V$ such that $d(f(z), f(z_0)) < \varepsilon_0$ and consequently an ε_0-chain from x to z_0, gives an ε_0-chain $x = x_1, x_2, \dots, x_k, z_0, f(z)$ from x to $f(z)$. Hence $A = \bigcap_{n \geq 0} f^n(cl(V))$ is an attractor with isolating neighborhood V. By assumption either $x \in A$ or $x \in A^*$. Since there is no ε_0-chain from x to x, $x \notin A$. On the other hand, if $\omega(x)$ denotes the limit points of $\{f^n(x) \mid n \geq 0\}$, then clearly $\omega(x) \subset V$, but this is not possible if $x \in A^*$ since A^* is closed and $x \in A^*$ would imply $\omega(x) \subset A^*$. Thus we have contradicted the assumption that $x \notin \mathbf{R}$. q.e.d.

If we define a relation $\sim$ on $\mathbf{R}$ by $x \sim y$ if for every $\varepsilon > 0$ there is an ε-chain from x to y and another from y to x, then it is clear that $\sim$ is an equivalence relation.

(1.4) **Definition.** The equivalence classes in $\mathbf{R}(f)$ for the equivalence relation $\sim$ above are called the *chain transitive components* of $\mathbf{R}(f)$.

(1.5) PROPOSITION. *If $x, y \in \mathbf{R}(f)$, then x and y are in the same chain transitive component if and only if there is no attractor A with $x \in A$, $y \in A^*$ or with $y \in A$, $x \in A^*$.*

PROOF: Suppose first that x and y are in the same chain transitive component, i.e., $x \sim y$, and $x \in A$. If U is an open isolating neighborhood for A, let $\varepsilon =$ dist$(X - U, cl(f(U)))$. There can be no $\varepsilon/2$-chain from a point in $f(U)$ to a point in $X - U$, hence none from a point in A to a point in A^*. By (1.3) $y \in A \cup A^*$, but $x \sim y$ implies $y \notin A^*$, so $y \in A$. This proves one direction of our result.

To show the converse, suppose that for every attractor A, $x \in A$ iff $y \in A$ (and hence $x \in A^*$ iff $y \in A^*$). Given $\varepsilon > 0$ let $V = \Omega(x, \varepsilon) =$ the set of all points z in $\mathbf{R}$ for which there is an ε-chain from x to z. Since x is chain recurrent $x \in V$.Also as in the proof of (1.3) V is an isolating neighborhood for an attractor A_0. Since $x \in A_0 \cup A_0^*$ and $x \in V$ we have $x \in A_0$. Thus $y \in A_0 \subset V$ so there is an ε-chain from x to y. A similar argument shows there is an ε-chain from y to x so $x \sim y$. q.e.d.

We are now prepared to present Conley's proof of the existence of a complete Lyapounov function.

(1.6) **Definition.** A complete Lyapounov function for $f : X \to X$ is a continuous function $g : X \to R$ satisfying:

(1) If $x \notin \mathbf{R}(f)$, then $g(f(x)) < g(x)$
(2) If $x, y \in \mathbf{R}(f)$, then $g(x) = g(y)$ iff $x \sim y$ (i.e., x and y are in the same chain transitive component
(3) $g(\mathbf{R}(f))$ is a compact nowhere dense subset of R.

By analogy with the smooth setting, elements of $g(\mathbf{R}(f))$ are called *critical values* of g.

(1.7) LEMMA. *There is a continuous function $g : X \to [0,1]$ such that $g^{-1}(A) = 0$, $g^{-1}(A^*) = 1$ and g is strictly decreasing on orbits of points in $X - (A \cup A^*)$.*

PROOF: Define $g_0 : X \to [0,1]$ by

$$g_0(x) = \frac{d(x,A)}{d(x,A) + d(x,A^*)}.$$

Let $g_1(x) = \sup\{g_0(f^n(x)) \mid n \geq 0\}$. Then $g_1 : X \to [0,1]$ and $g_1(f(x)) \leq g_1(x)$ for all x. We must show g_1 is continuous. If $\lim x_i = x \in A$, then clearly $\lim g_1(x_i) = 0$ so g_1 is continuous at points of A and the same argument shows it is continuous at points of A^*. If U is an open isolating neighborhood as above, let $N = U - f(cl(U))$. Let $x \in N$ and $r = \inf\{g_0(x) \mid x \in N\}$. Since $f^n(N) \subset f^n(cl(U))$ and $\bigcap_{n \geq 0} f^n(cl(U)) = A$, it follows that there is $n_0 > 0$ such that $g_0(f^n(N)) \subset [0, r/2]$ whenever $n > n_0$. Hence for $x \in N$,

$$g_1(x) = \max\{g_0(f^n(x)) \mid 0 \leq n \leq n_0\}$$

so g_1 is continuous on N. Since $\bigcup_{n=-\infty}^{\infty} f^n(N) = X - (A \cup A^*)$, g_1 is continuous. Finally, letting

$$g(x) = \sum_{n=0}^{\infty} \frac{g_1(f^n(x))}{2^{n+1}}$$

we obtain a continuous function $g : X \to [0,1]$ such that $g^{-1}(0) = A$, $g^{-1}(1) = A^*$. Also

$$g(f(x)) - g(x) = \sum_{n=0}^{\infty} \frac{g_1(f^{n+1}(x)) - g_1(f^n(x))}{2^{n+1}}$$

which is negative if $x \notin A \cup A^*$, since $g_1(f(y)) \leq g_1(y)$ for all y and g_1 is not constant on the orbit of x. q.e.d.

The following theorem is essentially a result of [C]. We have changed the setting from flows to homeomorphisms.

(1.7) THEOREM. *If $f : X \to X$ is a homeomorphism of a compact metric space, then there is a complete Lyapounov function $g : X \to R$ for f.*

PROOF: By (1.2) there are only countably many attractors $\{A_n\}$ for f. By (1.7) we can find $g_n : X \to R$ with $g_n^{-1}(0) = A_n$, $g_n^{-1}(1) = A_n^*$ and g_n strictly decreasing on $X - (A_n \cup A_n^*)$. Define $g : X \to R$ by

$$g(x) = \sum_{n=1}^{\infty} \frac{2g_n(x)}{3^n}.$$

The series converges uniformly so $g(x)$ is continuous. Clearly if $x \notin \mathbf{R}(f)$, then there is an A_i with $x \notin (A_i \cup A_i^*)$ so $g(f(x)) < g(x)$.

Also, if $x \in \mathbf{R}(f)$, then $x \in (A_n \cup A_n^*)$ for every n, so $g_n(x) = 0$ or 1 for all n. It follows that the ternary expansion of $g(x)$ can be written with only the digits 0 and 2, and hence $g(x) \in C$, the Cantor middle third set. Thus $g(\mathbf{R}(f)) \subset C$ so $g(\mathbf{R}(f))$ is compact and nowhere dense. This proves (3) of the definition.

Finally, if $x, y \in \mathbf{R}(f)$ and $g(x) = g(y)$, then it is clear that for all n, $g_n(x) = g_n(y)$ since $2g_n(x)$ is the n^{th} digit of the ternary expansion of $g(x)$ and, as above, this expansion will contain only the digits 0 and 2 iff there is no n with $x \in A_n$, $y \in A_n^*$ or with $x \in A_n^*$, $y \in A_n$. Thus by (1.5), $g(x) = g(y)$ iff x and y are in the same chain transitive component. q.e.d.

§2. BOUNDARY TWIST MAPS

In this section we consider homeomorphisms of the annulus A which are homotopic to the identity map and satisfy a boundary twist condition.

(2.1) **Definition.** A homeomorphism $f : A \to A$, homotopic to the identity satisfies a *boundary twist condition* provided there is a lift of f, $\tilde{f} : \tilde{A} \to \tilde{A}$ to the universal covering space $\tilde{A} = R \times I$ satisfying $\tilde{f}_1(x, 0) < x$ and $\tilde{f}_1(x, 1) > x$, where $\tilde{f}(x, s) = (\tilde{f}_1(x, s), \tilde{f}_2(x, s))$.

The key lemma for finding fixed points is the following, which is essentially from [F].

(2.2) LEMMA. *Suppose $f : R^2 \to R^2$ is an orientation preserving homeomorphism and the chain recurrent set $\mathbf{R}(f)$ is not empty, then for any $\varepsilon > 0$ there is a point $x \in R^2$ with $\|f(x) - x\| < \varepsilon$.*

PROOF: We briefly recall the argument from [F]. If the conclusion is false there is an ε_0 with $\|f(x) - x\| \geq \varepsilon_0$ for all $x \in R^2$. A result of [Ox] asserts that there is a $\delta > 0$ such that for any finite set of pairs $\{(x_i, y_i)\}$ with $\|x_i - y_i\| < \delta$ there is a pairwise disjoint set of piecewise linear arcs γ_i from x_i to y_i with the length of each $< \varepsilon_0/2$. If $z \in \mathbf{R}(f)$ let $z_1 = z, z_2, \ldots, z_n = z$ be a δ-chain from z to z. Setting $y_i = z_i$, $x_i = f(z_{i-1})$ we see that there are pairwise disjoint arcs γ_i from $f(z_{i-1})$ to z_i all of length $< \varepsilon_0/2$. By isotoping in a neighborhood of these arcs we can produce a perturbation g of f satisfying $g(z_{i-1}) = g(z_i)$ and $\|f(x) - g(x)\| < \varepsilon_0$ for all x.

Now g has a periodic point, namely z. Hence by results of [Br] or [Fa] g has a fixed point p. Thus $\|f(p) - p\| \leq \|f(p) - g(p)\| + \|g(p) - p\| < \varepsilon_0$ which is a contradiction. q.e.d.

(2.3) PROPOSITION. *Suppose $f : A \to A$ is a homeomorphism homotopic to the identity and satisfies the boundary twist condition. If for every ε there is an ε-chain for f from $S^1 \times \{0\}$ to $S^1 \times \{1\}$ and one from $S^1 \times \{1\}$ to $S^1 \times \{0\}$, then f has at least one fixed point.*

PROOF: We first observe that it is possible to extend f to a homeomorphism of $S^1 \times [-\delta, 1+\delta]$ in such a way that f is a rotation on $S^1 \times \{-\delta\}$ and $S^1 \times \{1+\delta\}$, and that the lift $\widetilde{f} : R \times [-\delta, 1+\delta] \to R \times [-\delta, 1+\delta]$ satisfies:

(1) $\widetilde{f}(x,t) = (\widetilde{f}_1(x,t), t)$ for $t \in [-\delta, 0] \cup [1, 1+\delta]$

(2) $\widetilde{f}_1(x,t) < x$ for $t \in [-\delta, 0]$

(3) $\widetilde{f}_1(x,t) > x$ for $t \in [1, 1+\delta]$.

From these properties it is easy to check that $S^1 \times [-\delta, 0]$ is contained in a single chain transitive component as is $S^1 \times [1, 1+\delta]$. Since there can be no fixed points in $S^1 \times [-\delta, 0] \cup [1, 1+\delta]$, it suffices to prove the result for the enlarged annulus for which f is a rotation on each boundary component. We proceed to do this referring to the annulus as $A = S^1 \times [0,1]$ rather than $S^1 \times [-\delta, 1+\delta]$.

Consider the lift $\widetilde{f} : \widetilde{A} \to \widetilde{A}$ ($\widetilde{A} = R \times [0,1]$) satisfying

$$\widetilde{f}(x,1) = (x + r_1, 1) \qquad \text{and} \qquad \widetilde{f}(x,0) = (x - r_2, 0)$$

for some $r_1, r_2 > 0$. We will show that every point of $R \times \{0\}$ is chain recurrent for $\widetilde{f}$; before doing so, however, we show that this will suffice to complete the proof. We can extend $\widetilde{f}$ to R^2 by setting $\widetilde{f}(x,t) = (x + r_1, t)$ for $t > 1$ and $\widetilde{f}(x,t) = (x - r_0, t)$ for $t < 0$. Thus by (2.2), for every $\varepsilon > 0$ there is an (x,t) with $\|\widetilde{f}(x,t) - (x,t)\| < \varepsilon$. Since this inequality also holds for f and the projection of (x,t) on the compact annulus A, it follows that f has a fixed point.

It remains to show that points of $R \times \{0\}$ are chain recurrent. Let $x_0 \in S^1$. By our remarks above, for every ε there is an ε-chain for f from $(x_0, 0)$ to $(x_0, 1)$ and likewise one from $(x_0, 1)$ to $(x_0, 0)$. If $(x, 0)$ is a lift of $(x_0, 0)$, then lifting the ε-chain on A we obtain an ε-chain for $\widetilde{f} : \widetilde{A} \to \widetilde{A}$ from $(x,0)$ to $(x+a, 1)$ for some $a \in Z$. The fact that $\widetilde{f}(y,1) = (y + r_1, 1)$ for all $y \in R$ implies there is an ε-chain from $(x,1)$ to $(x+p, 1)$ for any sufficiently large $p \in Z$. Likewise there is an ε-chain from $(x,1)$ to $(x+b, 0)$ for some $b \in Z$. Using the fact that $\widetilde{f}(x,t) + (0,1) = \widetilde{f}(x+1, t)$ to translate these ε-chains we can piece together an ε-chain from $(x,0)$ to $(x+a,1)$ to $(x+a+p,1)$ to $(x+a+p+b, 0)$. If p is sufficiently large $m = a+p+b > 0$.

We also know that $\widetilde{f}(x,0) = (x - r_2, 0)$ so there is an ε-chain from $(x,0))$ to $(x-n,0)$ for some integer $n > 0$. Once again piecing together we obtain an ε-chain from $(x,0)$ to $(x+m,0)$ to $(x+2m,0)$ to ... to $(x+mn,0)$ to $(x+mn-n,0)$ to ... to $(x,0)$. Thus $(x,0)$ is chain recurrent. q.e.d.

We can now give the proof of our main result.

(2.4) THEOREM. *If $f : A \to A$ is a homeomorphism homotopic to the identity and satisfying the boundary twist condition, then either f has at least one fixed point or there is a smoothly embedded essential closed curve $C \subset A$ such that $f(C) \cap C = \emptyset$.*

PROOF: As in the proof of (2.3) we enlarge the annulus to $A_0 = S^1[-\delta, 1+\delta]$ and extend f so that it is a rotation on each boundary component and so that the circles $\{S^1 \times \{t\} \mid t \in [-\delta, 0] \cup [1, 1+\delta]\}$ are invariant and contain no fixed points. It follows that if $A^+ = \{(x,t) \in A_0 \mid t \in [1, 1+\delta]\}$ and $A^- = \{(x,t) \in A_0 \mid t \in [-\delta, 0]\}$, then A^+ and A^- are chain recurrent and each is a subset of a chain transitive component, (i.e., between any two points of A^+ there is an ε-chain and similarly for A^-).

If A^+ and A^- are in the same chain transitive component for f, then f has at least one fixed point by (2.3). So suppose this is not the case. Let $g : A_0 \to R$ be a complete Lyapounov function for f. Then g is constant on A^+ and A^- and $g(A^+) \neq g(A^-)$. Choose $c \in R$ which is not a critical value of g and which lies between $g(A^+)$ and $g(A^-)$. The set $g^{-1}(c) \subset A$ has the property that it separates A and $f(g^{-1}(c)) \cap g^{-1}(c) = \emptyset$. It is possible to construct g in such a way that it is C^∞ (see [W]), though that is not necessary for our purposes. In fact, if g_0 is a C^∞ function which is a sufficiently close C^0 approximation to g, then $g_0^{-1}(c)$ will separate A and $f(g_0^{-1}(c)) \cap g_0^{-1}(c) = \emptyset$. By Sard's theorem we can pick a regular value of g_0 arbitrarily close to c. We choose a regular value c_0 such that $g_0^{-1}(c_0)$ separates A and $f(g_0^{-1}(c_0)) \cap g_0^{-1}(c_0) = \emptyset$. The components of $g_0^{-1}(c_0)$ are smoothly embedded circles, at least one of which must be essential since $g_0^{-1}(c_0)$ separates A. Let C be such an essential component, then $f(C) \cap C = \emptyset$. q.e.d.

References

[B] G. D. Birkhoff, *Proof of Poincaré's Last Geometric Theorem*, Trans. Amer. Math. Soc. **14** (1913), 14–22.

[B2] G. D. Birkhoff, *An extension of Poincaré's last geometric theorem*, Acta Math. **47** (1925), 297–311.

[Br] M. Brown, *A New Proof of Brouwer's Lemma on Translation Arcs*, Houston Jour. of Math. **10** (1984), 35–41.

[BN] M. Brown and W. D. Neumann, *Proof of the Poincaré-Birkhoff Fixed Point Theorem*, Michigan Math. J. **24** (1977), 21–31.

[Ca] P. Carter, *An Improvement of the Poincaré-Birkhoff Fixed Point Theorem*, Trans. Amer. Math. Soc. **269** (1982), 285–299.

[C] C. Conley, *Isolated Invariant Sets and the Morse index*, C.B.M.S. Regional Conference Series in Math. **38**, Amer. Math. Soc., Providence, RI, 1978.

[Fa] A. Fathi, An orbit closing proof of Brouwer's lemma on translation arcs, L'enseignement Mathématique, **33** (1987), 315-322.

[F] J. Franks, *Recurrence and Fixed Points of Surface Homeomorphisms*, to appear in Ergodic Theory and Dyn. Systems.

[Ox] J. Oxtoby, *Diameters of Arcs and the Gerrymandering Problem*, Amer. Math. Monthly **84** (1977), 155–162.

[W] W. Wilson, *Smoothing Derivatives of Functions and Applications*, Trans. Amer. Math. Soc. **139** (1969), 413–428.

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986261

Contemporary Mathematics
Volume 81, 1988

AN ANALOG OF SHARKOVSKI'S THEOREM FOR TWIST MAPS

Philip Boyland[1]

ABSTRACT: In this exposition, we present a partial order on the periodic orbits of twist maps of the annulus. It is the analog of Sharkovski's theorem for maps of the line. A periodic orbit is specified by its "orbit type" which is essentially the isotopy class of the map rel the orbit. This isotopy class is then analyzed using the Thurston-Nielsen theory of surface automorpisms. If this class has a pseudo-Anosov component, then the given orbit type dominates the infinite number of orbit types of the canonical representive in the class. The use of the partial order as a tool in understanding the birth of periodic orbits after the loss of a invariant circle is also discussed.

Section 1. Introduction:

The study of twist maps of the annulus was initiated by Poincare in his investigation of the restricted three body problem. These maps were also a central focus in Birkhoff's work on dynamical systems. In the ensuing years, twist maps have been found to occur in numerous situations of mathematical and physical interest. As maps they are relatively simple, yet they can give rise to very complicated dynamics. For this reason, they provide a nice model problem for Hamiltonian dynamical systems. In recent years there has been great progress in understanding the behavior of these maps (for a survey, see [Mr] or [Mc]), but there is much to do before a complete understanding is available.

The range of dynamical behavior in twist maps is from simple in the integrable case, to very complex, stochastic type behavior in the unstable case. For a typical twist map, the dynamics are composed of a combination of these two behaviors. The characteristic configuration consists of invariant circles which wrap around the annulus and divide it into "zones of instability". It is in these zones that the apparently random motion occurs. The transition from simple to complex dynamics in parametrized families is therefore

1

0271-4132/88 $1.00 + $.25 per page

closely connected to the phenomenon of the loss of invariant circles. Consequently, this phenomenon has been the subject of much study resulting in deep results, most notably, the Kolmogorov, Arnol'd, Moser (KAM) theory and the Aubry-Mather Theorem.

Very roughly, the KAM theory gives conditions which ensure that an invariant circle is preserved under perturbation. The Aubry-Mather Theorem gives information about what remains after an invariant circle is lost. What remains is a "circle with gaps", or more precisely, an invariant Cantor set (called an Aubry-Mather set). It has been known since Birkhoff that when an invariant circle disappears, there is transit in both directions through the region formerly occupied by the circle. This results in a kind of local circulation around the Aubry-Mather set which gives rise to, among other things, new periodic orbits ([BH]) and invariant Cantor sets ([Ma]). One would like to understand the type of periodic orbit created and the order of their creation when an invariant circle is lost. This is the question we wish to begin addressing here by developing an analog of Sharokovski's Theorem for twist maps.

This question forms a special case of a more general strategy: One attempts to understand complicated dynamics by understanding how the transition to this behavior occurs in parametrized families ([cf.[E]). It is somewhat remarkable that in certain (admittably special) instances there are sequences of events (i.e. bifurcations) which must always occur in this transition. In the current language of physics, there exist "universal structures in the transition to chaos." It is worth noting that the study of structures intrinsic to the transition from simple behavior to complexity is not restricted to dynamics. It occurs, for example, in the theory of critical phenomena (as critical exponents), in the study of the transition to turbulence, in unified quantum field theories (as sequential spontaneous symmetry breaking) and even in recent proposals for the identification of pure consciousness with the unified field (as the sequential unfolding of the unmanifest relationships between observer, observed and process of observation [Ha]).

Within dynamics, this strategy has been most successful to date when applied to one dimensional maps. The first example of a qualitative "universal structure" was provided by Sharkovski's Theorem ([Sh] and [St]). (Quantitative universality usually arises from the use of renormalization techniques in situations such as Feigenbaum doubling). Sharkovski's Theorem consists of a total order on the natural numbers, denoted "$\succ$", with the property that $n \succ m$ if and only if any continuous map of the real line which has a periodic orbit of period n also has one of period m. One implication for bifurcation theory is the following: For $\mu \in [0,1]$, let f_μ be a parametrized family such that f_0 has no periodic orbits and f_1 has periodic orbits of all periods (eg. $f_\mu = 4\mu x(1-x)$). If μ_n denotes the infimum of the parameter values for which f_μ has a period n orbit then $n \succ m$ implies $\mu_n \geq \mu_m$. (The stronger result $\mu_n > \mu_m$ is true by [BkHt]. One also has $\mu_n > \mu_m$ implies $n \succ m$ since the order is total). In this way, the partial order on $\mathbf{N}$ yields an order on parameter space, i.e. information on the order in which periodic orbits must be born.

One obvious limitation of Sharkovski's Theorem is that it only involves the period of the orbits. One may gain more information about the sequence of bifurcations by generalizions of the following form: First, some type of specification is assigned to periodic orbits. Then, an order relation is put on the set of specifications which is defined by the property that a first specification dominates a second if and only if any map that has a periodic orbit of the first type also has one of the second type. As with Sharkovski's Theorem, one then obtains information about the order in which periodic orbits (with given specifications) can be born in parametrized families.

One such generalization bears useful parallels to the two dimensional theory we develop here. In this generalization one specifies a periodic orbit by the cyclic permutation induced by the action of the map on the points of the orbit as ordered in **R** ([Ba],[J2],[Bt] and [BkHt]). The resulting order relation is now partial and not total. The main tool for computing in this order is the *primitive map.* The primitive map is the simplest piecewise linear map that has a periodic orbit with the given permutation (see fig. 4 below). Using what is essentially a Markov partition argument (the segments between adjacent points on the orbit act as Markov boxes), one can show that the permutation of any periodic orbit of the primitive map must be represented by a periodic orbit in any map that has a periodic orbit with the given permutation structure . Thus using the primitive map, one can compute all the permutations dominated by the given one. One sees then that the existence of a periodic orbit with a given permutation structure implies that the map must be at least as complicated as the primitive map of the given permutation. What is important then about the existence of a periodic orbit with a given permutation structure is what is implied about the action of the map on the complement of the orbit. The passage from the permutation of a periodic orbit to the implied behavior of the map is made possible by the one dimensionality of the system. In two dimensions a similar passage is only possible with addition hypotheses such as the twist condition.

The overall form of a two dimensional "Sharkovski's Theorem" is the same as the form given above for one dimensional maps. The question is, of course, what to chose as a specification of periodic orbits. The notion of permutation is not well defined in two dimensions and the period as a specification yields a trivial order relation (for example, in maps of the annulus no two numbers would be related). An appropriate two dimensional notion of permutation would describe how the points on the orbit are moved around the surface topologically by the map. One way to capture this notion is by the isotopy class of the map rel the orbit, or equivalently, the isotopy class of the map restricted to the surface minus the orbit ([cf [Bn]). The specification we assign to periodic orbits is called the *orbit type* and is basically this isotopy class. (This idea is essentially due to Bowen [Bo]. We will review the history of this circle of ideas at the end of the introduction.)

Continuing the analogy with the one-dimensional permutation order, we now seek the simplest map in the isotopy class rel the orbit and hope it will play the role of the primitive map. Specifically, we want its periodic orbits to be present, in the appropriate sense, in any map contained in the isotopy class rel the orbit (and thus in any map that has a periodic orbit with the given orbit type). In addition, the periodic orbits of this simplest map should be computable via a Markov partition. In short, the simplest map should allow us to compute precisely which orbit types are dominated by the given one.

A simplest map with these properties does, in fact, exist as is shown in Thurston's work on isotopy classes of diffeomorphisms of surfaces of negative Euler characteristic ([T],cf [FLP],[C]). Although Thurston's work lies at the foundation of all we develop here, there is regretably insufficient space to describe it in any detail. A brief outline is included in section 3.

The main difficulty in implementing this program lies in computing the behavior of the simplest map when given a combinatorial description of an isotopy class (or orbit type). In addition, the set of orbit types is large and the partial order appears to be quite complex. For these reasons it is useful to restrict consideration to certain more easily understood (but still interesting) subcollections of orbit types. In this paper we present the intial stages of an investigation into the partial order restricted to orbit types that arise from periodic orbits of monotone twist maps of the annulus. Complete proofs will appear elsewhere.

The monotone twist hypothesis is quite convenient in this setting as it allows the passage from permutation information about an orbit to the behavior of the map on its complement. (Recall that this passage is the main principle underlying the one dimensional permutation order.) More precisely, if one knows the ordering of the iterates of a periodic orbit around the annulus (i.e. its permutation in the angular coordinate) then as a consequence of the twist hypothesis, one can deduce the nature of the isotopy class on its complement (i.e. its orbit type). This allows one to use an angular permutation to represent the orbit type. This in turn, allows one to use the more easily understood partial order on periodic orbits of circle maps as a tool in studying the partial order on orbit types. This is possible because the orbit type order relationis a subset of the permutation order relation in the sense that a first orbit type dominating a second implies that a permutation representing the first dominates one representing the second. The problem basically reduces to discovering in which cases and in what sense the converse of this statement is true.

We have thus far presented the two dimensional theory as a generalization of the one dimensional theory. This was primarily for the purpose of motivation and exposition and is not an accurate representation of the histoy of the ideas we have presented. Put somewhat loosely, the basic principle that underlies these ideas is that certain dynamical behavior is caused by topological complications of a map and is thus preserved under isotopy. This principle goes back at least to Lefschetz and Nielsen and has arisen in numerous guises in the ensuing years. The instance of this principle applied here is its application to pseudo-Anosov isotopy classes on surfaces. Our utilization of this follows that of Birman and Williams who used it in a similar manner in the context of flows in their study of the planetary orbits of a fibered link ([BW]). The investigation of periodic orbits using the isotopy class rel the orbit is the other main idea employed here. As noted above, this idea seems to be due to Bowen ([Bo]). It has been applied by numerous authors, for example, [BF],[Hd] and [Fr]. The use of the knot type in the suspension as in [HW] is a closely related notion. The combination of these basic ideas to yield a partial order on periodic orbits of the disk or annulus appears to have been in the folklore for some time. The idea occured to J.Smillie, the author and certainly others. It does not seem to have been the subject of a sustained investigation or published exposition. This is our justification for writing this rather lenghthy introduction. Matsuoka has investigated the partial order related to period three orbits using techniques connected with the Bureau representation of the braid group ([Ms2], cf. [Ms1]). In the context of monotone twist maps, the work of Jungries [J1] should also be noted.

Section 2. Order Structures

We begin by defining the object that will specify the order structure of a periodic orbit around the annulus. It is essentially just the permutation of the points on the orbit under the map. However, because we are interested in what the orbit implies about the topological complications of the map, it turns out to be necessary to work with the lift of the orbit in the universal cover and to describe the permutation structure there.

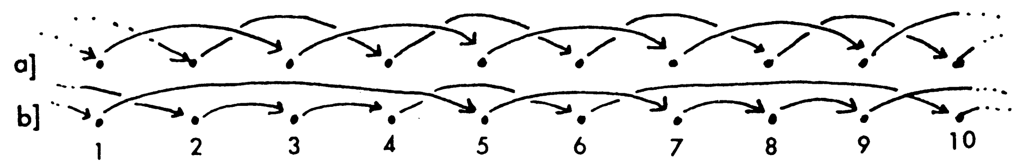


Fig. 1

In figures 1a) and b), we show the lifts of two different orbits to the universal cover. The forward iterate of each point is indicated by an arrow. Both orbits have rotation number 2/5. Note that a single orbit in the base can be covered by several orbits (in this case, by two orbits).

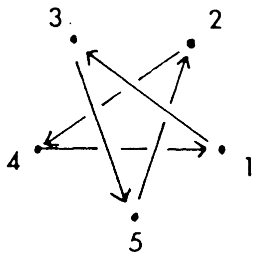
DEFINITION: If $\sigma : \mathbf{Z} \to \mathbf{Z}$ is a bijection and p and q are positive integers, then the triple (σ, p, q) is called a (p,q)-*order structure* if for all $n \in \mathbf{Z}$ the following hold:

1) $\sigma(n+q) = \sigma(n) + q$
2) $\sigma^q(n) = n + pq$
3) $\sigma^j(n) \neq n + kq$ for $0 < j < q$ and $0 \leq k < p$.

Statement (1) implies that σ projects to a permutation $\sigma' : \mathbf{Z}/q\mathbf{Z} \to \mathbf{Z}/q\mathbf{Z}$. In other words, our "fundamental domain" is q points long. Statements (2) and (3) imply that σ' is cyclic and the resulting periodic orbit has "rotation number" p/q. Note that we are not assuming that p and q are relatively prime. In figures 2(a) and 2(b), we show the projection of the orbits in figures 1(a) and (b) to $\mathbf{Z}/5\mathbf{Z}$.

Because we do not want an order structure to depend on the placement of zero, we need to introduce an equivalence relation. If (σ, p, q) is an order structure, let $\sigma \sim T^m \sigma T^{-m}$ for all $m \in \mathbf{Z}$, where T is the translation, $T(n) = n + 1$. The resulting equivalence classes will also be called order structures and we will not be unnecessarily careful in distinguishing between the two.

A given bijection, σ, may be part of a (p,q)-order structure for many p's and q's. For example, if $\sigma_B(n) = n + p$, then for each relatively prime pair, (p,q), the triple



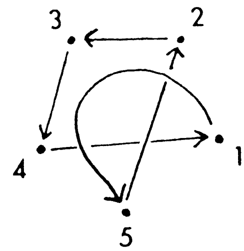


Fig. 2

(σ_B, p, q) is a (p, q)-order structure called the *monotone* or *Birkhoff* (p, q)-order structure. (cf. [K],[H]). Figures 1(a) and 2(a) show the monotone $(2, 5)$-order structure and its projection. Note that the monotone order structures are the only ones in which σ is order preserving. This means that the monotone (p, q)-order structure is associated with periodic orbits of circle homeomorphisms. Put in another way, if the projection of σ_B is σ'_B which acts on on $\mathbf{Z}/q\mathbf{Z}$, then σ'_B preserves the "radial order". However, it is not the only order structure with this property. For example, if $\tau : \mathbf{Z} \to \mathbf{Z}$ is defined by

$$\tau(n) = \begin{cases} n + p & \text{for } n \not\equiv 0 \bmod q; \\ n + 2p & \text{otherwise,} \end{cases}$$

then τ and σ_B project to the same permutation on $\mathbf{Z}/q\mathbf{Z}$. This loss of information after projecting is what necessitates the use of the cover in the definition of order structure.

Before defining what we mean by the order structure of a given periodic orbit, we introduce some notation. The annulus $A = S^1 \times [0, 1]$ has universal cover $\tilde{A} = \mathbf{R} \times [0, 1]$ with projections $\pi_x : \tilde{A} \to \mathbf{R}$ and $\pi_y : \tilde{A} \to [0, 1]$ onto the first and second coordinates respectively. The maps we shall consider are diffeomorphisms of A that are orientation and boundary preserving. The set of all such maps is denoted $\mathrm{Diff}_+(A)$. Given $f \in \mathrm{Diff}_+(A)$, a map $F : \tilde{A} \to \tilde{A}$ is called the *lift* of f if it is a lift in the usual sense and in addition $F(0, 0) \in [0, 1) \times \{0\}$. The *orbit* of a point $\beta \in A$ under f is denoted $o(\beta, f)$ and the *extended orbit* of β, denoted $eo(\beta, f)$, is the set of points in $\tilde{A}$ which cover a point of $o(\beta, f)$, i.e. $eo(\beta, f) = p^{-1}(o(\beta, f))$ where $p : \tilde{A} \to A$ is the projection. If $o(\beta, f)$ is a periodic orbit with period q, its rotation number, denoted $\rho(x, f)$, is p/q, where p satisfies $\pi_x(F^q(\tilde{\beta})) = \pi_x(\tilde{\beta}) + (p, 0)$ for $\tilde{\beta}$ and F lifts of β and f respectively.

DEFINITION: Let $f \in \mathrm{Diff}_+(A)$ have a periodic orbit $o(\beta, f)$. Fix a $\beta_0 \in eo(\beta)$ and if we assume that π_x restricted to $eo(\beta)$ is injective we may label the elements of $eo(\beta)$ as $\{\beta_i : i \in \mathbf{Z}\}$ with $\pi_x(\beta_i) < \pi_x(\beta_j)$ if and only if $i < j$. Now let $\sigma : \mathbf{Z} \to \mathbf{Z}$ be the bijection induced on $\mathbf{Z}$ by the action of F on $eo(\beta)$, i.e. $F(\beta_i) = \beta_{\sigma(i)}$. If $o(\beta)$ has period q and rotation number p/q then (σ, p, q) will be a (p, q)-order structure and will be called the *order structure* of the periodic orbit of $o(\beta, f)$. We denote this as $os(\beta, f) = (\sigma, p, q)$.

In order to have this definition make sense we needed to assume that π_x restricted to $eo(\beta)$ was injective. If this is not the case it may always be arranged by a small change of coordinates which keeps the map monotone twist. However, two choices of new coordinates can yield different order structures for the given periodic orbit. This is illustrated in figure 3(a) where we show the lift of a periodic orbit with two points on the same vertical line. The lines labeled (b) and (c) are transformed into vertical lines under the changes of coordinates indicated in 3(b) and 3(c) respectively. From this we see that if we try to use the order structure to specify a periodic orbit in a partial order of the Sharskovski type then we will find differing specifications for the two orbits that should be the same.

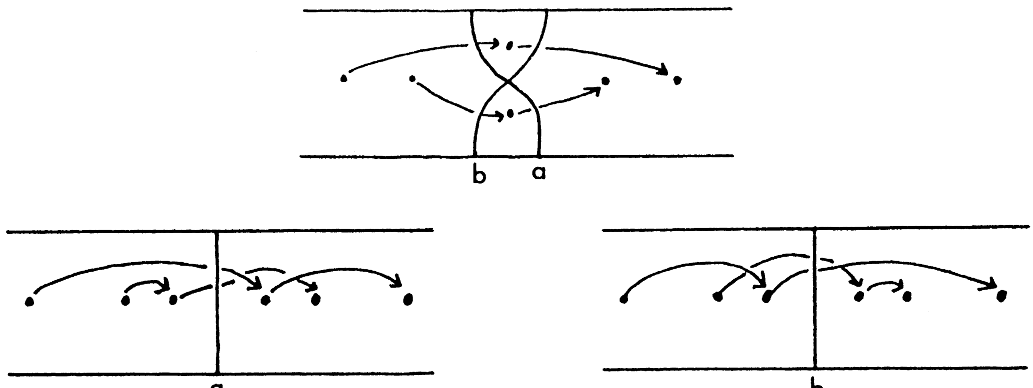


Fig. 3

Section 3. Orbit Types

The situation discussed at the end of the previous section is resolved by introducing the following equivalence relation on periodic orbits. Because we are going to use the equivalence classes under this relation to specify a periodic orbit, we also require that the action of the appropriate maps on the complement of the orbits be isotopic. Although we restrict our attention here to periodic orbits of monotone twist maps, the definition is made for any $f \in \mathrm{Diff}_+(A)$.

DEFINITION: Let $f, g \in \mathrm{Diff}_+(A)$ have periodic orbits $o(x,f)$ and $o(y,g)$ respectively. The orbits $o(x,f)$ and $o(y,g)$ have the same *orbit type* if there exists a homeomorphism $h : A \to A$ with $h \cong id$, $h(o(x,f)) = o(y,g)$, and $hfh^{-1} \cong g$ rel $o(y,g)$.

The conjugation in the definition of equivalence means that periodic orbits which correspond under changes of coordinates are equivalent. The requirement of isotopy rel $o(y,g)$ ensures that the action of f and g on the complement of their respective orbits are topologically the same. The equivalence class of $o(x,f)$ under orbit types is denoted $ot(x,f)$ and for $f \in \mathrm{Diff}_+(A)$, $ot(f)$ denotes the collection of orbit types of periodic orbits of f, i.e. $ot(f) = \{ot(x,f) : x \text{ is a periodic orbit of } f\}$. The set of all possible orbit types for maps of the annulus is defined as $OT = \{ot(x,f) : f \in \mathrm{Diff}_+(A) \text{ and } x \text{ is a periodic orbit of } f\}$.

For a fixed f with periodic orbit, $o(x,f)$, the isotopy class of f rel $o(x,f)$ gives an isotopy class (or mapping class) on the q-punctured annulus, where q is the period of $o(x,f)$. The orbit type of $o(x,f)$ can thus be thought of a conjugacy class in the mapping class group of the q-punctured annulus (or more properly, the $q+2$-punctured sphere since we allow isotopies that move the boundary of the annulus). The set OT is the union of such classes as q goes to infinity.

We now define the order relation on OT that is the analog of Sharkovski's order.

DEFINITION: If $\gamma, \eta \in OT$ let $\gamma \geq \eta$ if for all $f \in \mathrm{Diff}_+(A)$, $\gamma \in ot(f)$ implies $\eta \in ot(f)$.

THEOREM 1: The order relation $(OT, \geq)$ is a partial order.

We sketch the proof. To be a partial order, an order relation must be reflexive ($\alpha \geq \alpha$), transitive ($\alpha \geq \beta$ and $\beta \geq \gamma$ implies $\alpha \geq \gamma$) and antisymmetric ($\alpha \geq \beta$ and $\beta \geq \alpha$ implies $\alpha = \beta$). The first two properties are obvious. The main ingredient in the proof of the third is the fact that any map of the disk is isotopic to the identity. This means that any periodic orbit can be isotoped away. In particular, using somewhat standard techniques (cf [Br]) one can find an isotopy that forces a given pair of orbit types to "disappear " at different points in the isotopy. This implies that distinct orbit types are never co-related, in other words, $(OT, \geq)$ is antisymmetric.

The first issue to confront is the nontriviality of this order relation. Do most (or any) elements dominate infinitely many others? For this we need to briefly (and informally) describe the origins of the partial order in Thurston's work on surfaces (for more details, see [T], [FLP], [C]). Thurston shows that isotopy classes of diffeomorphisms on surfaces of negative Euler characteristic are either simple (= finite order), very complicated (= pseudo-Anosov), or are built up from components of these types. In the pseudo-Anosov classes, the infinite collection of periodic orbits of the pseudo-Anosov representative are unremovable in the sense of Asimov and Franks ([AF]). If one can show that the isotopy class of f rel $o(x, f)$ is of psuedoAnosov type, this implies that $ot(x, f)$ will dominate the infinitely many orbit types of the pseudo-Anosov representative. Since most (in a certain sense) mapping classes have at least one pseudo-Anosov component, one sees that most orbit types dominate infinitely many others. An orbit type with all finite order components will dominate a finite (usually nonzero) number of orbit types.

The general structure of $(OT, \geq)$ thus appears to be quite rich. One general strategy for computing in $(OT, \geq)$ follows the discussion above. First decide the Thurston type of one of the isotopy classes within the orbit type (the Thurston type is invariant under conjugacy). One then computes the dynamics of the canonical representative in the class. The first task is relatively straight forward, especially for certain restricted classes of orbits. The theory of train tracks ([Kn], [HP], [BS]) provides a powerful tool for attacking the second task but the actual computation of a train track appears to be highly trivial in general.

There is an analogous partial order for periodic orbits of diffeomorphisms of the disk and sphere. The obvious generalization to other two manifolds fails due to the existence of Anosov or psuedo-Anosov isotopy classes on the entire manifold and not just for classes rel a periodic orbit. In these classes, as noted above, many periodic orbits are present in every map in the isotopy class and so the order relation on orbit types fails to be antisymmetric.

Section 4. Monotone Twist Maps

Since computation in $(OT, \geq)$ appears to be rather difficult in general, we restrict our attention to orbit types arising in monotone twist maps (defined below). Given permutation information about a periodic orbit of a monotone twist map (i.e. its order structure) one can immediately compute the action of the map on the complement of the orbit up to isotopy (i.e. the orbit type). Thus within the class of monotone twist maps, the order structure provides a convenient combinatorial tool for working with orbit types. These notions are made precise in the follwing lemma and definitions. A map $f \in \mathrm{Diff}_+(A)$ is called *monotone twist* if

$$\frac{\partial \pi_y \circ f}{\partial x} > 0.$$

The set of all monotone twist maps is denoted MT.

LEMMA 2: Let $f, g \in MT$ have periodic orbits $o(x, f)$ and $o(y, g)$, respectively. If $os(x, f) = os(y, g)$ then $ot(x, f) = ot(y, g)$. In addition, given any order structure (σ, p, g), there exists an $f \in MT$ with periodic orbit $o(x, f)$ with $os(x, f) = (\sigma, p, q)$.

Lemma 2 allows us to make the following definition.

DEFINITION: Given an order structure (σ, p, q), choose an $f \in MT$ with periodic orbit $o(x, f)$ with $os(x, f) = (\sigma, p, q)$. Define the *orbit type* of (σ, p, q) as $ot(\sigma, p, q) = ot(x, f)$.

The converse of the first statement of Lemma 2 is false. Many order structures can give rise to the same orbit type. The two order structures arising from the coordinate changes illustrated in figure 3 are an example of this. (It would be nice to have an algorithm which decides when two order structures give rise to the same orbit type. This appears to be difficult.) There is one case in which the corespondance between order structures and orbit types is one to one. It follows from Hall([H1]) that if (σ_B, p, q) is the monotone (p, q)-order structure and $f \in MT$ then $ot(x, f) = ot(\sigma_B, p, q)$ if and only if $os(x, f) = (\sigma_B, p, q)$. We also note that $ot(\sigma_B, p, q)$ is a finite order class in the sense of Thurston and thus does not dominate anything in $(OT, \geq)$. In fact, the minimal elements of $(OT, \geq)$ (i.e. the elements that are smaller than anything with which they are comparable) are precisely the orbit types of (σ_B, p, q) for relatively prime pairs (p, q).

If (σ, p, q) is not monotone the situation is quite different as is indicated in the following theorem. It is rather easily proven by showing the isotopy class in question is pseudo-Anosov.

THEOREM 3: If (σ, p, q) is a (p, q)-order structure that is not monotone and p and q are relatively prime, then $ot(\sigma, p, q)$ dominates infinitely many elements in $(OT, \geq)$.

We now move on to the question of computing in $(OT, \geq)$ using an order structure to specify an orbit type. Given an order structure, we construct a map of the circle, called the *tight circle map*, which is, in a certain sense, the simplest map of the circle that has a periodic orbit with the given order structure. This map plays the role of the primitive map discussed in the introduction.The order structures of periodic orbits of this tight circle map are precisely the order structures dominated by the given one in maps of the circle. In pushing everything up a dimension, one finds that the orbit types of these order structures contain all the orbit types dominated by the orbit type of the given order structure. However, the containment is often proper and the problem becomes one of determining which of the dominated order structures "go up a dimension". Our primary success to date has been in finding order structures where all the order structures dominated in the one dimensional setting "go up" to orbit types dominated in the two dimensional case. We have recently made some progress on the general case.

Before stating our theorems, we need to develop a fair amount of combinatoric machinery for describing order structures.

DEFINITION: Given an order structure (σ, p, q), extend $\sigma : \mathbf{Z} \to \mathbf{Z}$ to a piecewise linear $F_\sigma : \mathbf{R} \to \mathbf{R}$ that satisfies $F_\sigma(x+q) = F_\sigma(x) + q$ and for all $x \notin \mathbf{Z}$, $dF(x)/dx$ exists. The *tight circle map* of (σ, p, q) is the projection of F_σ to $f_\sigma : S^1 \to S^1$, where $S^1 = \mathbf{R}/q\mathbf{Z}$. Since f_σ is a continuous, degree one circle map, the set of rotation numbers of f_σ, denoted $\rho(f_\sigma)$, is a closed interval ([I]). We define the *rotation band* of (σ, p, q) as $RB(\sigma, p, q) = \rho(f_\sigma)$. If (σ, p, q) is such that F_σ has exactly two turning points in each fundamental domain $[x, x+q)$ and these turning points are adjacent in $\mathbf{Z}$ (as in fig. 4), we say that (σ, p, q) is a *bimodal order stucture*.

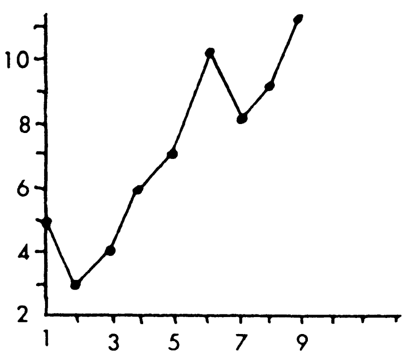


Fig. 4

Figure 4 shows the tight circle map of the order structure shown in Figure 1(b). The rotation band of this order structure is $[1/3, 1/2]$. This was shown in [BH] where the rotation band was used in a criterion for the nonexistence of invariant circles (*cf.* [Bd2] and [Bd3]).

The next definition formalizes the notion of two periodic orbits fitting together in a prescribed manner. Figure 5 shows one way in which a monotone $(1,2)$ and a monotone $(1,3)$ order structure can fit together to make what we call a monotone, joint $(1,3)-(1,2)$

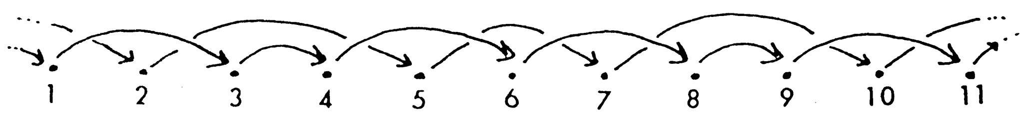


Fig. 5

structure. This joint order structure is closely related to the order structure of Figure 1(b) in a way we will explain shortly.

DEFINITION: A bijection $\tau : \mathbf{Z} \to \mathbf{Z}$ is called a *joint order structure* if it satisfies (1) and (2) in the definition of a (p, q) order structure with $p = r + m$ and $q = s + n$ and further:

(a) The projected map $\tau' : \mathbf{Z}/q\mathbf{Z} \to \mathbf{Z}/q\mathbf{Z}$ has exactly two periodic orbits which we denote as $o(x_1)$ and $o(x_2)$;

(b) If $\bar{\pi} : \mathbf{Z} \to /q\mathbf{Z}$ is the projection, τ restricted to $\bar{\pi}^{-1}(o(x_1))$ and $\bar{\pi}^{-1}(o(x_2))$ are required to be (r, s) and (m, n) order structures, respectively (after relabeling).

The bijection τ is called a *monotone* joint $(r, s) - (m, n)$ order structure if the restricted (r, s) and (m, n) order structures are both monotone. It is not difficult to show (using, for example, techniques of [Bd1] section 2) that there is exactly one *bimodal* monotone joint $(r, s) - (m, n)$ order structure when the integers within each pair are relatively prime. By a Markov partition arugument (as in [BGMY]) one may constuct a periodic orbit of period $s + n$ that "shadows" the pair of orbits described by the joint order structure. This periodic orbit behaves like a monotone (r, s)-orbit for s iterates and then like a monotone (m, n)-orbit for n iterates. The order structure of this orbit is called a $(r/s, m/n)$ *concatenation* order structure and is defined precisely below. It turns out to be more convenient to define the concatenation order structure in terms of how it can be transformed into a joint monotone order structure. In the example of Figure 1(b), the point 6 is a local maximum of the order structure (more precisely, of F_σ) and the point 7 is a local minimum. If ϕ is the map that transposes these two points (as well as their translates), then $\sigma \circ \phi$ is the order structure of Figure 5. This means that Figure 1(b) is what we call the $(1/3, 1/2)$ concatenation order structure.

DEFINITION: Given rationals in lowest terms $0 < r/s < m/n < 1$, let $p = r + m$ and $q = s + n$. The $(r/s, m/n)$ *concatenation order structure* is a bimodal (p, q)-order structure (σ, p, q) with the property that $\sigma \circ \phi$ is a monotone, joint $(r, s) - (m, n)$ order structure where $\phi : \mathbf{Z} \to \mathbf{Z}$ is defined using a local maximum, m_0, of F_σ as $\phi(n) = n + 1$ and $\phi(n + 1) = n$ for $n \equiv m_0 \bmod q$ and $\phi(n) = n$, otherwise.

One can show that these properties define a unique order structure whose orbit type which we denote as $ot(r/s, m/n)$. These concatenation order structures are well behaved in the sense discussed above. All the orbits dominated in one dimension are also dominated in two dimensions. One proves this by constructing the train track for the isotopy class rel the orbit and comparing the action on it to the tight circle map. This is the basis of the first three theorems of the next section. In Fig 6, we illustrate the invariant track and its image for the orbit type of Fig. 1(b).

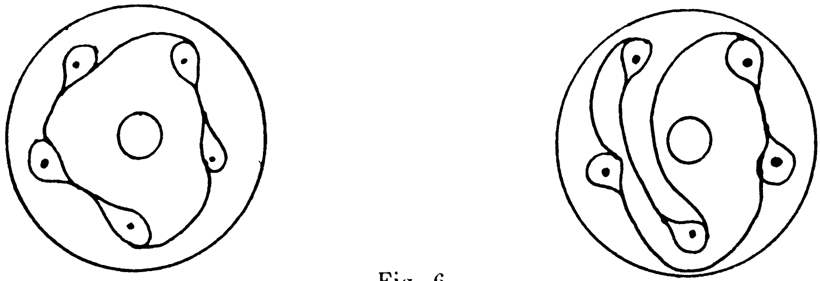


Fig. 6

Section 5. Results

THEOREM 4: If (σ, p, q) is a bimodal order structure and $0 < r/s < m/n < 1$ are fractions in lowest terms with $RB(\sigma, p, q) \subset (r/s, m/n)$, then $ot(\sigma, p, q) \leq ot(r/s, m/n)$.

The next theorem states that the partial order restricted to the concatenation orbits looks like set-theoretic inclusion.

THEOREM 5: If $0 < r/s < m/n < 1$ and $0 < r'/s' < m'/n' < 1$ are rationals in lowest terms then $(r'/s', m'/n') \subset (r/s, m/n)$ if and only if $ot(r'/s', m'/n') \leq ot(r/s, m/n)$.

The topological entropy of a map is denoted $h(f)$. It is a measure of the complication of the dynamical system induced by the map (see, for example, [Bo] for precise definitions). The proof of the next theorem relies on the fact that the entropy of a pseudo-Anosov map is a lower bound for the entropies of all maps isotopic to it ([FLP] exposé 11). Because of the nice property of the concatenation orbits noted above, the entropy of the tight circle map is the same as the entropy of the pseudo-Anosov prepresentative in the isotopy class. The entropy of the tight circle map can be computed using techniques of [BGMY].

THEOREM 6: If $0 < r/s < m/n < 1$ are rationals with $|rn - ms| = 1$ (such rationals are called Farey adjacent) and $ot(r/s, m/n) \in ot(f)$ for some $f \in \mathrm{Diff}_+(A)$ then the topological entroppy satisfies $h(f) \geq \log(\lambda)$ where λ is the biggest positive root of $(x^s - 2)(x^n - 2) = 3$. In particular,

$$h(f) \geq \frac{log(7 + 4\sqrt{3})}{n + s}.$$

Thus, for example, any twist map f that has a periodic orbit with the order structure illustrated in Figure 1(b) has entropy $h(f) \geq \log(\lambda) \doteq log(1.72208) \doteq .5435$ where we numerically estimated the largest positive root of $(x^3 - 2)(x^2 - 2) = 3$. The estimate in the last line of the theorem is a consequence of an elementary computation of a lower bound for λ.

We note that the previous three theorems apply to all diffeomorphisms $f \in \mathrm{Diff}_+(A)$. We have only used the monotone twist hypothesis to constuct orbit types out of order structures. The notion of orbit type is independent of the twist hypothesis. For example, a periodic orbit of an arbitrary diffeomorphism could appear quite complicated but actually be the same orbit type as a bimodal concatenation. We should also note that orbit type information does not give complete order structure data since the passage from order structure to orbit type was not one-to-one. This means that the conclusion that $ot(p'/q', m'/n') \in ot(f)$ from Theorem 5 does not imply that f has a periodic orbit with the $(p'/q', m'/n')$ concatenation order structure even if f is monotone twist.

The next theorem states that periodic orbits with monotone orbit types and the appropriate rotation numbers are always present when one has a bimodal orbit type with nontrivial rotation band. The corresponding fact about monotone order structures under the monotone twist hypothesis was proved in [Bd3] (*cf* [J2]). We note that one piece of the Aubry-Mather Theorem is that area preserving monotone twist maps always have periodic orbits with monotone (p, q)-order structures for all the expected rationals p/q. Also, Hall ([H1]) shows that a twist map with any p/q-periodic orbit has a monotone one.

THEOREM 7: If (σ, p, q) is a bimodal order structure and r/s is a fraction in lowest terms with $r/s \in RB(\sigma)$, then $ot(\sigma_B, r, s) \le ot(\sigma, p, q)$ where (σ_B, r, s) is the monotone (r, s)-order structure.

The hypothesis that (σ, p, q) is bimodal is probably unnecessary. The proof of this theorem uses different techniques than those discussed previously. By "fattening up" the tight circle map as in the appendix of [Fk], one obtains an Axiom A representative in the appropriate isotopy class. One then shows directly that the desired orbits satisfy the hypothesis of [AF] and are thus unremovable under isotopy.

Our final conjecture concerns the question posed in the introduction. What happens to the dynamics of an area-preserving, monotone twist map when an invariant circle breaks? It is known that the loss of an invariant circle is always accompanied by the birth of nonmonotone periodic orbits whose rotation bands include the rotation number of the broken circle ([BH], [H2], [J2]). The order in which these new orbits are born is, of course, restricted by the partial order on OT. Our conjecture is that the simplest nonmonotone orbit types, namely, those arising from bimodal concatenations, must be among the orbits born.

CONJECTURE: If f is an area-preserving monotone twist map that has no invariant circle with rotation number equal to the irrational ω, then there exists an $n \in \mathbf{N}$ with $ot(p_n/q_n, p_{n+1}/q_{n+1}) \in ot(f)$ where p_n/q_n is the n^{th} convergent of the continued fraction of ω.

Further, if f_μ is a generic, one parameter family of area preserving monotone twist maps in which f_0 has an invariant circle with rotation number ω, which is absent when $\mu > 0$, then $n(\mu) \to \infty$ monotonically as $\mu \to 0$ for $\mu \in (0, \epsilon)$ for some $\epsilon > 0$.

If the first sentence of the conjecture is true, one could then use Theorems 4 and 5 to infer the existence of other orbits. The proof of the second sentence will involve techniques different from those discussed here. It appears likely that this conjecture (if true) could be proved using the constrained variational techniques recently developed by Mather.

REFERENCES

[AF] Asimov, D. and Franks, J., Unremoveable closed orbits, *Springer LNIM*, 1007, 1983

[Ba] Baldwin, S., Generalizations of a theorem of Sharkovski's on orbits of real valed functions, preprint.

[Be] Bernhardt, C., The ordering on peremutations induced by continuous maps of the real line, *Ergod. Th. and Dynam. Sys.*,7, 155-160, 1987.

[Bn] Birman, J., *Braids, Links, and the Mapping Class Group*, Ann. of Math Studies, 81, Princeton Univ. Press, 1974.

[BS] Birman, J. and Series, C.,Algebraic linearity for automorphisms of a surface, preprint.

[BW] Birman, J. and Williams, R., Knotted periodic orbits in dynamical systems II: knot holders for fibered knots, *Cont. Math.*, 20, 1983.

[BF] Blanchard, P. and Franks, J., The dynamical complexity of orientation reversing homeomorphisms of surfaces, *Inv. Math.*, 62, 333-339, 1980.

[BGMY] Block, L., Guckenheimer, J., Misiurewicz, M. and Young, L.S., Periodic points and topological entropy of one dimmensional maps, *Springer LNIM*, 819, 1980.

[BlHt] Block,L.and Hart, D., Orbit types for maps of the interval, *Ergod. Th. and Dynam. Sys.*, 7, 161-164, 1987.

[BH] Boyland, P. and Hall, G.R., Invariant circles and the order structure of perioidic orbits in monotone twist maps, *Topology*, 26, 21-35, 1987.

[Bd1] Boyland, P.,Bifurcations of circle maps: bistability, and rotation intervals, *Comm. Math. Phys.*, 106, 353-381, 1986.

[Bd2] Boyland, P., Rotation bands and invariant circles in monotone twist maps, *Comm. Math. Phys.*, to appear.

[Bd3] Boyland, P., Rotation Sets and Morse decompositions for twist maps, *Erg. Th. and Dyn. Sys.*, to appear.

[Bo] Bowen, R., Entropy and the fundamental group, *Springer LNIM*, 668, 1978.

[Br] Brunovsky, P., On one-parameter families of diffeomorphisms I and II, *Comment. Math. Univ. Caroline*, 11, 559-582, 1970 and 12, 765-784, 1970.

[C] Casson, A., Automorphisms of surfaces after Nielsen and Thurston, lecture notes by S. Bleirer.

[E] Eckman, J.P., *Rev. Mod. Phys*, 53, 643-661, 1981.

[Fk] Franks, J., Knots, Links and Symbolic Dynamics, *Ann. of Math.*, 113, 529-552, 1981.

[FLP] Fathi, A., Laudenbach, F. and Poenaru, V., Travaus de Thurston sur les surfaces, *Asterique*, 66-67, 1979.

[Fr4] Fried, D., Entropy and twisted cohomology, preprint.

[Ha] Hagelin, J., Is consciousness the unified field? A field theorist's perspective, *Vedic Sci. and Mod. Sci.*, 1, 29-87, 1987.

[H1] Hall, G.R., A topological version of a theorem of Mather on twist maps, *Erg. Th. and Dyn. Sys.*, 4, 585-603, 1984.

[H2] Hall, G.R., A topological version of a theorem of Mather's on shadowing in twisi maps, preprint.

[Hd] Handel, M., The entropy of orientation reversing homeomorphisms of surfaces, *Topology*, 21, 291-296, 1982.

[HP] Harer, J. and Penner, R., Combinatorics of train tracks, preprint.

[HW] Holmes, P. and Williams, R., Knotted periodic orbits in the suspension of Smale's horseshoe, *Arch. Rat. Mech. Anal.*, 90, 115-194, 1985.

[I] Ito, Rotation sets are closed, *Math. Proceed. Cambridge Ohilos. Soc.*, 89, 107-110, 1981.

[J1] Jungries,I.,Some results on the Sharkovski ordering of permutations, preprint.

[J2] Jungries, I., Some results on periodic orbits of twist maps, preprint.

[K] Katok, A., Some remarks on the Birkoff and Mather twist theorems, *Ergod. Th. and Dynam. Sys.*, 2, 185-194, 1982.

[Ku] Kuhn, N., Train track notes.

[Mc] MacKay, R., Introduction to the dynamics of area-preserving maps, Proceedings of the US Particle accelerator school, SLAC, 1985, to appear.

[Ma] Mather, J., More Denjoy minimal sets for area preserving diffeomorphisms, *Comment. Math. Helvetci*, 60, 508-557, 1985.

[Ms1] Matsuoka, T., The number and linking of periodic orbits, *Inv. Math.*, 70, 319-340, 1983.

[Ms2] Matsuoka, T., On a generalization of Sharkovski's theorem to maps of the disk, preprint.

[Mr] Moser, J., Recent developments in the theory of Hamiltonian systems, *SIAM Review*, 28, 459-485, 1986.

[Sh] Sharkovski, A., Coexistence of cycles of a continuous map of a line into itself, *Ukr. Mat. Z.*, 16, 61-71, 1964.

[St] Stefan, P., A theorem of Sharkovski on the existence of periodic orbits of continuous endomorphisms of the real line, *Comm. Math. Phys.*, 54, 237-248, 1977.

[T] Thurston, W., On the geometry and dynamics of surface difffeomorphisms, preprint

Department of Mathematics
Maharishi International University
Fairfield, IA 52556

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986262

Contemporary Mathematics
Volume **81**, 1988

SOME PROBLEMS ON DYNAMICS OF ANNULUS MAPS

Glen R. Hall*

In this exposition we will discuss some recent results on the dynamics of maps of the annulus. Since the health of an area is closely related to the availability of open problems, we will organize what follows around a collection of problems. Personal preference has guided the choice of these problems and the discussions which follow and hence there is a heavy bias towards "topological" theorems and techniques. Excellent discussions of the "variational" approach and its relation to differential geometry can be found in Bangert [Ba1] and Moser [M1]. Many of these questions arose in a topics class in twist maps taught by the author in Winter, 87 at the University of Minnesota. Thanks are due for the comments and the patience of those that attended: D. Aronson, M. Jolly, R. McGehee, R. Moeckel, D. Norton and B. Peckham. Also conversations with P. Boyland, C. Golé and D. Goroff have been a useful and most enjoyable part of thinking about these problems.

Notation: We collect here a small amount of notation which will be used throughout.

Let $\mathcal{A} = \mathbf{R}/\mathbf{Z} \times [0,1]$ be the annulus $\mathcal{C} = \mathbf{R}/\mathbf{Z} \times \mathbf{R}$ be the cylinder. They have covers $A = \mathbf{R} \times [0,1]$ and $\mathbf{R}^2$ respectively with natural projections: $\prod : A \to \mathcal{A}$ (or $\mathbf{R}^2 \to \mathcal{C}$). A *lift* of a point $\zeta \in \mathcal{A}$ will be any point in $\prod^{-1}(\zeta)$. A *lift* of a subset $\mathcal{B} \subseteq \mathcal{A}$ will be the set $\prod^{-1}(\mathcal{B})$.

We consider maps $f : \mathcal{A} \to \mathcal{A}$ which have lifts $\bar{f} : A \to A$ satisfying the following assumptions which hold throughout the paper

$$(1)\ f \text{ is degree one (i.e. } \forall (x,y) \in A,\ f(x+1,y) = f(x,y) + (1,0))$$

* Partially supported by the Sloan Foundation.

(2) f is injective and orientation preserving.

We let π_1, π_2 denote the usual projections onto coordinates, i.e. $\pi_1(x,y) = x$, $\pi_2(x,y) = y$. Given $f : \mathcal{A} \to \mathcal{A}$ and $\zeta \in \mathcal{A}$ we define the rotation number of ζ by choosing a lift $z \in A$ of ζ and setting

$$\text{rotation } \# \text{ of } \zeta \text{ under } f = \rho(z,f) = \lim_{n\to\infty} \frac{\pi_1 \bar{f}^n(z)}{n} \text{ if it exists}$$

where $\bar{f}$ is a lift of f. For definiteness we take $\bar{f}$ so that

$$\pi_1 \bar{f}(0,0) \in [0,1).$$

We note that a periodic point ζ of period q for a map $f : \mathcal{A} \to \mathcal{A}$ will satisfy $f^q(\zeta) = \zeta$, so there will exist $p \in \mathbf{Z}$ so that if $z \in A$ is any lift of ζ then

$$\bar{f}^q(z) = z + (p, 0)$$

and the rotation number of ζ will be p/q. Finally we give the following definition of two special types of maps.

Definition: A map $f : \mathcal{A} \to \mathcal{A}$ will be called a *monotone twist map* if it is a diffeomorphism (satisfying 1 and 2 above) and it satisfies $\dfrac{\partial(\pi_1 \circ f)}{\partial y} > 0$.

A diffeomorphism $f : \mathcal{A} \to \mathcal{A}$ will be called a *boundary twist map* if $\rho((0,0),f) \neq \rho((0,1),f)$.

Hence the monotone twist condition says the change in the angular component increases as radial component increases while a boundary twist condition says only that one boundary rotates more than the other. We can display these conditions by looking at the images of typical radial segments in $\mathcal{A}$ or in A

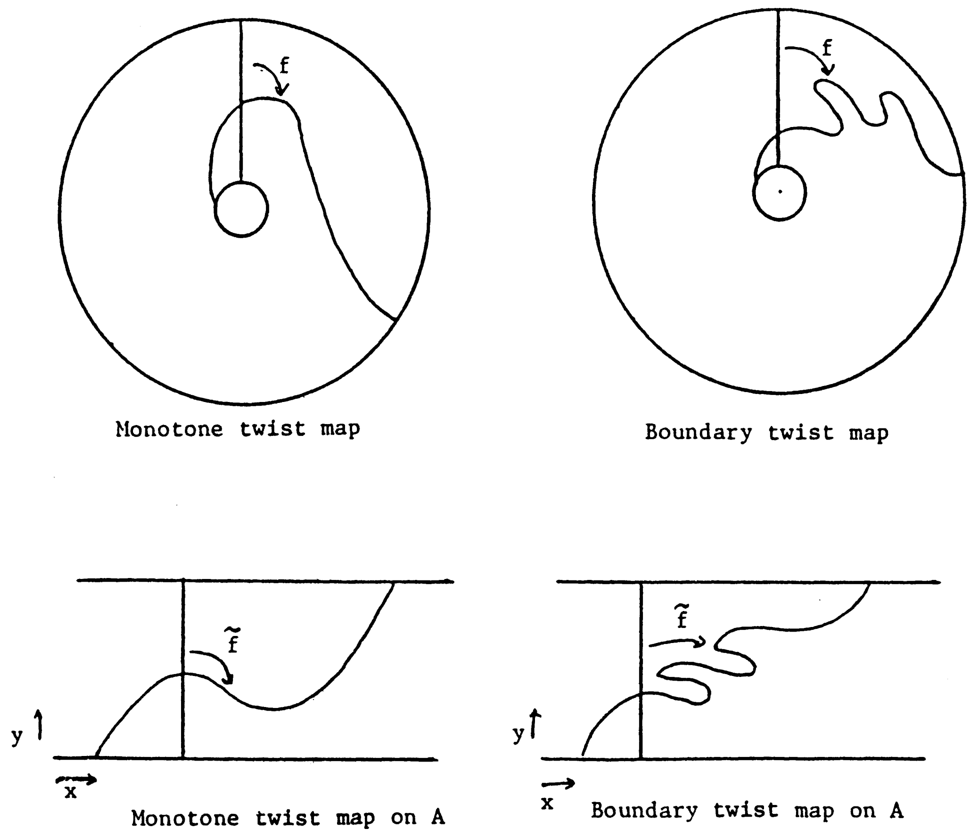


Fig. 0

Problem 1: Given a map $f : \mathcal{A} \to \mathcal{A}$ with no periodic orbits and a neighborhood U of f in the analytic topology, does U contain a map $g : \mathcal{A} \to \mathcal{A}$ which has a periodic orbit? (Are maps with periodic orbits dense in the analytic topology?)

There are many theorems on periodic orbits of annulus maps dating back at least as far as Poincaré and Birkhoff (see [B1]). The recent work of Franks [F1,F2] has given some of the best insight into the existence of fixed points. We state an example:

THEOREM (Franks [F1]): Suppose $f : \mathcal{A} \to \mathcal{A}$ is a homeomorphism and Λ is a chain transitive compact invariant set for f then if $\tilde{\Lambda}$ is the lift of Λ to A either

(1) $\tilde{f}$ has a fixed point (so f has a fixed point), or

(2) $\lim_{n\to\infty} \pi_1(\tilde{f}^n(z) - (z)) = +\infty$ for all $z \in \Lambda$ or $-\infty$ for all $z \in \Lambda$ and the limits are uniform in z.

COROLLARY (Franks [F1]): If f has no periodic points then for Λ a chain transitive compact invariant set, every point $\zeta \in \Lambda$ has the same irrational rotation number.

The motto seems to be that if there is recurrent behavior in the map (which can't be avoided by compactness) and points in the same recurrence component are moved by the map on the lift at (asymptotically) different speeds left or right then there must be periodic points. Hence there seem to be strong restrictions on maps without periodic points. In particular, if we take $f : \mathcal{A} \to \mathcal{A}$ a diffeomorphism then f will preserve the boundary circles of $\mathcal{A}$. If for any $\omega \in \mathbf{R}$ we let $R_\omega : \mathcal{A} \to \mathcal{A}$ be the map whose lift is given by $\tilde{R}_\omega(x,y) = (x+\omega, y)$ then there will exist arbitrarily small ω such that $R_\omega \circ f$ has periodic orbits (on the boundaries). When f is not a diffeomorphism onto $\mathcal{A}$, but has $f(\mathcal{A}) \subseteq$ interior $(\mathcal{A})$ then we might hope that the attractor formed by $\bigcap_{n\geq 0} f^n(\mathcal{A})$ will retain enough of the topology of the circle (Conley [C1]) or at least enough one dimensional character that a one dimensional approach could produce periodic points (see Young [Y1]). However, the map on the resulting attractor and the attractor itself can be quite non-standard (see for example, Handel [H2], Hall [Hl], Aronson et al. [A1]), motivating the following:

Problem 1A: Given a diffeomorphism $f : \mathcal{A} \to \mathcal{A}$ with no periodic points and a neighborhood U of f in the analytic topology is there a diffeomorphism $g : \mathcal{A} \to \mathcal{A}$ in U with interior periodic points?

Problem 1B: Given a map $f : \mathcal{A} \to \mathcal{A}(f(\mathcal{A}) \subseteq \mathcal{A})$ with no periodic points is there an (arbitrarily small) $\omega \in \mathbf{R}$ such that $R_\omega \circ f$ has periodic points (where $\tilde{R}_\omega(x,y) = (x+\omega,y)$ as above)?

We note that if we replace "analytic" with "C^1" in the above then the answers to 1 and 1A are yes by the C^1 closing lemma ([P]). We don't know the answers to the above in the C^2 case and recent work of Grutierenze [G1] indicates the situation for the C^2 closing lemma is quite delicate. Of course the above questions do not require we "close" a particular orbit so they do not imply the closing lemma.

If we add additional hypotheses to the mapping $f : \mathcal{A} \to \mathcal{A}$ then the situation regarding fixed points becomes more clear. For example, Boyland [Bd1] has classified fixed point free twist maps of $\mathcal{A}$. He shows that for twist maps the hypotheses of Franks' theorem may be considerably weakened.

THEOREM (Boyland [Bd1]). *If $f : \mathcal{A} \to \mathcal{A}$ is a twist map (or positive tilt map) and there exist $z_1, z_2 \in A$ such that for $\tilde{f}$ the lift of f we have*

$$\pi_1(z_1) < \pi_1(\tilde{f}(z_1)) \textit{ and } \pi_1(z_2) > \pi_1(\tilde{f}(z_2))$$
$$\pi_1(\tilde{f}(z_1)) > \pi_1(\tilde{f}^2(z_1)) \qquad \pi_1(\tilde{f}(z_2)) < \pi_1(\tilde{f}^2(z_2)) \text{ (See Fig.1)}$$

then f has a fixed point.

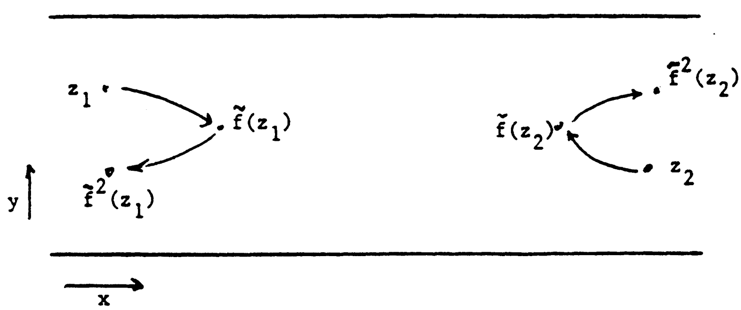


Fig. 1

The proof shows that the box in Figure 1 connecting the boundaries between $\tilde{f}(z_1)$ and $\tilde{f}(z_2)$ will contain a set which has non-zero index under $\tilde{f}$. The twist condition forces the topological complication to appear after just one iteration. The topological approach exhibited here could be useful for the other problems and we will return to it later.

Finally, if we assume $f : \mathcal{A} \to \mathcal{A}$ is area preserving then constructing examples of periodic point free maps (besides rotations by irrationals) becomes quite difficult –See Anosov-Katok [A,K], Fathi-Herman [F H] and Handel [H1].

Problem 2: Given a diffeomorphism $f : \mathcal{A} \to \mathcal{A}$ show $\{\rho(z, f) : z \in A\}$ is closed.

[Added in proof: M. Handel has proven that the rotation set is closed via techniques of [H3].]

The spark for much recent work in the dynamics of annulus maps and twist maps in particular is the theorem of Aubry and Mather which states

THEOREM (Aubry, et al. [Ay1], Mather [Ma1]): If $f : \mathcal{A} \to \mathcal{A}$ is an area preserving monotone twist map then $\{\rho(z, f) : z \in A\}$ is a closed interval.

The above problem is a natural generalization of the theorem (this question is also posed by Botelho [Bo]). From Franks' theorem (or the Poincaré Birkhoff theorem) such an f has periodic orbits of all rational rotation numbers between the rotation number of f on the boundary of $\mathcal{A}$. Hence the new information in the Aubry-Mather theorem is the existence of points with irrational rotation number. (These are the "ghosts" of invariant circles of the KAM theory, see Moser [M2].)

The original proofs of Aubry and Mather were variational, i.e., they defined an energy function related to the given map on a space of possible orbits, then showed that the minimum of this energy was an orbit of the map. Moreover, this minimum orbit could be found for any rotation number and had very nice properties which we discuss below (see Katok [K1]). Very similar ideas had been used by Hedlund in studying geodesics on tori (see Bangert [Ba1]). It is unclear how to extend the above approach to maps without a twist condition and/or without an invariant measure. Using topological techniques one can recover versions of these theorems without the area preservation hypothesis which points the way for generalizations to maps without twist conditions.

First we need to define some special types of orbits for a map $f : \mathcal{A} \to \mathcal{A}$.

Step 1: Given a diffeomorphism $f : \mathcal{A} \to \mathcal{A}$, find a suspension ϕ of f, i.e. find a map $\phi : \mathcal{A} \times \mathbf{R} \to \mathcal{A}$ such that ϕ is continuous and

$$\phi : \mathcal{A} \times \{0\} \to \mathcal{A} \text{ is the identity}$$

$$\phi : \mathcal{A} \times \{1\} \to \mathcal{A} \text{ is equal to } f$$

$$\phi(z,t) = \phi(f^{[t]}(z), t - [t]) \text{ where } [\cdot] \text{ denotes the greatest integer function}$$
$$\text{and } \phi : \mathcal{A} \times \{t\} \to \mathcal{A} \text{ is a diffeomorphism for every } t \in [0,1].$$

Step 2: Let $\tilde{f} : A \to A$ be the lift of f and $\tilde{\phi} : A \times [0,1] \to A$ be the lift of ϕ. We can display $\tilde{\phi}$ as in Fig. 2.

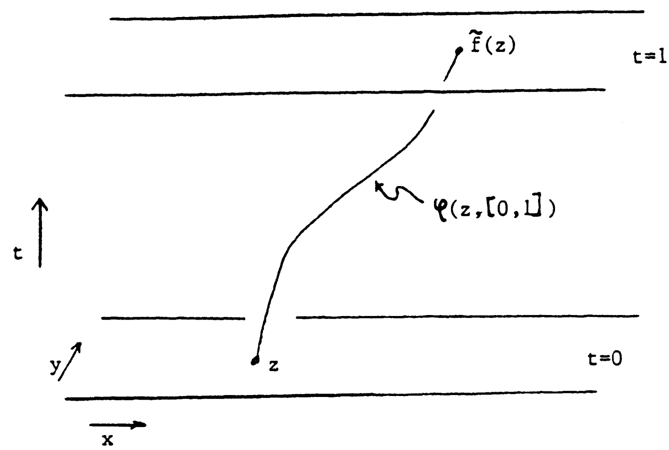


Fig. 2

where (x, y) are the coordinates in A and t is the new time coordinate in $[0,1]$. (So f is the "time one" map of ϕ.)

Step 3: Define the extended orbit of $z \in A$ by

$$eo(z) = \{\tilde{f}^i(z) + (j,0) : i,j \in \mathbb{Z}\}.$$

Now we may define a monotone periodic orbit as follows.

Definition If $\zeta_0 \in \mathcal{A}$ is a p/q periodic point of f and $z_0 \in A$ is a lift of ζ_0 (so $f^q(z_0) = z_0 + (p,0)$) then we call ζ_0 a *monotone* p/q periodic orbit if the curves

$$\gamma_z = \{\{(\phi(z,t),t) : t \in [0,q]\} : z \in eo(z_0)\} \subseteq A \times [0,q]$$

(i.e., the flow lines of points in $eo(z_0)$ under ϕ) are unlinked.

Here by "unlinked" we mean the following:

If we let $\{\tilde{\gamma}_z : z \in e0(z_0)\}$ be the image of the curves γ_z under the map $h : A \times [0,q] \to A \times [0,q] : (x,y,t) \to (x - t \cdot p/q, y, t)$ then the end points of each $\tilde{\gamma}_z$ are directly above each other. If we can deform the $\tilde{\gamma}_z$'s to straight vertical lines, leaving the end points fixed, without intersectng any two of them then they are called unlinked and the associated orbit is monotone. A picture makes this clearer, here $q = 2,\ p = 1$

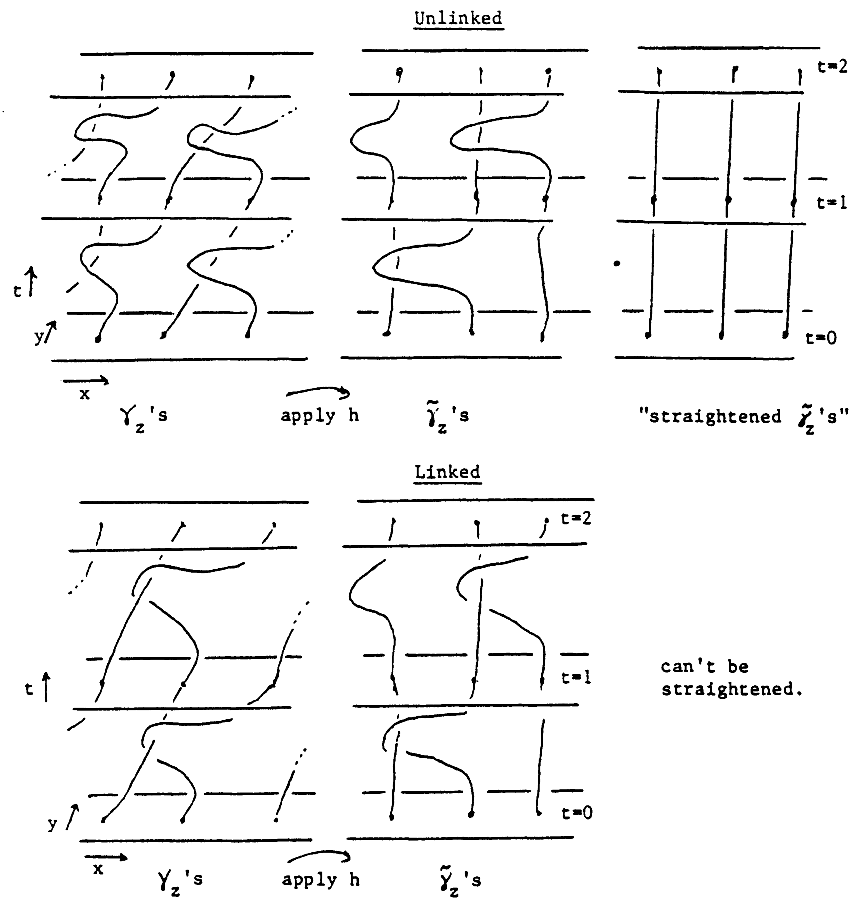


fig. 3.

Of course the definition above requires proof that the notion of monotone is independent of the choice of suspension, an instructive exercise for the reader. The remarkable fact about monotone twist maps is that an equivalent definition of monotone orbit can be given which completely avoids the construction above.

THEOREM (see Hall [Hl1]): If $f : \mathcal{A} \to \mathcal{A}$ is a monotone twist map with p/q periodic point $\zeta \in \mathcal{A}$ and $\tilde{f} : A \to A$, $z \in A$ are lifts of f and ζ then ζ is monotone if and only if

$$\forall z_1, z_2 \in eo(z), \pi_1(z_1) < \pi_1(z_2) \Rightarrow \pi_1(\tilde{f}(z_1)) < \pi_1(\tilde{f}(z_2)) \qquad (*)$$

Hence we can use condition $(*)$ as a definition of monotone.

This condition is equivalent to saying that $\tilde{f}$ preserves the x-coordinate ordering on $eo(z)$. Since condition $(*)$ makes sense on any orbit we say

Definition: Given $f : \mathcal{A} \to \mathcal{A}$ a monotone twist map with lift $\tilde{f}$, we say $\zeta \in \mathcal{A}$ is *monotone* if $(*)$ is satisfied for $z \in A$ any lift of ζ.

Monotone orbits behave nicely under limits, i.e., we have the following facts (see Katok [K1]):

(1) If $f : \mathcal{A} \to \mathcal{A}$ is a monotone twist map, $\{\zeta_i\}_{i=1}^{\infty}$ a sequence of monotone points and $\zeta_i \to \zeta_0$ as $i \to \infty$ then ζ_0 is monotone,

and

(2) the rotation number exists for every monotone point and $\rho(\zeta_i, f) \to \rho(\zeta_0, f)$ as $i \to \infty$

Hence to prove the Aubry, Mather theorem it suffices to find monotone periodic points of all rotation numbers, then take limits. One can either use the variational techniques and show that the minimal energy orbit of a given rotation number for an area preserving monotone twist map is infact monotone. Alternately, one can use a more topological approach to show

THEOREM (Hall [Hl1]): If $f : \mathcal{A} \to \mathcal{A}$ is a monotone twist map and f has a p/q periodic point (p, q relatively prime) then f has a p/q monotone periodic point.

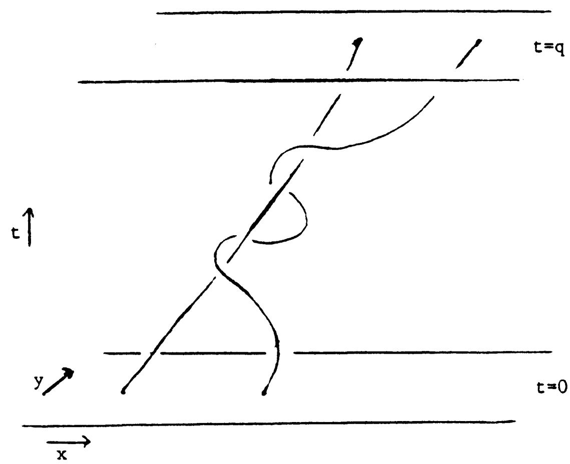
Using the fixed point theorems already available for area preserving monotone twist maps, the above provides all the required monotone orbits (see also Le Calvey [LC]).

The definition of monotone given earlier does not require a twist condition, hence we can ask

Problem 2A: If $f : \mathcal{A} \to \mathcal{A}$ has a p/q periodic point does it have a monotone p/q-periodic point?

Problem 2B: If $f : \mathcal{A} \to \mathcal{A}$ has a sequence of monotone periodic points $\{\zeta_i\}_{i=1}^{\infty}$ what can be said about their limit points?

For example, each ζ_i in 2B might look complicated (see Fig. 3) but actually can be "straightened out". However, if the rotation numbers p_i/q_i of ζ_i have $q_i \to \infty$ as $i \to \infty$ then the limit point of the ζ_i's might have a quite complicated orbit (see the example of Handel [H1]).



this is "unlinked"

Fig. 4

Finally we note that another possible definition of monotone (see Mather [Ma2]) is the following.

Definition: If $f : \mathcal{A} \to \mathcal{A}$ is a diffeomorphism and ζ a p/q periodic point of f then ζ is called monotone if there exists a one-parameter family of diffeomorphisms $f_t : \mathcal{A} \to \mathcal{A}$ and a one-parameter of points ζ_t for $t \in [0,1]$, continuous in t such that $f_1 = f$, $\zeta_1 = \zeta$, for each t, ζ_t is a p/q point for f and f_0 is a map with lift of the form $(x,y) \to (x + cy, y)$ for some constant $c \in \mathbf{R}$.

The property of "linked" or "unlinked" isolates a set of orbits (the monotone orbits) from others and only monotone orbits occurs near simple maps like the integrable map above (where "near" depends on the rotation number). Hence the feeling is that a fixed point theorem (e.g. Poincaré's Last Geometric Theorem) should give this simplest fixed point. This leads to the following not very well posed problem

Problem 2C: Given a path connected set of maps X for which there is a fixed point theorm (i.e. every map in X has a fixed point and "one proof works for all of X") then for $f, g \in X$, does there always exist fixed points z for f, w for g and a one parameter family of maps f_t in X with fixed points z_t such that $f_0 = f$, $z_0 = z$ and $f_1 = g$, $z_1 = w$?

An affirmative answer to the above with X = area preserving diffeomorphisms of $\mathcal{A}$ with a boundary twist condition and Poincaré's Last Geometric theorem as the fixed point theorem would imply the Aubry, Mather theorem.

Problem 3: Given a (monotone twist) map $f : \mathcal{A} \to \mathcal{A}$ with a non-monotone p/q periodic point what other periodic orbits must f have?

There has been considerable progress on this problem recently by Boyland [Bd2] and Jungries [J1,J2]. Boyland in particular uses techniques of "Thurston train tracks" so that the twist conditions do not play as vital a role. We refer the reader to Boyland [Bd2] for the latest developments, noting only that the motto "topologically complicated orbits imply lots of periodic orbits" should be very useful in problems 1, 1A-B.

Problem 4: Given $f : \mathcal{A} \to \mathcal{A}$ a monotone twist map with lift $\tilde{f} : A \to A$ what is the minimal set $\mathcal{B} \subseteq \mathcal{A}$ with lift $B \subseteq A$ such that $\forall z \in A \exists w \in B : |\tilde{f}^n(z) - \tilde{f}^n(w)| < 1$ for all $n \in \mathbf{Z}$?

That is, what is a description of a set which "globally shadows" every orbit of f (see Handel [H2]). Since homoclinic and heteroclinic behavior are to be expected in twist maps there could be many points without rotation numbers and whose orbits move about the annulus or along the strip in wild ways (see Aronson, et al. [A1], Hockett, Holmes [HH]). The set $\mathcal{B}$ would "catch all the rotation" behavior of the map.

For area preserving monotone twist mappings a candidate for $\mathcal{B}$ would be "locally energy minimizing orbits." Mather [see Hl2] has shown that any sequence of periodic orbits in an area preserving monotone twist map without invariant circles (see below) can be "globally shadowed" by locally energy minimizing orbits. The existence of such global shadows in twist maps can also be implied topologically as follows:

We say a monotone twist map $f : \mathcal{A} \to \mathcal{A}$ with lift $\tilde{f} : A \to A$ satisfies condition B if for every $\epsilon > 0$ there exist $z_1, z_2 \in A$ and $n_1, n_2 > 0$ such that $\pi_1(z_1) < \epsilon$, $\pi_2(z_2) > 1 - \epsilon$ and $\pi_2(\tilde{f}^{n_1}(z_1)) > 1 - \epsilon$, $\pi_2(\tilde{f}^{n_2}(z_2)) < \epsilon$.

THEOREM (Hall [Hl2]): Suppose $f : \mathcal{A} \to \mathcal{A}$ is a monotone twist map and satisfies condition B and has lift $\tilde{f} : A \to A$. Let $\{\zeta_i\}_{i=-\infty}^{\infty}$ be a sequence of monotone periodic points (see Problem 2) and $\{n_i\}_{i=-\infty}^{\infty}$ a sequence of positive integers. Then there exists $\zeta \in \mathcal{A}$ with a lift $z \in A$ such that for each i, there exists K_i and z_i a lift of ζ_i with $|\tilde{f}^j(z_i) - \tilde{f}^{K_i+j}(z)| < 1$ for $j = 0, \ldots, n_i$, and $K_{i+1} - K_i > n_i$ for all i.

That is, a sequence of monotone periodic orbits can be "globally shadowed." The proof of the above theorem indicates that a possible candidate for the set $\mathcal{B}$ or its lift B is the following:

$$\{z \in A : \forall n, D\tilde{f}^n(z)\binom{0}{1} \cdot \binom{1}{0} \geq 0 \text{ for all } n \in \mathbf{Z}\}$$

i.e., the set of points whose vertical tangent vector always has image which points to the right. (This set is used for showing nonexistence of invariant circles (see MacKay and Percival [MP]). An obvious obstruction to building orbits which wander about the annulus is the existence of invariant curves which separate the annulus into invariant regions, hence we give the following;

Definition: A map $f : \mathcal{A} \to \mathcal{A}$ will be said to have an *invariant circle* if there exists a set $\Gamma \subseteq$ interior of $\mathcal{A}$ homeomorphic to S^1 which is homotopically nontrivial in $\mathcal{A}$ and $f(\Gamma) = \Gamma$.

An area preserving monotone twist map with no invariant circles will automatically satisfy condition B and infact even stronger versions of condition **B** (see Le Calvey [LC2]).

There are several very important questions concerning area preserving monotone twist maps which may be related to the above, for example

Problem 4A: For generic area preserving monotone twist mappings are "elliptic islands" (invariant sets surrounding elliptic periodic orbits) dense?

Problem 4B: Does an area preserving monotone twist map without invariant circles have an invariant set of positive measure on which the map is ergodic?

These measure theoretic questions are of a different flavor than the above, but an understanding of possible choices might lead to some insight (or not). See Wojtkowski [W1,W2] for recent progress on these problems.

Finally, we remark that it is possible to think of much weaker "shadowing sets" $\mathcal{B}$, i.e. to require that the shadows don't have to stay as close. So we might ask

Problem 4C: Given $f : \mathcal{A} \to \mathcal{A}$ with lift $\tilde{f} : A \to A$ and $\alpha \in (0,1)$ can we describe sets $\mathcal{B}_\alpha \subseteq \mathcal{A}$ with lift $B_\alpha \subseteq A$ such that $\forall z \in A$, $\exists w \in B_\alpha : |\tilde{f}^n(z) - \tilde{f}^n(w)| < |n|^\alpha$ for all $n \in \mathbf{Z}$

Such sets $\mathcal{B}_\alpha$ would catch the rotation numbers, but miss some of the subtle twisting (see the example of Handel [H1]).

Problem 5: For the "standard" one parameter family $s_k : \mathcal{C} \to \mathcal{C}$ with lift $\tilde{s}_k : \mathbf{R}^2 \to \mathbf{R}^2$ given by $\tilde{s}_k(x,y) = (x + y + \frac{k}{2\pi}\sin(2\pi x), y + \frac{k}{2\pi}\sin(2\pi x))$ what is the largest value of k^* for which s_k has an invariant circle?

(Here, by invariant circle we mean a homotopically nontrivial invariant curve, as defined above.)

The KAM-theory implies that this map will have many invariant circles for k small, but these circles will disappear as k grows until they are all gone (see Moser [M3], Mather

[Ma3], Herman [He1]). The solution of this problem which would be best is a "closed form" expression for k^* (e.g., k^* is the value of some definite integral or the root of some polynomial) and an answer of this form is probably extremely difficult to obtain. There has been considerable numerical work estimating the value of k^* (see MacKay, Percival [MP], Greene [G], Olvera, Simo [OS]) using various techniques. Recently, Jungries [J2] has shown that it suffices to show that s_k has certain types of pseudo-orbits to show that s_k does not have invariant circles. This is related to the topological criterion for non-existence of invariant circles and existence of monotone (see problem 2) periodic points.

THEOREM (Boyland,Hall [BH]): An area preserving monotone twist map $f : \mathcal{A} \to \mathcal{A}$ does not have an invariant circle with irrational rotation number ω if and only if f has a non-monotone p/q periodic orbit for p/q some convergent of the continued fraction of ω.

This allows a quantitative relationship to be established between existence of non-monotone periodics and non-existence of invariant circles. Other quantitative techniques to show non-existence of invariant circles have been derived by Mather [Ma3,Ma4,Ma5] via variational techniques.

The subtle interplay between smoothness and number theory which arises in the proof of existence of invariant circles (KAM theory) for "near integrable" maps (see Moser [M2], [He1]) is discouraging for topological techniques on problem 5, however, related problems have a more topological flavor. The following closely related problems may also be reminiscent of recent results in complex dynamics (see [DH]).

Problem 5A: For the standard family $s_k : C \to C$, is the set of k such that s_k has an invariant circle an interval?

Problem 5B: For the standard family does the area of each "zone of instability" (i.e. the annular region between two invariant circles containing no invariant circles) grow monotonically with k?

Recent results of Bullett [Bu] show the answers to 5A and B are no for a piecewise linear model of the standard family!

Problem 6: Do any of the above results hold for the analogous higher dimensional maps on products of annulii?

This is almost too vague to be meaningful, which is an accurate indication of the state of affairs. First it is clear that we must be clear about "analogous, higher dimensional" maps, and we will surely have to add conditions which restrict more strongly the type of maps we consider.

We note that the variational arguments of Aubry and Mather seem to have their natural generalization in the work of Bangert and Moser [Ba1,M1,M3] on variational problems. This connects well with the work of Hedlund on geodesics [Ba1].

Another generalization for area preserving monotone twist mappings of the annulus or cylinder is exact symplectic mappings f of $(\mathbf{R}/\mathbf{Z})^n \times \mathbf{R}^n = (\mathbf{R}/\mathbf{Z} \times \mathbf{R})^n$ or a product of cylinders (see Conley Zehnder [CZ]). The condition analogous to the twist condition is that the projection of $f(\{\underline{\theta}\} \times \mathbf{R}^n)$ onto $(\mathbf{R}/\mathbf{Z})^n$ is a diffeomorphism. While the condition of being exact symplectic is much stronger than area preservation in higher dimensions, it turns out that time one maps of Hamiltonian systems satisfy this condition automatically and hence it has many applications.

Among the theorems which are known for these exact symplectic twist maps are the following:

(1) KAM theory remains true, i.e. perturbations of integrable maps have many invariant tori. However, these tori no longer separate the space and there is the chance for orbits to wander long distances around these tori (i.e. "Arnol'd diffusion", see Arnold and Avez [AA]).

(2) The Poincaré's Last Geometric fixed point theorem has its analog in the Birkhoff-Lewis and particularly the recent Conley-Zehnder theorem [CZ].

(3) A few "regularity" results are known for orbits (Bernstein and Katok [BK]) which state that not all periodic orbits and invariant tori can be situated in the space in a completely arbitrary manner.

The first step to generalizing the theorems above is to decide on the proper definitions of, e.g., monotone orbit, and so forth. This is far from a trivial problem since we have lost the topological restrictions for maps on the (much smaller) annulus. However, there is still topological information available. For example, an invariant torus in an exact symplectic map will form after suspension by a flow (see problem 2) a three dimensional invariant set in five dimensional space and hence there is a possibility that orbits of points might "link" with it.

Perhaps the only thing which is clear concerning these higher dimensional maps is that they will provide many interesting problems in the future.

References

[AK] D.V. Anosov, A. Katok, New examples in smooth ergodic theory. *Ergodic diffeomorphisms, Trans. Moscow Math. Soc.* Vol. 23 (1970), p.1.

[AA] V.I. Arnold, A. Avez, *Ergodic Problems of Classical Mechanics*, W.A. Benjamin, New York, Amsterdam, 1968.

[A1] D. Aronson, M.A. Chory, G.R. Hall, R.P. McGehee, Bifurcations from an invariant circle for two parameter families of maps of the plane: A computer assisted study, *Com. Math. Phys.* **83** (1982), p.303.

[Au] S. Aubry, P.Y. LaDaeron, The discrete Frenkel-Kontorova model and its extensions, *Physica 8D* (1983), p.381.

[Ba1] V. Bangert, Mather sets for twist maps and geodesics on tori, to appear in *Dynamics Reported.*

[Bk] D. Bernstein, A. Katok, Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian Systems with convex Hamiltonians, *Erg. Th. and Dyn. Sys.* **88** (1987), p.225.

[B1] G.D. Birkhoff, Proof of Poincaré's geometric theorem, *Trans. Amer. Math. Soc.* **14** (1913), p.14.

[Bd1] P. Boyland, Rotation sets and Morse decompositions in twist maps, to appear in *Ergod. Thr. and Dyn. Sys.*

[Bd2] P. Boyland, Lecture, Boulder, Colo. 1987 (see this volume)

[BH] P. Boyland, G.R. Hall, Invariant circles and the order structure of periodic orbits for monotone twist maps, *Topology* **26** (1987), p.21.

[Bo] F. Botelho, Rotation sets of maps of the annulus, preprint.

[Br] M. Brown, W.D. Newman, Proof of the Poincaré-Birkhoff fixed point theorem, *Michigan Math. J* **24** (1977), p.21.

[Bu] S. Bullett, Invariant circles for the piecewise linear standard map, *Com. Math. Phys.* **107** (1986), p.241.

[C] C. Conley, *Isolated Invariant Sets and the Morse Index*, CBMS Regional Conference Series in Math, 38, AMS, Providence, 1978.

[C-Z] C. Conley, E. Zehnder, The Birkhoff-Lewis fixed point theorem and a conjecture of V.I. Arnold, *Invent. Math.* **73** (1983), p.33.

[DH] A. Douady, J. Hubbard, Dynamique des Polynomes Complexes, preprint, Universite Paris-Sud (1984).

[F1] J. Franks, Recurrence and fixed points of surface homeomorphisms, to appear, *Erg. Th. and Dyn. Systems.*

[F2] J. Franks, Generalization of the Poincaré-Birkhoff Theorem, preprint.

[FH] A. Fathi, M. Herman, Existence de diffeomorphismes minimaux, *Asterisque*, **49** (1977), p.37.

[G] J.M. Greene, A method for determining a stochastic transition, *J. Math. Phys.* **20** (1979), p.1183.

[G1] C. Gutierrez, On the C^n-closing lemma for flows on the torus T^2, *Erg. Th. and Dyn. Sys.*. **6** (1986), p.45.

[Hl] G.R. Hall, Bifurcation of an attracting invariant circle: A Denjoy attractor, *Erg. Th. and Dyn. Sys.*, **3** (1983), p.87.

[Hl1] G.R. Hall, A topological version of a theorem of Mather on twist maps, *Ergod. Th. and Dyn. Sys.*, **4** (1984), p.585.

[Hl2] G.R. Hall, A topological version of a theorem of Mather's on shadowing in monotone twist maps, to appear, proceedings of semester in Dynamics, Banach Institute.

[H1] M. Handel, A pathological area preserving C^∞ diffeomorphism of the plane, *Proc. A.M.S.*, **86** (1982), p.163.

[H2] M. Handel, Global shadowing of pseudo-Anosov homeomorphisms *Erg. Th. and Dyn. Sys.* **5** (1985), p.373.

[H3] M. Handel, Zero entropy surface diffeomorphisms, preprint.

[Ha1] J. Harrison, Continued fractals and the Seifert conjecture, *Bul. A.M.S.* **13** (1985), p.147.

[He1] M. Herman, Sur les Courbes invariant par les diffeomorphisms de l'anneau, Vol.1 *Asterisque* **103-104** (1986).

[HH] K. Hockett, P. Holmes, Josephson's junction, annulus maps, Birkhoff attractors, horseshoes and rotation sets, *Erg. Th. and Dyn. Sys.* **6** (1986), p.205.

[K1] A. Katok, Some remarks on Birkhoff and Mather twist map theorems, *Erg. Th. and Dyn. Sys.* **2** (1982), p.185.

[J1] I. Jungries, Some results on periodic points of twist maps, preprint.

[J2] I. Jungries, A method for proving that monotone twist maps have no invariant circles, preprint.

[LC1] Le Calvez, Proprietes des attracteurs de Birkhoff, to appear in *Erg. Th. and Dyn. Sys.*.

[LC2] Le Calvez, Proprietes dynamiques des regions d'instabilite, preprint.

[MP] R.S. MacKay, I.C. Percival, Converse KAM: Theory and practice, *Com. Math. Phys.* **98** (1985), p.469.

[Ma1] J. Mather, Existence of quasi-periodic orbits for twist homeomorphisms of the annulus, *Topology*, **21** (1982), p.457.

[Ma2] J.N. Mather, Amount of rotation about a point and the Morse index, *Com. Math. Phys.* **94** (1984), p.141.

[Ma3] J.N. Mather, Non-existence of invariant circles, *Erg. Th. and Dyn. Sys.* **2** (1982), p.397.

[Ma4] J.N. Mather, A criterion for the non-existence of invariant circles, *Pub. Math. I.H.E.S.* **63** (1986), p.153.

[Ma5] J.N. Mather, Destruction of invariant circles, preprint.

[M1] J. Moser, Recent developments in the theory of Hamiltonian Systems, *SIAM Review* **28** (1986), p.459.

[M2] J. Moser, Minimal solutions of variational problems on a torus, *Annales de L'Institute Henri Poincare, Analyse Nonlineaire* **3** (1986), p.229.

[M3] J. Moser, *Stable and Random Motions in Dynamical Systems*, Annals of Mathematical Studies #77, Princeton Univ. Press, Princeton, 1973.

[OS] A. Olvera, C. Simo, An obstruction method for the destruction of invariant curves, to appear in Physica D.

[P] C. Pugh, An improved closing lemma and general density theorem, *Amer. J. Math.* **89** (1967), p.1010.

[W1] M. Wojtkowski, Invariant families of cones and Lyapunov exponents, *Erg. Th. and Dyn. Sys.* **4** (1984).

[W2] M. Wojtkowski, Principles for the design of billiards with non vanishing Lyapunov exponents, preprint.

[Y1] L-S Young, A closing lemma on the interval, *Invent. Math.* **54** (1979), p.179.

http://dx.doi.org/10.1090/conm/081/986263

Contemporary Mathematics
Volume **81**, 1988

AREA PRESERVING HOMEOMORPHISMS OF TWO MANIFOLDS

Edward E. Slaminka

ABSTRACT. Using the methods of free modification due to M. Brown and its extension to area preserving, orientation preserving homeomorphisms by Pelikan-Slaminka, we present the tools needed to prove a collection of theorems which are the topological analogs of existence and removal theorems for fixed points in the differential category.

The theorems cited concern the bound on the index of an isolated fixed point for area preserving, orientation preserving homeomorphisms of two manifolds; the removal of index zero isolated fixed points for area preserving, orientation preserving homeomorphisms (and C^k-diffeomorphisms) on two manifolds by an area preserving orientation preserving homeomorphism (C^k-diffeomorphism which is a local perturbation; the existence of n+1 stable and unstable compact, connected, simply connected, zero area sets for index -n isolated fixed points for area preserving, orientation preserving homeomorphisms of two manifolds; and, the Conley-Zehnder theorem for area preserving, orientation preserving homeomorphisms of the two torus.

1. INTRODUCTION. In this paper we present a technique which has proven useful in understanding the dynamics of area preserving homeomorphisms of two manifolds. This method, called free modification, is an extension of the method of the same name developed by M. Brown and used by him to prove the Brouwer Translation Arc Lemma [7]. Free modifications give us the tool needed to understand the local behavior of a homeomorphism near an isolated fixed point of index not equal to one. We use this technique to prove a few fixed point theorems which were first stated assuming differentiability. We employ free modifications to compensate for the fact that we have neither Jacobians to measure area nor linearizations to compute the fixed point index. It must be noted that our technique only applies (at present) to orientation preserving homeomorphisms of two manifolds. This is due mainly to the fact that an essential ingredient in the development of our technique is the Brouwer Translation Arc Lemma for which no known applicable higher dimensional analog has been found.

1980 Mathematics Subject Classification (1985 Revision). 58F.

In Section 2 we will describe the theorems which we have been able to prove using this technique. In Section 3 we present the definitions and lemmas required in the sequel. The main body of this paper is encapsulated in Section 4 wherein we describe the technique of free modification in detail. In Section 5 we present the technical arguments needed to modify the homeomorphism into a setting for which Section 4 is applicable.

The following results and constructions apply equally well to homeomorphisms which preserve a Lebesgue-Stieltjes measure, that is, a measure which is bi-absolutely continuous with respect to Lebesgue measure.

I wish to thank Ken Meyer and Don Saari for organizing this conference and providing me with the venue to present my results. I also wish to thank Mort Brown who introduced me to these methods and for the support he has shown.

2. A SURVEY OF RESULTS. The first theorem gives an upper bound to the index for isolated fixed points on two manifolds.

THEOREM (Pelikan-Slaminka [20]). Let $h:M^2 \to M^2$ be an area preserving, orientation preserving homeomorphism of an orientable two manifold, and let p be an isolated fixed point for h. Then the index of p with respect to h is less than or equal to 1.

In 1975 Simon [22] proved the above theorem with the proviso that h is smooth.

Using this result and the technique of free modifications Boucher and Brown have shown the following.

THEOREM (Boucher, Brown [4]). Let $h:D^2 \to D^2$ be an area preserving, orientation preserving homeomorphism of a two disc having n stable, n unstable fixed points on $bd(D^2)$. The h possesses at least $n + 1$ fixed points in the $int(D^2)$.

A fixed point $p \in S^1$ under an orientation preserving homeomorphism $h:S^1 \to S^1$ is <u>stable</u> (resp. <u>unstable</u>) if there exists a neighborhood N of p such that if $x \in N$ then $h^n(x) \to p$ (resp. $h^{-n}(x) \to p$) as $n \to \infty$.

This theorem extends one of Montgomery which states that if $h:int(D) \to int(D)$ is an area preserving, orientation preserving homeomorphism of an open two disc to itself, then h possesses at least one fixed point in the $int(D)$.

Though the local fixed point index of Brouwer gives information about the existence of fixed points when the index is non zero, very little is known about index zero fixed points. Schmitt [21] proved that if $h:R^2 \to R^2$ is an orientation preserving homeomorphism of the plane and p is an isolated fixed point of index zero, then p can be removed by a local perturbation. Simon-Titus [24] proved a similar theorem when h is C^k, by constructing a C^1 local perturbation. We have shown that our technique gives the above two results, extends Simon's to a C^k local perturbation, and also proves an analogous theorem for area preserving homeomorphisms and C^k-diffeomorphisms.

THEOREM (Slaminka [25]). Let $h:M^2 \to M^2$ be an area preserving, orientation preserving homeomorphism (C^k or C^∞ -diffeomorphism) of an orientable two manifold M^2 having an isolated fixed point p of index zero. Given any open neighborhood N of p such that $N \cap \text{Fix}(h) = p$, there exists an area preserving homeomorphism (C^k or C^∞ -diffeomorphism) $\hat{h}$ such that:

1) $\hat{h} = h$ on $\overline{M - N}$; and,
2) $\hat{h}$ is fixed point free on N.

As corollaries we prove the existence of the second fixed point for the Birkhoff Twist Theorem (cf. Birkhoff [2, 3] and Brown-Neumann [9] for earlier proofs) which relies only upon the fact that if the first fixed point had index zero, then it could be removed; and, we prove the existence of the second fixed point for the Conley-Zehnder theorem for area preserving homeomorphisms (Franks [17] proved the existence of the first fixed point).

A standard result in dynamics is the stable/unstable manifold theorem. However, it requires that the map be smooth and that the fixed point be hyperbolic. We present a version for area preserving homeomorphisms of two manifolds which requires only that the index of the fixed point be not equal to one. Though our version does not generate manifolds, the continua which we do construct will be useful in the next theorem.

THEOREM (Baldwin, Slaminka [1]). Let $h:M^2 \to M^2$ be an area preserving, orientation preserving homeomorphism of an orientable two manifold having an isolated fixed point p with index equal to $-n$ for $n \geq 0$. Let D be an open disc with $p \in D$, such that $D \cap \text{Fix}(h) = p$. There exists $2(n+1)$ compact, connected, simply connected, area zero sets $U_1, U_2, \ldots, U_{n+1}$, and $S_1, S_2, \ldots, S_{n+1}$ such that:

1) $U_i, S_i \subset D$, and U_i, S_i meet $\text{bd}(D)$ for all i;
2) $p \in U_i, S_i$ for all i;
3) $h(S_i) \subset S_i$, $h^{-1}(U_i) \subset U_i$ for all i;
4) $x \in S_i$ implies that $h^m(x) \to p$ as $m \to \infty$
$x \in U_i$ implies that $h^{-m}(x) \to p$ as $m \to \infty$ for all i; and,
5) $S_i \cap S_j = p = U_i \cap U_j$ for all $i \neq j$.

In 1945 Montgomery [18] proved that if $h:R^n \to R^n$ is a measure preserving homeomorphism and A is a compact connected set such that $h(A) \subset A$, then if U is an open set with compact closure which includes A, there exists a compact connected set K of which A is a proper subset and such that K is in $h^{-1}(U)$ and $h(K) \subset K$.

In 1985 Ding [12] proved the existence of a Lagrange stable set for each fixed point of an area preserving homeomorphism of the plane.

Let $h:T^2 \to T^2$ be an orientation preserving homeomorphism of a two torus and let $\tilde{h}:R^2 \to R^2$ be a lift of h to its universal cover. Also, let m be Lebesgue measure in R^2. The mean translation vector of h is defined to be $\int_D \tilde{h}(\vec{z}) - \vec{z}\, dm$ where D is a fundamental domain for h. Following an argument of Franks [17] it can be shown that free modifications preserve mean translation vectors.

Using the above we prove the following analog of the Conley-Zehnder Theorem.

THEOREM (Slaminka [26]). Let $h:T^2 \to T^2$ be an area preserving orientation preserving homeomorphism of the two torus with mean translation vector $\vec{0}$. Then h has at least three fixed points.

Conley-Zehnder [10] proved that every symplectic C^1-diffeomorphism h on the 2n-torus which is generated by a globally Hamiltonian vectorfield, possesses at least $2n + 1$ fixed points. John Franks [17] initiated the research into the topological setting by proving that one fixed point existed for the two torus.

Our hope is that these same techniques can be used to generalize this result to surfaces (the two-dimensional Arnol'd conjecture). The differentiable versions have been proven by Flöer [16], Sikorav [23], and Eliashberg [14].

3. DEFINITIONS AND LEMMAS. We present here the following definitions and lemmas (without proof) which will be needed in the sequel. These include the concepts of local fixed point index, the Brouwer Translation Arc Lemma, free modification of a homeomorphism and the construction of area preserving homeomorphisms.

If $h:R^2 \to R^2$ is an orientation preserving homeomorphism, C is a simple closed curve with $C \cap \mathrm{Fix}(h) = \emptyset$, then the index of C with respect to h, $\mathrm{ind}(h,C)$, is defined to be $\deg_{x\in C} \dfrac{h(x) - x}{\|h(x) - x\|}$.

One of the main properties of index is an isotopy condition which states that if C_1 and C_2 are two simple closed curves bounding an annulus A such that $\mathrm{Fix}(h) \cap A = \emptyset$, then $\mathrm{ind}(h,C_1) = \mathrm{ind}(h,C_2)$ (cf. Dold [13]). Thus an isolated fixed point p inherits a well-defined index which we call the fixed point index of p. Brown [6] and Schmitt [21] observed that if $C \cap h(C)$ is finite, then the $\mathrm{ind}(h,C)$ can be computed by considering the orientation of the vector from x to $h(x)$ for $h(x) \in C \cap h(C)$, that is, how the vector rotates as $h(x)$ moves through an intersection point of C and $h(C)$.

As mentioned earlier, our work depends heavily upon the following Lemma.
LEMMA (Brouwer [5]). Let $h:R^2 \rightarrow R^2$ be an orientation preserving homeomorphism, and let D be a disc such that $h(D) \cap D = \emptyset$. If $h^n(D) \cap D \neq \emptyset$ for some $n \neq 0$, then there exists a simple closed curve C with $\mathrm{ind}(h,C) = 1$.

More recent proofs of this lemma can be found in Brown [7] and Fathi [15]. This recurrence type lemma has found applications in the recent work of Franks [17].

The technique of free modification is the cornerstone upon which our theorems are built. Let $h:M^2 \rightarrow M^2$ be an orientation preserving homeomorphism of an orientable two manifold, and let D be a disc such that $h(D) \cap D = \emptyset$. Let $g:M^2 \rightarrow M^2$ be a homeomorphism supported on D. Then $f = hg$ is a free modification of h.

Note that the following properties hold for free modifications:

1) $\mathrm{Fix}(f) = \mathrm{Fix}(h)$;
2) $\mathrm{Ind}(f,C) = \mathrm{Ind}(h,C)$ for simple closed curves C with $\mathrm{Fix}(h) \cap C = \emptyset$; and,
3) f is isotopic to h by and isotopy j_t where $h^{-1}j_t$ is supported on D (by Alexander's isotopy lemma).

Our use of free modifications alters the orbit structure of h only on "small" sets which do not contain the fixed point. In the sequel we will use the Brouwer Lemma which requires passing to the universal cover of $(M^2 - \mathrm{Fix}(h)) \cup p$ where p is an isolated fixed point. By Montgomery's theorem we see that this cover must be R^2 (any aréa preserving, orientation preserving homeomorphism of a two-sphere must have at least two fixed points).

Our use of free modifications will require that either the modification preserves area or a measure which is equivalent to area (i.e. Lebesgue-Stieltjes measure). The following lemma can be proved either using a result of Oxtoby-Ulam [19] or can be found in Slaminka [25].
LEMMA. Let $h:\mathrm{bd}(D) \rightarrow \mathrm{bd}(E)$ be an orientation preserving homeomorphism (C^2-diffeomorphism) where D, E are discs (C^k discs) in R^2 such that $\mathrm{area}(D) = \mathrm{area}(E)$. There exists an area preserving homeomorphism (C^k-diffeomorphism) $\hat{h}:D \rightarrow E$ such that $\hat{h} = h$ on $\mathrm{bd}(D)$.
This next lemma involves constructing an area preserving homeomorphism which moves points along an arc in the interior of a disc.
LEMMA. Let D be a disc (C^k disc) in R^2, let $I \subset \mathrm{int}(D)$ be an arc (C^k arc) with endpoints x, y and let $J \subset I$ be an arc with endpoint x. There exists an orientation preserving, area preserving homeomorphism (C^k-diffeomorphism) $f:D \rightarrow D$ such that:

1) $f = id$ on $bd(D)$

2) $f(I) = J$.

To prove this lemma one merely cuts the disc into two discs so that the arc I is on the boundary of both discs. Now apply the previous lemma and paste the discs together again.

Free modifications do not necessarily preserve area. However there is a sufficiently large class of free modifications which will preserve a Lebesgue-Stieltjes measure.

PROPOSITION. Let h be an orientation preserving homeomorphism of R^2 preserving a Lebesgue-Stieltjes measure u and let $Fix(h)$ be isolated with index of $p = n \neq 1$ for each $p \in Fix(h)$. Suppose that hg is a free modification of h (where g is a supported on the disc D) and that g is either C^2 or bi-Lipschitz. Then there exists a Lebesgue-Stieltjes measure v which is preserved by hg. Construction of v: For a measureable set $A \subset R^2$ define $A_j = h^{-j}(D) \cap A$ for $j = 0, 1, 2, \ldots$ and set $A_c = A - \bigcup A_j$. The $A = A_c \cup \bigcup A_j$. By the Brouwer Lemma we have expressed A as a disjoint union. Define the measure v as follows:

$$v(A) = u(A_c) + \sum u((hg)^{j+1}(A_j)) \quad \text{with} \quad j = 0,1,2,\ldots$$

One can then show that v is thus a lebesgue-Stieltjes measure. The condition that g be C^2 or bi-Lipschitz is sufficient to ensure that the measure v is non-atomic.

4. REDUCTION TO CANONICAL FORM. Let $h:R^2 \rightarrow R^2$ be an orientation preserving homeomorphism having isolated fixed points p_i such that the index of $p_i = n$ for all i, with $n \neq 1$. Note that this implies that given any simple closed curve C with $Fix(h) \cap C \neq 0$, then $ind(h,C) \neq 1$.

Let D be a disc with $p \in int(D)$, $Fix(h) \cap D = p$, and let $C = bd(D)$. We will assume the following simplifications in this section. The more general case will be considered in section 5.

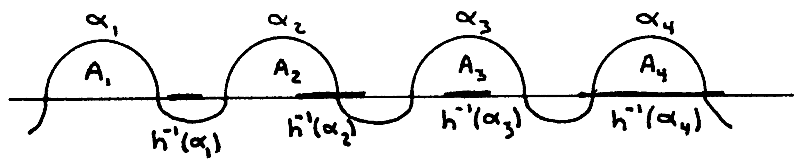
1) C intersects $h(C)$ transversely;

2) $C \cap h(C)$ is finite; and,

3) $D \cap h(D)$ is connected.

By a result of Curtis-Dugundji [11], if h is C^2 then 3) is true for "small" discs.

Let $\alpha_1, \ldots, \alpha_m$ be the connected arcs of $\overline{h(C)-D}$, let $\beta_1, \ldots, \beta_m$ be the arcs on C which separate each α_i from p in $h(D)$, and let A_i be the discs bounded by $\alpha_i \cup \beta_i$ (See figure below).



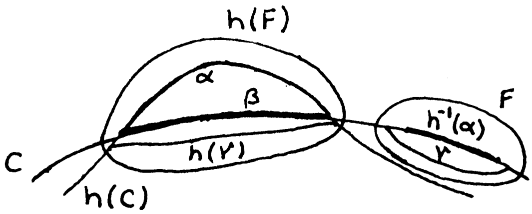
The arcs will be partitioned into the following four types:

Type 1: $h^{-1}(\alpha_i) \cap A_i = \emptyset$;
Type 2: $h^{-1}(\alpha_i)$ meets precisely one endpoint of β_i;
Type 3: $h^{-1}(\alpha_i) \subset \beta_i$ (elliptic); and
Type 4: $h^{-1}(\alpha_i) \supset \beta_i$ (hyperbolic).

We say that a simple closed curve is in canonical form for a homeomorphism h if and only if all of the arcs α_i are hyperbolic, elliptic or $C \cap h(C) = \emptyset$. Given a simple closed curve our goal is to modify the curve and/or homeomorphism to obtain one which is in canonical form. Once the curve is in this form it will be relatively easy to prove the theorems quoted in section 2. Given a curve in canonical form it is quite simple to read off the index of that curve. The $\operatorname{ind}(h,C) = 1 + E - H$ where E is the number of elliptic arcs and H is the number of hyperbolic arcs. This is the topological analog of the Poincaré-Hopf theorem for flows. Arcs of type 1 and 2 will be removed showing that these types of intersections are inessential in the computation of the index. Elliptic/hyberbolic pairs of arcs will be cancelled, resulting in our canonical form

We now proceed to eliminate the Type 1 arcs. There are three different methods for eliminating Type 1 arcs. We present all three methods due to their applicability to a variety of settings.

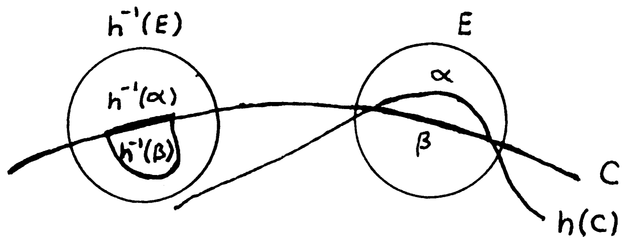
Method 1. This method yields only a C^0 modification and can be used where area preserving is not required. We will repress the subscripts on α_i, β_i, and A_i as we will be concentrating on one arc. There exists a disc F containing $h^{-1}(\alpha)$ such that $h(F)$ contains A (since $\operatorname{Fix}(h) \cap C = \emptyset$) and such that $F \cap h(F) = \emptyset$. Let γ be an arc in F with endpoints on $C - h^{-1}(\alpha)$ such that $h(\gamma) \subset D$ (see figure below).



Now we can describe the free modification which will remove the arc α. Let $g:R^2 \to R^2$ be a homeomorphism which is the identity off of F and maps $h^{-1}(\alpha)$ to γ. The composition hg will map $h^{-1}(\alpha)$ to $h(\gamma)$, thus removing the intersection.

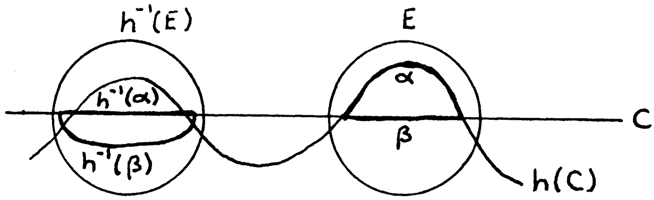
Method 2. This method is identical to method 1 except that we pick g to be C^2 or bi-Lipschitz in order that hg preserves a Lebesgue-Stieltjes measure. We no longer have an area preserving homeomorphism but we do have a measure preserving homeomorphism which is adequate to prove the theorem by Pelikan-Slaminka stated in Section 2.

Method 3. We will remove the intersection using an area preserving modification and will also construct a different simple closed curve. We consider the inverse of α and β. We shall first assume that $h^{-1}(\alpha) \cap h(C) = \emptyset$ and that $h^{-1}(\beta) \cap h(C) = \emptyset$. We construct a C^k disc E containing $\alpha \cup \beta$ such that $E \cap h^{-1}(E) = \emptyset$ and $h^{-1}(E) \cap h(C) = \emptyset$. (See figure below.)

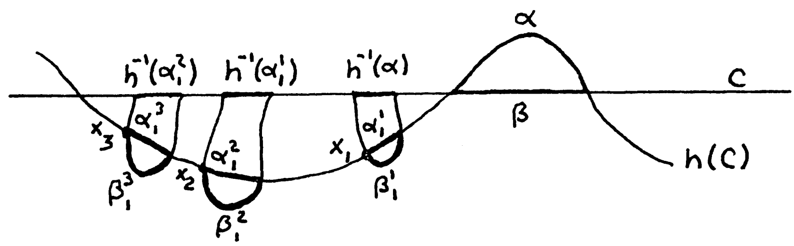


Let $C' = \mathrm{bd}(\overline{D - h^{-1}}(E))$. The simple closed curve C' bounds a disc D' such that $p \in \mathrm{int}(D')$ and card $|C' \cap h(C')|$ is two less than card $|C \cap h(C)|$. In effect, we have "removed" the arc α.

Assume that $h^{-1}(\alpha) \cap h(C) \neq \emptyset$. If card $|h^{-1}(\alpha) \cap h(C)| \geq$ card $|h^{-1}(\beta) \cap h(C)|$ then construct a C^k-disc E containing $\alpha \cup \beta$ such that card $|E \cap h^{-1}(\beta)|$ = card $|h(C) \cap h^{-1}(\beta)|$ and "cut out" E as before (see figure below).

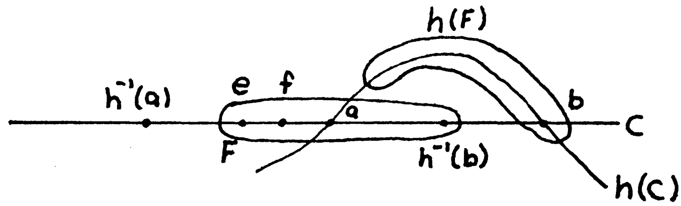


Finally, we assume that card $|h^{-1}(\beta) \cap h(C)| > \text{card}|h^{-1}(\alpha) \cap h(C)| \geq 0$. We may also assume that card$|h^{-1}(\beta) \cap h(C)|$ is finite by the results of Section 5. Let F^0 be the disc bounded by $\alpha \cup \beta$, and F_i^1 be the discs bounded by $h^{-1}(\beta)$ and $h(C)$ with $p \notin F_i^1$. We construct arcs α_i^1, β_i^1 such that $\alpha_i^1 \cup \beta_i^1 = \text{bd}(F_i^1)$ where $\beta_i^1 \subset h^{-1}(\beta)$ and $\alpha_i^1 \subset h(C)$. (See figure below).



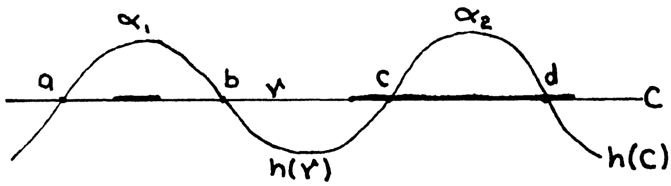
For each simple closed curve $\alpha_i^1 \cup \beta_i^1$ we consider $h^{-1}(\alpha_i^1) \subset C$ and $h^{-1}(\beta_i^1) \subset D$, and if possible, apply the above analysis to cut out the disc E_i^1 containing $\alpha_i^1 \cup \beta_i^1$ and then cut out the disc E. The only obstruction to this procedure would be if, for each $n > 0$, $h^{-1}(\beta_1^n) \cap \text{bd}(C) \neq \emptyset$. Note that $h^n(\beta_i^n) \subset \beta$. Choose an orientation for C and let x_n be the right hand endpoint of β_i^n. The sequence $\{x_n\} \subset h(C)$ and hence must have a cluster point $x \in h(C)$. However x cannot be in the fixed point set for h, thus by the Brouwer Lemma (or use the proof of Lemma 3.4 of Brown [8]) we arrive at a contradiction. Hence the construction of the arcs α_i^n, β_i^n must terminate for finite n. We then proceed to "cut out" the finite number of discs E_i^n to obtain a new simple closed curve C' in which α has been removed.

We can now eliminate Type 2 arcs. For simplicity assume the orientation as shown in the figure below where a, b are the endpoints of α.



Let e be a point in $h^{-1}(\alpha)$ such that e does not meet β and $e \neq h^{-1}(a)$. Then there exists a disc F containing the arc $eh^{-1}(b)$ such that $F \cap h(F) = \emptyset$. Let f be a point in the arc $eh^{-1}(b)$ which does not meet β and is distinct from e. Construct an area preserving homeomorphism g (via a Lemma in Section 3) which is the identity outside of F and which maps the arc ef to the arc $eh^{-1}(b)$ along C. The free modification hg then maps $h^{-1}(a)f$ to the arc $h^{-1}(\alpha)$. The arc α is now a Type 1 arc which can be removed using one of the methods noted above.

We will finally remove pairs of elliptic and hyperbolic arcs. Again, for simplicity, assume the orientation given in the figure below where the arc α_1 is elliptic and the arc α_2 is hyperbolic. Let the endpoints of α_1 be a and b, and let the endpoints of α_2 be c and d. Also let γ be the arc on C between $h^{-1}(\alpha_1)$ and $h^{-1}(\alpha_2)$ which intersects b.

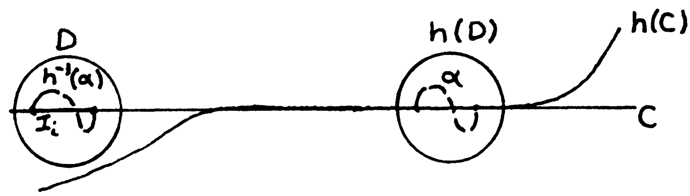


Viewing h as a homeomorphism of S^2 with ∞ as a fixed point (note that h is not now area preserving) one can see that $h(\gamma)$ is an arc of Type 2. By removing this arc the elliptic and hyperbolic arcs coalesce into another arc of Type 2 (with respect to the original homeomorphism with p as a fixed point), and can then be removed.

After performing the above reductions a finite number of times we obtain a simple closed curve C and homeomorphism h which is in canonical form. Depending upon which method we use, we either have a C^0-homeomorphism, a homeomorphism which preserves a Lebesgue-Stieltjes measure or an area preserving homeomorphism.

5. GENERAL CASE. In the general case the intersection of C and $h(C)$ is not as simple as presented in section 4. In this section we show that this more general setting can be reduced to that of the previous section. As before we assume that D is a disc containing a fixed point p in its interior and that C is the boundary of D.

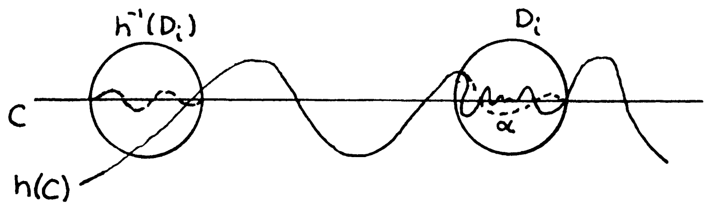
NON TRANSVERSE INTERSECTIONS. Assume that $h(C) \cap C$ consists of non transverse intersections. We consider a Lebesgue number δ for h restricted to C such that if I is any connected arc on C with diameter less δ, then $h(I) \cap I = \emptyset$. Pick a finite collection of such arcs which cover C. If $h(I_i)$ forms a non-transverse intersection with C construct a disc D containing I_i in its interior such that $h(D) \cap D = \emptyset$. Let α be an arc in $h(D)$ with endpoints the same as $h(I_i)$ and such that the area of the discs in $h(D)$ separated by $h(C)$ equal the area of the corresponding discs in $h(D)$ separated by $h(C-I_i) \cup \alpha$ (See figure below).



Now construct an area preserving homeomorphism $g: D \to D$ which maps I_i to $h^{-1}(\alpha)$. Extend g to R^2 so that $g = id$ off of D. Perform this construction for each of the appropriate I_i and obtain a modification which has transverse interesections.

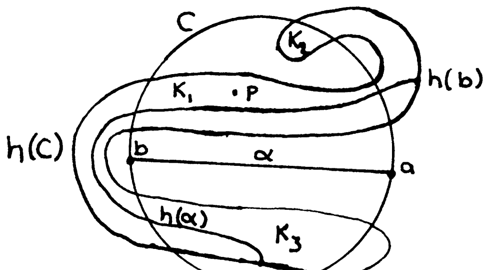
INFINITE NUMBER OF INTERSECTIONS. Assume that $h(C) \cap C$ consists of an infinite number of transverse intersections. Cover $h(C)$ with a finite number of discs $D_1, D_2, \ldots D_k$ having the properties that:

1) $h^{-1}(D_i) \cap D_i = \emptyset$, for each i; and,
2) $h(C) \cap D_i$ is connected for each i. (See figure below).



Connect the endpoints of $h(C) \cap D_i$ by a smooth arc α which lies in $\mathrm{int}(D_i)$, intersects $C \cap D_i$ a finite number of times, and such that the area of each of the two discs bounded by α in D_i are the same as those bounded by C in $h^{-1}(D_i)$. Define $g_i : h^{-1}(D_i) \to h^{-1}(D_i)$ such that g_i is an area preserving homeomorphism which is the identity on the boundary of $h^{-1}(D_i)$ and takes $C \cap h^{-1}(D_i)$ to $h^{-1}(\alpha)$ (use the Lemma in Section 2 or constructing such a map). Now extend g_i to all of R^2 so that $g_i = \mathrm{id}$ on the complement of $h^{-1}(D_i)$. By considering $h \circ g_1 \circ \ldots \circ g_k$ we obtain an orientation preserving, area preserving homeomorphism under which C intersects its image a finite number of times.

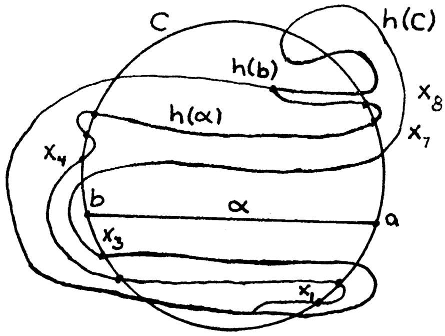
CONNECTED COMPONENT. By the above we can assume that $D \cap h(D)$ is the union of a finite number of components K_i, $i=1,2,\ldots n$, with the fixed point $p \in K_1$. We will construct a simple closed curve C_1 bounding a disc $D_1 \subset D$ with $p \in \mathrm{int}(D_1)$ and a homeomorphism h' such that $h'(D_1) \cap D_1$ has fewer components (See figure below).



Pick $a,b \subset C \cap (\bigcup K_i)^c$ such that a,b, are endpoints of an arc α in $D \cap (\bigcup K_i)^c$ and such that $\alpha \cap \mathrm{bd}(D \cap h(D)^c) = \{a,b\}$, with the property that α separates K_1 from at least one K_i $(i \neq 1)$ in D. Since $h(\alpha) \cap \alpha = \emptyset$, we may assume that $h(\alpha)$ intersects C a finite number of times (otherwise we can modify the map h on a disc containing to achieve this).

Let β_1 and $\beta_2 \subset C$ be the two arcs with endpoints a and b. If $h(\alpha)$ K_i is connected for each i, let D_1 be the disc bounded by α and either β_1 or β_2 (whichever is such that $K_1 \subset D_1$). Then $D_1 \cap h(D_1)$ has less than n components.

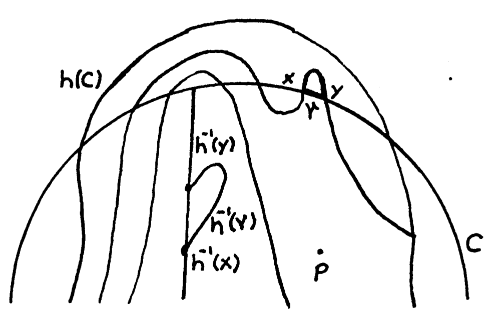
Thus we assume that $h(\alpha) \cap K_i$ is not connected for at least some i (See figure below).



Let $x_1, x_2, \ldots, x_m$ be the intersection points of $h(\alpha)$ with C. Assume that the subscripts give an order to these points which is inherited from α. Let $x_i x_{i+1} \subset h(\alpha)$ be the arc with endpoints x_i and x_{i+1}.

Consider only those arcs which lie in the complement of the interior of D. Since $h(D)$ is contractible there exists at least one such arc $x_j x_{j+1}$ with endpoints lieing in $bd(K_i)$ for some i. Pick x_j so that no other x_k lies between x_j and x_{j+1} on $C \cap bd(K_i)$, where $x_k x_{k+1}$ is another such arc. We will modify the map h in such a way as to move the arc $x_j x_{j+1}$ into the interior of D. Since we will focus upon this particular arc, we will rename the arc xy, and the component K.

Let γ be the arc on $C \cap bd(D)$ with endpoints x and y. We observe that $(\gamma \cup xy)\ h^{-1}(\gamma \cup xy) = \emptyset$. Thus there exists a disc E containing $h^{-1}(\gamma \cup xy)$ in its interior such that $E \cap h(E) = \emptyset$. We now cut out the arc xy as we did in Section 2, Method 3 (or if only a measure preserving map is required we modify the homeomorphism using Method 2). This procedure reduces the number of components, and, by finite induction, we can reduce our example to a single component.



BIBLIOGRAPHY

1. Baldwin S. and Slaminka E. E., "A topological analog to the stable/unstable manifold theorem for area preserving homeomorphisms of two manifolds", in preparation.

2. Birkhoff, G.D., "Proof of Poincaré's geometric theorem", Trans. Amer. Math. Soc., 14 (1913), 14-22.

3. Birkhoff, G. D., "An extension of Poincaré's last geometric theorem", Acta Math., 47 (1925), 297-311.

4. Boucher, K. and Brown, M., private communication.

5. Brouwer, L. E. J., "Beweis des ebenen Translationssatzes", Math. Ann., 72 (1912), 27-54.

6. Brown, M., private communication.

7. Brown, M., "A new proof of Brouwer's lemma on translation arcs", Houston J. of Math., 10 (1984), 35-41.

8. Brown, M., "Homeomorphisms of two-dimensional manifolds", Houston J. of Math., 11 (1985), 455-469.

9. Brown, M. and Neumann, W. D., "Proof of the Poincaré-Birkhoff fixed point theorem", Mich. Math. J., 24 (1977), 21-31.

10. Conley, C. and Zehnder, E., "The Birkhoff-Lewis fixed point theorem and a conjecture of Arnold", Invent. Math. 74 (1983), 33-49.

11. Curtis, M. L. and Dugundji, J., "A proof of deRham's theorem", Fundamenta Math., 68 (1970), 265-269.

12. Ding, T., "Boundedness of solutions of Duffing's equation", J. Diff. Eq., 61 (1986), 178-207.

13. Dold, A., Lectures on Algebraic Topology, Springer Verlag, Berlin, 1972.

14. Eliashberg, preprint.

15. Fathi, A., "An orbit closing proof of Brouwer's lemma on translation arcs", preprint.

16. Flöer, A., "Proof of the Arnold conjecture for surfaces and generalizations for certain Kähler manifolds", Duke Math. J., 53 (1986), 1-32.

17. Franks, J., "Recurrence and fixed points of surface homeomorphisms, preprint.

18. Montgomery, D., "Measure perserving homeomorphisms at fixed points", Bull. Amer. Math. Soc., (1945), 949-953.

19. Oxtoby, J. C. and Ulam, S. M., "Measure-preserving homeomorphisms and metrical transitivity", Annals of Math., 42 (1941), 874-920.

20. Pelikan, S. and Slaminka, E. E., "A bound for the fixed point index of area preserving homeomorphisms of two manifolds", to appear in Ergodic Th. and Dyn. Systems.

21. Schmitt, B., "Sur les plongements, admettant zero ou un point fixe, du disque dans le plan", Topology, 14 (1975), 357-365.

22. Simon C., "A bound for the fixed-point index of an area-preserving map with applications to mechanics", Invent. Math., 26 (1974), 187-200.

23. Sikorav, J. C., "Points fixes d'un symplectomorphisme homologue de l'identite", J. Differential Geom., 22 (1985), 49-79.

24. Simon, C. and Titus, C., "Removing index-zero singularities with C^1-small perturbations", Dynamical Systems - Warwick 1974 (Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick, Coventry 1973/74), Lecture Notes in Math., Vol. 468, Springer, Berlin, 1975, 278-286.

25. Slaminka, E. E., "Removing index 0 fixed points for area preserving maps of two manifolds", submitted.

26. Slaminka, E. E., "A Conley-Zehnder theorem for area preserving homeomorphisms of the two torus", in preparation.

MATHEMATICS - FOUNDATIONS, ANALYSIS, AND TOPOLOGY
AUBURN UNIVERSITY
AUBURN, ALABAMA 36849

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986264

Contemporary Mathematics
Volume 81, 1988

AN ANOSOV TYPE STABILITY THEOREM FOR ALMOST PERIODIC SYSTEMS

Kenneth R. Meyer[1]

ABSTRACT. In this paper I discuss a natural generalization of the structural stability theorem for Anosov diffeomorphisms i.e. diffeomorphisms which have a global hyperbolic structure. The maps discussed define skew product dynamical systems over a discrete almost periodic system. This is the natural generalization for almost periodic systems of the Poincaré map for periodic systems. This follows from the Miller-Sell method of embedding an almost periodic system of differential equations in a flow. Generalizations are given of the shadowing lemma, the expansive property, and the openness and the structural stability of Anosov systems.

I. Introduction. Recently, George Sell and I have been developing a geometric theory of systems of almost periodic (a.p.) differential equations along the lines suggested by Smale (1967) for autonomous or periodic systems. Smale's program seeks global stability results and rest heavily on the concept of a hyperbolic structure. One of the main tools of this theory is the shadowing lemma of Anosov (1967) and Bowen (1975).

Miller (1965) and Sell (1967) showed how to embed the solutions of an almost periodic system of differential equations in a dynamical system . This dynamical system is a skew product flow over the translation flow on the hull of the a.p. equations. This embedding introduces geometric techniques into the theory of a.p. systems.

In Meyer and Sell (1987a), we present a simple analytic proof of the classical shadowing lemma which easily generalizes to the skew product systems of Miller and Sell. In Meyer and Sell (1987c) we present a slightly different generalization of the shadowing lemma. In Meyer and Sell (1987b,c), we give a generalization the Smale horseshoe basic set and Melnikov's method to a.p. systems. This paper will give a generalization to a.p. systems of the Anosov

1980 Mathematics Subject Classification (1985 Revision). 58F15, 58F27

[1]This research was supported by a grant from the Applied and Computational Mathematics Program of DARPA.

(1967) stability theorem. The proof of this stability theorem is a simple application of the generalized shadowing lemma given in our previous papers once we establish that the generalized Anosov systems are open in the appropriate topology.

Although one usually thinks of Smale's program as dealing solely with dissipative systems, both the horseshoe and Anosov systems appear in Hamiltonian systems. In fact Poincaré (1899) discussed transverse homoclinic orbits, which imply horseshoes, in the restricted three body problem. Geodesic flows on manifolds with negative curvature are Hamiltonian Anosov systems -- see Anosov (1967). Markus and Meyer (1974) give another example of Hamiltonian Anosov system.

The Section II gives a brief introduction to some of the basic geometric results about almost periodic systems. In particular the hull of an a.p. function, the translation flow on the hull, the existence of cross sections, and almost periodic suspensions are defined and discussed. Section III gives the Miller-Sell embedding of the solutions of a system of a.p. equations into a skew product dynamical system. It also gives the definitions of a skew Anosov system, skew equivalence and skew structural stability. With these definitions the main theorem says the skew Anosov systems are skew structurally stable. Section IV contains a discussion of the shadowing lemma for skew Anosov systems, the proof of the openness of skew Anosov systems and the proof of the structural stability of skew Anosov systems using these two facts.

II. The Hull, Cross Sections, and Suspensions. Throughout this paper *almost periodic* (a.p.) will be in the sense of Bohr(1959). Besicovitch (1932), Bohr (1959), Favard (1933) and Fink (1974) are good general references on almost periodic functions and differential equations. The examples and some of the other elementary facts given here are discussed in more detail in Meyer and Sell (1987c). Let $C = C(R,R^n)$ (or $C(R,C^n)$) denote the space of continuous functions from R into R^n (or C^n) with the topology of uniform convergence on compact set -- the compact open topology. Translations define a flow on C as follows

$$(1) \qquad \pi : C \times R \longrightarrow C : (f,\tau) \rightarrow f_\tau$$

where $f_\tau(t) = f(t+\tau)$. For any $f \in C$ the orbit closure of f is called the *hull* of f and is denoted by H(f). If f is a.p. then H(f) is a compact minimal set; each element $g \in H(f)$ is a.p. with $H(f) = H(g)$; $\pi|H(f)$ is equicontinuous; and

$H(f)$ can be given a compact, connected Abelian group structure. We also let **AP** denote the space of a.p. functions with the sup norm on R. The hull is defined in this space and the above results hold there also -- see Sell (1971).

If f is a.p., its associated Fourier series will be denoted by

$$f \sim \sum a_k \exp i\omega_k t. \tag{2}$$

It follows that $f_\tau \sim \sum a_k \exp i\omega_k(t+\tau)$. If $f_{\tau_n} \longrightarrow g$, use the Cantor diagonal procedure to select a subsequence if necessary such that

$$\tau_n \longrightarrow \alpha_k \mod 2\pi/\omega_k \text{ as } n \longrightarrow \infty, \text{ for all } k. \tag{3}$$

Then the Fourier coefficients of f_τ converge to the Fourier coefficients of g so

$$g \sim \sum a_k \exp i\omega_k(t+\alpha_k). \tag{4}$$

Thus, if $g \in H(f)$ there are angles α_k defined mod $2\pi/\omega_k$ such that (4) holds.

Example 1: Consider a quasi-periodic function of the form

$$q(t) = a_1\exp i\omega_1 t + a_2\exp i\omega_2 t \tag{5}$$

where ω_1/ω_2 is irrational and a_1, a_2 are real. In this case

$$H(q) = \{ a_1\exp i\omega_1(t+\alpha_1) + a_2\exp i\omega_2(t+\alpha_2) \colon \alpha_i \text{ defined mod } 2\pi/\omega_i \}.$$

Thus the two angles α_1, α_2 are coordinates for $H(q)$, or $H(q)$ is homeomorphic to the two torus.

Example 2: Consider a limit periodic function of the form

$$\ell(t) = \sum_0^\infty a_k \exp i2\pi \left(\frac{t}{2^k} \right) \tag{6}$$

where the a_k are chosen so that the series converges absolutely and uniformly. In this case $g \in H(\ell)$ if and only if

$$g(t) = \sum_0^\infty a_k \exp i2\pi \left(\frac{t + \alpha_k}{2^k} \right) \tag{7}$$

where the angles α_k are defined mod 2^k and satisfy $\alpha_k \equiv \alpha_{k+1} \mod 2^k$. In this case $H(\ell)$ is homeomorphic to the standard solenoid -- see Figure 1.

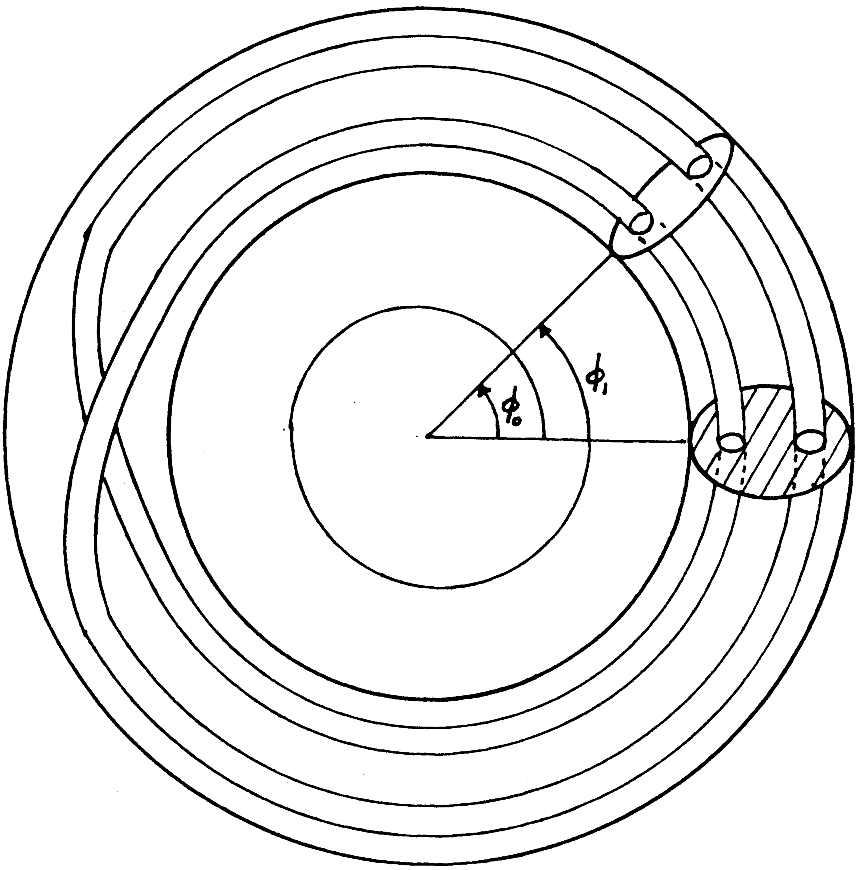


Figure 1. $H(\ell)$ -- The solenoid.

A flow $\sigma : X \times R \longrightarrow X$, X a compact metric space, admits a (*global*) *cross section* Z if i) Z is a closed subset of X, ii) all trajectories meet Z, and iii) there is a positive continuous function $T : Z \longrightarrow R$ such that $\sigma(z, T(z)) \in Z$ for all $z \in Z$ and $\sigma(z,t) \notin Z$ for $0 < t < T(z)$. The function T is called the *first return time*. The *Poincare map* (or *section map*) is the map

$$P : Z \longrightarrow Z : z \longrightarrow \sigma(z, T(z)). \tag{8}$$

The translation flow on the hull of an almost periodic function always admits a cross section. Let f be a.p. and have a Fourier series given in (2) then if $g \in H(f)$

$$g_\tau \sim \sum a_k \exp i\omega_k(t+\alpha_k+\tau) \sim \sum a_k \exp i\omega_k(\alpha_k+\tau) \exp i\omega_k t \tag{9}$$

thus the Fourier coefficient corresponding to the frequency ω_k is $a_k \exp i\omega_k(\alpha_k+\tau)$ which is has a constantly changing argument as τ varies provided $\omega_k \neq 0$. Thus a cross section to the translation flow on $H(f)$ is

$$Z = \{\ g \in H(f) : \arg (\ a_k \exp i\omega_k(\alpha_k+\tau)\ ) = 0\ \}. \tag{10}$$

In this case the first return time is $2\pi/\omega_k$ and the Poincaré map defines a discrete a.p. dynamical system.

Example 1. A cross section to the translation flow on $H(q)$ is $\alpha_1 \equiv 0 \bmod 2\pi/\omega_1$ and α_2 can be used as a coordinate it this cross section. In this case the Poincaré map is the irrational rotation of the circle $P : \alpha_2 \longrightarrow \alpha_2 + (\omega_2/\omega_1)2\pi$.

Example 2. A cross section to the translation flow on $H(\ell)$ is $\phi_1 \equiv 0 \bmod 1$ -- the shaded disk in Figure 1. Topologically, this cross section is a Cantor set and the associated Poincaré map is equivalent to the classical adding machine. The adding machine is the dynamical system

$$\gamma : \prod_0^\infty \{0,1\} \longrightarrow \prod_0^\infty \{0,1\} : \ldots a_2a_1a_0. \longrightarrow \ldots a_2a_1a_0. + 1, \tag{11}$$

i.e. the space is all binary integers with the product topology and the map adds 1 to a binary integer. See Meyer and Sell (1987c) for more details.

Let $P : Z \longrightarrow Z$ be a discrete a.p. dynamical system, say the irrational rotation of the circle or the adding machine. Let $D : X \longrightarrow X$ be a discrete dynamical system, i.e. D is a homeomorphic of the topological space X. The *P - almost periodic suspension of* D is defined as the suspension of the product system $P \times D : Z \times X \longrightarrow Z \times X : (z,x) \longrightarrow (P(z), D(x))$. That is, first define the parallel flow

$$\gamma : (Z \times X \times R) \times R \longrightarrow (Z \times X \times R) : ((z,x,\tau),t) \longrightarrow (z,x,\tau+t)$$

and then drop this flow to the quotient space $(Z \times X \times R)/\sim$ where $\sim$ is the equivalence relation $(z,x,\tau) \sim (P(z),D(x),\tau+1)$.

III. Skew Product Flows and Skew Anosov Systems. Now let G be the space of functions f from $R^n \times R^1$ into R^n such that for every compact set $K \subset R^{2n}$, (i) the function is uniformly continuous on $K \times R$ and (ii) there is a constant k such that

$$|f(x,t) - f(y,t)| < k|x - y|, \quad t \in R, \ x,y \in K.$$

Let G be given the compact open topology. Define the flow $\pi : G \times R \longrightarrow G : (F,\tau) : \longrightarrow F_\tau$ where $F_\tau(x,t) = F(x,t+\tau)$ and define the hull as before. Let $L(x,t) \in G$ be almost periodic in t uniformly in x (u.a.p.). Consider the system of differential equation

$$\dot{x} = F(x,t), \quad F \in H(L). \tag{1}$$

This might be a Hamiltonian system on an even dimensional space. Let $\Phi(t,x,F)$ be the solution of (1) such that $\Phi(0,x,F) = x$. Assume that Φ is defined for all $t \in R$, $x \in R^n$, $F \in H(L)$. Miller (1965) defined a flow on $R^n \times H(L)$ by

$$\begin{aligned} \Pi : (R^n \times H(L)) \times R &\longrightarrow R^n \times H(L) \\ : ((x, F), t) &\longrightarrow (\Phi(t,x,F), F_t). \end{aligned} \tag{2}$$

This is an example of a skew product flow, where the space is a product and the flow acting on the second factor is a flow in its own right. Under the general assumption of smooth F in (1) the function Φ and its first partial with respect to x will be continuous on $R^n \times H(L)$, but it makes no sense to speak of a partial derivative of Φ with respect to F because $H(g)$ is not a manifold in general. See Sell (1971) for a general discussion and more details.

Example 3. Consider the differential equation

$$\dot{x} = f(x) + \varepsilon p(t), \qquad p \in H(r) \tag{3}$$

where $f : R^n \longrightarrow R^n$ is smooth and $r : R \longrightarrow R^n$ is a.p. Also assume that solutions are defined for all t and for all values of the small parameter ε. Let Z be any cross section for the flow on the hull of r with constant first return time T and Poincaré map P. Let $\phi(t,x)$ be the solution of (3) such that

$\phi(0,x) = x$ when $\varepsilon = 0$. In order to do a perturbation analysis one considers (3) as an a.p. system even when $\varepsilon = 0$. That being the case, when $\varepsilon = 0$ the dynamical system defined by (2) equivalent to the P-almost periodic suspension of $\phi(T,x)$. Thus we can consider (3) as a perturbation problem where the unperturbed system is a P-almost periodic suspension. Notice that in this example the perturbation would not change the flow on the base, i.e. the translation flow on the hull of r would be the same for all values of the perturbation parameter ε. This is the motivation for the definitions given below.

Let $P : Z \longrightarrow Z$ be a discrete a.p. dynamical system and M a smooth, connected, compact manifold. Then $A : M \times Z \longrightarrow M \times Z$ will be called a *skew Anosov system (over P)* if

i) A is a skew product system over P, i.e. $A(m,z) = (B(m,z),P(z))$;

ii) $B : M \times Z \longrightarrow M$, has a continuous partial derivative with respect to its first argument, denoted by D_1B;

iii) there exist subspaces $E^s_{(m,z)}$ and $E^u_{(m,z)}$ such that

(4) $$T_mM = E^s_{(m,z)} \oplus E^u_{(m,z)} \text{ for all } (m,z) \in M \times Z$$

and this splitting is continuous;

iv) $D_1B(m,z) : E^s_{(m,z)} \longrightarrow E^s_{A(m,z)} \quad D_1B(m,z) : E^u_{(m,z)} \longrightarrow E^u_{A(m,z)}$

v) there are constants $C > 0$ and $0 < \lambda < 1$ such that

(5)
$$\| D_1B^n(m,z)(u) \| \le C\lambda^n \| u \| \text{ for } u \in E^s_{(m,z)} \text{ and } n > 0$$
$$\| D_1B^{-n}(m)(u) \| \le C\lambda^n \| u \| \text{ for } u \in E^u_{(m,z)} \text{ and } n > 0,$$

and all $(m,z) \in M \times Z$.

Let $A_i : M \times Z \longrightarrow M \times Z : (m,z) \longrightarrow (B_i(m,z),P(z))$ $i = 1,2$ be two skew product systems over the same base $P : Z \longrightarrow Z$. We say A_1 and A_2 are *skew equivalent* if there is a homeomorprism $H : M \times Z \longrightarrow M \times Z$ $: (m,z) \longrightarrow (h(m,z),z)$ such that the following diagram commutes:

$$
\begin{tikzcd}
M \times Z \arrow[r, "A_1"] \arrow[d] & M \times Z \arrow[d, "H"] \\
M \times Z \arrow[r, "A_2"'] & M \times Z
\end{tikzcd} \tag{6}
$$

Note that here and several times below we treat the second variable differently. Here we require that H be the identity map on the second factor. Thinking about the differential equation examples given above these seems natural since the second factor corresponds to the time translate of the equations. Thus H does not change the clock.

Let $C_P^1 = C_P^1(M \times Z, M \times Z)$ be the space of functions $\Phi : M \times Z \to M \times Z : (m,z) \to (\phi(m,z), P(z))$ where ϕ has a continuous first partial with respect to it first argument and we place the topology of uniform convergence of the functions and there first partial with respect to its first argument. That is two such functions are close if their values are close and their first partials are close. We say that $\Phi \in C_P^1$ is *skew structurally stable* if the is a neighborhood **N** of Φ in C_P^1 such that if $\Psi \in \mathbf{N}$ then Φ and Ψ are skew equivalent. The main result of this note is:

Theorem: Skew Anosov systems are skew structurally stable.

IV. The Shadowing Lemma, Openness, and the Proof of Structural Stability. Let $P : Z \to Z$ be a discrete a.p. dynamical system, M a smooth compact, connected manifold and $A : M \times Z \to M \times Z : (m,z) \to (B(m,z), P(z))$ be a discrete skew product dynamical system. For $\alpha > 0$ a *(skew) α-pseudo-orbit* for A is a bisequence $\{(m_i, z_i)\}$, $-\infty < i < \infty$, with $z_{i+1} = P(z_i)$ and $d(m_{i+1}, B(m_i, z_i)) < \alpha$ for all i. Here d is some distance function on M. Note that $\{z_i\}$ is a P-orbit and so we allow jumps of distance α in the M direction only. If we think in terms of the differential equation examples of the previous section this means we allow jumps in the solutions of one equation but do not allow a jump in the equations. An A-orbit $\{ A^i(m_0, z_0) = (m_i, z_i) \}$ *(skew) β-shadows* an α-pseudo-orbit $\{ (p_i, z_i) \}$ if $d(m_i, p_i) < \beta$ for all i and of course $z_{i+1} = P(z_i)$. Note that the base orbits are the same. In Meyer and Sell (1987c) we give a simple proof of:

Theorem (The skew shadowing lemma): If A *is an skew Anosov system, then for every* $\beta > 0$ *there is an* $\alpha > 0$ *such that every* α*-pseudo-orbit is* β*-shadowed by an* A*-orbit. Moreover, there is a* $\beta_0 > 0$ *such that if* $0 < \beta < \beta_0$ *then the* A*-orbit given above is uniquely and continuously determined by the* α*-pseudo-orbit.*

Continuity means that the map which sends $p_0 \longrightarrow m_0$ is continuous. The constant β_0 is a function of the constants C and λ in the definition of an Anosov system.

A is *(skew) expansive* if there is an $\varepsilon > 0$ such that given any two A-orbits $\{ A^i(m,z) \}$ and $\{ A^i(p,z) \}$ with $m \neq p$ there is some j such that $d(B^j(m,z),B^j(p,z)) > \varepsilon$. Note that the second argument is the same. Again thinking in terms of the differential equations the expansiveness is for the solutions of one equation. In Meyer and Sell (1987c) an immediate corollary of the proof of the skew shadowing lemma is:

Corollary: Skew Anosov systems are skew expansive.

In fact the ε can be taken as the β_0 of the shadowing lemma and therefore is a function of the constants C and λ in the definition of an Anosov system.

Here we shall give a new definition of skew Anosov which is different from the one given in the previous section. In the old definition the manifold M was given one Riemannian metric and the estimates in III.6 contained a constant C. In the new definition we assume that $A : M \times Z \longrightarrow M \times Z : (m,z) \longrightarrow (B(m,z),P(z))$ satisfies conditions i), ii), iii), and iv) of the old definition but change v). Now assume that for each $z \in Z$, M is given a metric $(\ ,\ )_z : TM \times TM \longrightarrow R$ which varies continuously with z and which in tern defines a norm $\|\ \|_z : TM \longrightarrow R$. Assume there is a constant $0 < \lambda < 1$ such that

$$\text{v}')\quad \| D_1B(m,z)(u) \|_{P(z)} < \lambda \| u \|_z \text{ for } u \in E^s_{(m,z)}$$

$$\| D_1B^{-1}(m,z)(u) \|_{P^{-1}(z)} < \lambda \| u \|_z \text{ for } u \in E^u_{(m,z)},$$

and all $(m,z) \in M \times Z$.

Lemma: The new and old definition of skew Anosov system are equivalent.

Proof. That the old definition implies the new is proved precisely in the same way as Proposition 4.2 of Shub (1987). Assume that A satisfies the new definition as given above and fix $w \in Z$. Since M and Z are compact and the metric varies continuously there is a constant $K \geq 1$ such that

$$(1) \qquad K^{-1} \| u \|_z < \| u \|_w < K \| u \|_z,$$

for all $u \in T_pM$, $p \in M$ and $z \in Z$. Iterating (1) for $u \in E^s_{(m,z)}$ gives

$$\| D_1B^n(m,z)(u) \|_{P^n(z)} < \lambda^n \| u \|_z$$

for $u \in E^s_{(m,z)}$ and using (2) gives

$$(2) \qquad K^{-1} \| D_1B^n(m,z)(u) \|_w < K \lambda^n \| u \|_w .$$

And similarly for $u \in E^u_{(m,z)}$. Thus the old definition holds with the single metric $(\ ,\ )_w$ on M with the constant $C = K^2$.

Theorem: The set of Anosov systems is an open set in $C^1_P(MxZ, MxZ)$.

Proof: Let $A : M \times Z \to M \times Z : (m,z) \to (B(m,z),P(z))$ be an Anosov diffeomorphism by the new definition given above and $A' : M \times Z \to M \times Z : (m,z) \to (B'(m,z),P(z))$ be close to A in the C^1_P topology. Let $\mathscr{F}^1 = \mathscr{F}^1(M,Z)$ be the space of C^1 vector field depending on a parameter $z \in Z$, i.e. $X \in \mathscr{F}^1$ if $X : M \times Z \to TM \times Z : (m,z) \to (Y(m,z),z)$ is continuous, has a continuous partial derivative with respect to it first argument, denoted by D_1X, and $Y(m,z) \in T_mM$ for all $(m,z) \in M \times Z$. Place on $\mathscr{F}^1$ the topology of uniform convergence of functions and their first partial derivative with respect to their first argument. Define mappings $F, F' : \mathscr{F}^1 \to \mathscr{F}^1$ by the formulas:

$$(3) \qquad \begin{aligned} F(X)(m,z) &= (D_1B(A^{-1}(m,z))(Y(A^{-1}(m,z)),z) = (G(X)(m,z),z) \\ F'(X)(m,z) &= (D_1B'(A'^{-1}(m,z))(Y(A'^{-1}(m,z)),z) = (G'(X)(m,z),z) \end{aligned}$$

The tangent bundle $TM \times Z = \cup\, T_mM \times Z$ (union on $m \in M$) admits a decomposition

$$(4)\qquad \begin{aligned} &TM \times Z = E^s \oplus E^u \\ &E^s = \cup\, E^s_{(m,z)}, \quad E^u = \cup\, E^u_{(m,z)} \end{aligned}$$

where the latter unions are over all $(m,z) \in M \times Z$. The first factor of F and F', G and G', are linear and so using the splitting (4) we can write

$$(5)\qquad G = \begin{bmatrix} G_{++} & 0 \\ 0 & G_{--} \end{bmatrix}, \qquad G' = \begin{bmatrix} G'_{++} & G'_{+-} \\ G'_{-+} & G'_{--} \end{bmatrix}$$

The matrix for G is diagonal since the splitting is invariant for A. By v') and the fact that we have taken A' close to A in the C^1_p topology it follows that

$$(6)\qquad \begin{aligned} &\| G_{++} u \| < \lambda \| u \| \quad \text{and} \quad \| G'_{++} u \| < \lambda \| u \| \quad \text{for } u \in E^s \\ &\| G^{-1}_{--} v \| < \lambda \| v \| \quad \text{and} \quad \| G'^{-1}_{--} v \| < \lambda \| v \| \quad \text{for } v \in E^u \\ &\| G'_{+-} v \| < \varepsilon \| v \| \quad \text{for } v \in E^u, \quad \| G'_{-+} u \| < \varepsilon \| u \| \quad \text{for } u \in E^s \end{aligned}$$

where $0 < \lambda < 1$ and ε can be taken arbitrarily small by taking A' close to A.

Let $\mathcal{L} = \mathcal{L}(E^s, E^u)$ be the space of continuous vector bundle maps with the sup norm, i.e. $L \in \mathcal{L}$, $L(m,z) : E^s_{(m,z)} \to E^u_{(m,z)}$ is linear. We want to find L so that $\{ (u, Lu) : u \in E^s \}$ is F' invariant subspace. Since

$$(7)\qquad G' \begin{bmatrix} u \\ Lu \end{bmatrix} = \begin{bmatrix} G'_{++} & G'_{+-} \\ G'_{-+} & G'_{--} \end{bmatrix} \begin{bmatrix} u \\ Lu \end{bmatrix} = \begin{bmatrix} G'_{++}u + G'_{+-}Lu \\ G'_{-+}u + G'_{--}Lu \end{bmatrix}$$

invariance takes the form

$$(8)\qquad L\, G'_{++} + L\, G'_{+-} L = G'_{-+} + G'_{--}\, L$$

or

$$(9)\qquad L = G'^{-1}_{--} \{ -G'_{-+} + L\, G'_{++} + L G'_{+-} L \}$$

Define an operator $\mathcal{T} : \mathcal{L} \longrightarrow \mathcal{L}$ by

$$(10) \qquad \mathcal{T}(L) = G'^{-1}_{--} \{ -G'_{-+} + L\, G'_{++} + LG'_{+-}L_{-} \}$$

so a fixed point of $\mathcal{T}$ solves (10). Let

$$(11) \qquad \mathsf{L} = \{ L \in \mathcal{L} : \| L \| = \sup_{(m,z)} \sup_{\|x\|=1} \| L(m,z)(x) \| \le 1 \}.$$

If $L \in \mathsf{L}$ then

$$(12) \qquad \begin{aligned} \| \mathcal{T}(L) \| &\le \| G'^{-1}_{--} \| (\| G'_{-+} \| + \| L \| \| G'_{++} \| + \| L \|^2 \| G'_{+-} \| \\ &\le \lambda (\varepsilon + \lambda + \varepsilon) \le 1 \end{aligned}$$

provided ε is sufficiently small, so $\mathcal{T} : \mathsf{L} \longrightarrow \mathsf{L}$. Furthermore, for L, $K \in \mathsf{L}$

$$(13) \qquad \begin{aligned} \| \mathcal{T}(L) - \mathcal{T}(K) \| &\le \| G'^{-1}_{--} \| \{ \| L - K \| \| G'_{++} \| + \| LG'_{+-}L - KG'_{+-}K \| \} \\ &\le \lambda \{ \lambda \| L - K \| + \| LG'_{+-}(L-K) \| + \| (K-L)G'_{+-}K \| \\ &\le \lambda \{ \lambda + 2\varepsilon \} \| L - K \| \end{aligned}$$

and so for ε sufficiently small $\mathcal{T}$ is a contracting map which has a unique fixed point L in L.

Thus we have constructed a bundle $E'^{s} = \{ (u,Lu) : u \in E^{s} \}$ which is F' invariant. The bundle $E'^{u} = \{ (Ku,u) : u \in E^{u} \}$ is constructed in a similar manner. By construction both K and L have norm less than 1 and the dimensions of the fibers of E'^{s} and E^{s} are the same as are those of E'^{u} and E^{u}. If $v = (v^{s}, v^{u}) \in E'^{s}_{(m,z)} \cap E'^{u}_{(m,z)}$ then $v^{u} = Lv^{s} = LKv^{u}$ but since the norms of L and K are less that 1 this implies $v^{u} = v^{s} = 0$. Thus $TM \times Z = E'^{s} \oplus E'^{u}$. The estimates of the form (1) follow at once from the inequalities (7).

Proof of the structural stability of Anosov systems.

Let A be an Anosov system where $A : M \times Z \longrightarrow M \times Z : (m,z) \longrightarrow (B(m,z),P(z))$ and first fix α so that all functions in this α-neighborhood of A are Anosov with the same constants C and λ. Let $\varepsilon > 0$ be the uniform expansive constant and β_0 the uniform constant of the shadowing lemma for all functions in this neighborhood. Let $\beta = \min(\ \varepsilon/3,\ \beta_0/3\ )$ and restrict α further if necessary so that the conclusion of the shadowing lemma holds for this α and β and $\alpha < \beta_0$. Let $D : M \times Z \longrightarrow M \times Z : (m,z) \longrightarrow (E(m,z),P(z))$ be within this α of neighborhood of A. Let $(m,z) \in M \times Z$ be arbitrary.

Then since A and D are α close $\{\ D^i(m,z)\ \}$ is an α-pseudo-orbit for A and so there exists a $y = h(m,z)$ such that the A-orbit $\{\ A^i(y,z)\ \}$ β-shadows $\{\ D^i(m,z)\ \}$. The function $h : M \times Z \longrightarrow M$ is continuous by the shadowing lemma and hence so is $H : M \times Z \longrightarrow M \times Z : (m,z) \longrightarrow (h(m,z),z)$. Let $(m,z) \neq (m',z')$. Clearly if $z \neq z'$ $H(m,z) \neq H(m',z')$ so let $z = z'$ and $m \neq m'$. By the expansive property of D there is a j such that $d(E^j(m,z),E^j(m',z)) > \varepsilon$. But $d(E^j(m,z),B^j(y,z)) < \beta \le \varepsilon/3$ and $(E^j(m',z),B^j(y',z)) < \beta \le \varepsilon/3$ and so $d(B^j(y,z),B^j(y',z)) > \varepsilon/3$ or $y \neq y'$. Therefore h and H are one to one. Thus for fixed $z \in Z$ the map $h(.,z) : M \longrightarrow M$ is a continuous, one-to-one mapping of a compact, connected Hausdorff space and so is a homeomorphism. This implies that H is a homeomorphism also.

Since $d(E^i(m,z),B^i(y,z)) < \alpha$ for all i we have

$$d(E^{i-1}(E(m,z),z),B^{i-1}(B(y,z),z) = d(E^i(m,z),B^i(y,z)) < \alpha < \beta_0.$$

Thus the A orbit through $A(y,z) = (B(y,z),z)$ β_0-shadows the D-orbit through $D(m.z) = (E(m,z),z)$and so by uniqueness $A(y,z) = H(D(m,z))$. But $(y,z) = H(m,z)$ so $A \circ H = H \circ D$ or H is a skew equivalence.

References

Anosov, D. V. 1967 : *Geodesic flows on compact Riemannian manifolds of negative curvature*, Proc. Steklov Math. Inst. 90.

Besicovitch, A.S. 1932: **Almost Periodic Functions**, Cambridge Univ. Press.

Bohr, H. 1959: **Almost Periodic Functions**, Chelsea Publ., New York.

Bowen, R. 1975: *ω - limit sets for Axiom A diffeomorphisms*, **J. Diff. Eqs.** 18, 333-9.

Favar, J. 1933: **Lecons sur les fonctions presque periodiques**, Gauthier-Villars, Paris.

Fink, A.M. 1975: **Almost Periodic Differential Equations**, Lect. Notes in Math. 377, Springer-Verlog, New York.

Markus, L. and Meyer, K.

1974: *Generic Hamiltonian systems are neither integrable nor dense*, **Mem. Amer. Math. Soc.** 144.

1980: *Periodic orbits and solenoids in generic Hamiltonian systems*, **Amer. J. Math.** 102, 25-92.

Mather, J. 1968: *Characterization of Anosov diffeomorphisms*, **Nederl. Akad. Wetensch. Indag. Math.** 30, 479-95.

Meyer, K. R. and Sell, G. R.

1987a: *An analytic proof of the shadowing lemma*, to appear in **Funkcialaj Ekvacioj**.

1987b: *Homoclinic orbits and Bernoulli bundles in almost periodic systems*, **Oscillations, Bifurcations and Chaos**, CMS-AMS 8, Providence, R.I., 527-44.

1987c: *Melnikov transforms, Bernoulli bundles, and skew product flows*, to appear.

Miller, R. K. 1965: *Almost periodic differential equations as dynamical systems with applications to the existence of a.p. solutions*, **J. Diff. Eqs.** 55, 337-45.

Poincaré, H. 1899: **Les methods nouvelles de la mechanique celeste III**, Gauthier-Villar, Paris.

Poinkoff, H. G. 1973: *An analytic closing lemma*, **Proc. Midwest Dyn. Sys. Sem.**, Northwestern Univ. Press, Evanston, Ill, 128-256.

Sell, G. R.

1967: *Nonautonomous differential equations and topological dynamics I, II*, **Trans. Amer. Math. Soc.** 127, 241-62 & 263-83.

1971: **Topological Dynamics and Ordinary Differential Equations**, Van Nostrand, New York.

Shub, M. 1987: **Global Stability of Dynamical Systems**, Springer-Verlog, New York.

Smale, S. 1967: *Differentiable dynamical systems*, **Bull. Amer. Math. Soc.** 73, 747-817.

Department of Mathematical Sciences
University of Cincinnati
Cincinnati, Ohio 45208

http://dx.doi.org/10.1090/conm/081/986265

Contemporary Mathematics
Volume 81, 1988

THE PRESCRIBED ENERGY PROBLEM FOR PERIODIC SOLUTIONS OF HAMILTONIAN SYSTEMS

Paul H. Rabinowitz

During the past several years there has been considerable progress in the study of the existence and multiplicity of periodic solutions of prescribed energy of Hamiltonian systems. The goal of this talk is to survey this work. Let $H = H(p,q)$ denote the Hamiltonian where $p,q \in \mathbb{R}^n$. For simplicity suppose H is smooth. (For the results below H continuously differentiable is generally sufficient.) The corresponding Hamiltonian system is

$$\dot{p} = -H_q(p,q) \qquad (1)$$
$$\dot{q} = H_p(p,q) \; .$$

Setting $z = (p,q)$ and $J = \begin{pmatrix} 0 & -id \\ id & 0 \end{pmatrix}$ where id denotes the $n \times n$ identity matrix, (1) can be written more succinctly as

$$\dot{z} = JH_z(z) \; . \qquad (2)$$

As is well known, if $z(t)$ is a solution of (2), $H(z(t)) \equiv$ constant. The basic question we will be concerned with here is what sort of geometrical or topological conditions must be imposed on an energy surface so that it contains a periodic solution. For definiteness we take the prescribed energy to be 1 and set $M \equiv H^{-1}(1)$ in what follows.

The first general result for (2) of the above type that we know of is due to Seifert [1] in 1948. Using geodesic ideas from geometry he proved

THEOREM 3: <u>Suppose</u> $H(p,q) = \sum_{i,j=1}^{n} a_{ij}(q)p_i p_j + V(q)$ <u>with</u> a_{ij}, V <u>smooth and satisfying</u>

1980 Mathematics Subject Classification (1985 Revision). 34C25, 58E30, 58F05, 70H05.

This research was sponsored in part by the United States Army under Contract No. DAAL03-87-K-0043, the Air Force Office of Scientific Research under Grant No. AFOSR-87-0202 and the National Science Foundation under Grant No. MCS-8110556.

(V_1) $\mathcal{D} \equiv \{q \in R^n | V(q) < 1\}$ is diffeomorphic to the closed unit ball in $\mathbf{R}^n$

and

(K_1) the matrix $(a_{ij}(q))$ is uniformly positive definite in $\mathcal{D}$.

Then there is a $T > 0$, a pair of points $Q_1, Q_2 \in \partial\mathcal{D}$, and a solution $(p(t),q(t))$ of (2) such that $(p(0),q(0)) = (0,Q_1)$, $(p(t),q(t)) = (0,Q_2)$, and $q(t)$ lies in the interior of $\mathcal{D}$ for $t \in (0,T)$.

Observing that H is even in p and reflecting p and q about 0 and T as odd and even functions $P(t)$, $Q(t)$ respectively, it is easy to see that (P,Q) is a $2T$ periodic solution of (2). Note also that $(P(t),Q(t)) \in M$ for all $t \in \mathbf{R}$.

The recent developments for (2) begin with work appearing in 1978 of Weinstein [2] and Rabinowitz [3]. In [2], Weinstein proved a generalization of Theorem 3 and used it to show if M bounds a compact convex neighborhood of 0 in R^{2n}, then (2) possesses a periodic solution on M. His techniques are in the spirit of Seifert's. In [3], minimax methods from the calculus of variations were used to prove that if M bounds a compact starshaped neighborhood of 0 in R^{2n}, then M contains a periodic solution of (2).

Subsequent to [2] and [3], other sufficient conditions have been given on H under which (2) possesses a periodic orbit on M. See e.g. Rabinowitz [4], Weinstein [5], Gluck and Ziller [6], Hayashi [7], and Benci [8]. In particular in [5], Weinstein observed that the problems studied in [1-4] all possessed a common differential geometric feature, namely M was of contact type in R^{2n} and $H^1(M) = 0$. He conjectured that more generally any compact hypersurface in a symplectic manifold with $H^1(M) = 0$ would contain a periodic trajectory. This conjecture was recently proved (September, 1986) by C. Viterbo [9] in the R^{2n} setting without the $H^1(M)$ condition. At a conference on Periodic Solutions of Hamiltonian Systems held in Il Ciocco, Italy in October of 1986, H. Hofer and E. Zehnder simplified Viterbo's argument and extended his work to obtain a surprising result [10]:

THEOREM 4. Suppose M is a compact hypersurface in $\mathbf{R}^{2n}$ (and in particular $H_z \neq 0$ on M). Then for any $\delta > 0$, there is an ε such that $|\varepsilon| < \delta$ and $H^{-1}(1+\varepsilon)$ contains a periodic solution z_ε of (2). Moreover, if T_ε is the period of $z_\varepsilon = (p_\varepsilon, q_\varepsilon)$, then there exists a constant β independent of δ such that

$$0 < A(z_\varepsilon) \equiv \int_0^{T_\varepsilon} p_\varepsilon \cdot \dot{q}_\varepsilon dt < \beta .$$

REMARK 5. It is not difficult to show - see [10] - that if M is of contact type, a periodic solution of (2) on $H^{-1}(1 + \varepsilon)$ for small ε leads to a periodic solution of (2) on M. Hence Viterbo's result follows from Theorem 4.

By Theorem 4, there exists a sequence $\varepsilon_m \to 0$ as $m \to \infty$ such that $H^{-1}(1 + \varepsilon_m)$ contains a periodic solution z_m of (2) with $0 < A(z_m) < \beta$. This fact naturally suggests trying to find a periodic solution of (2) on M as a limit of the sequence (z_m). This approach will succeed if the corresponding sequence of periods (T_m) of (z_m) is bounded (at least along a subsequence). (Note that (T_m) cannot tend to 0 since $H_z \neq 0$ on M.) To show that (T_m) is bounded, it suffices to prove that there is an $\alpha > 0$ such that $T_m^{-1}A(z_m) \geq \alpha$ for then $T_m \leq \alpha^{-1}\beta$ via Theorem 4. Such bounds are in fact known for a class of Hamiltonians including those treated in [1-8]. See Benci-Rabinowitz [11] and a slight generalization in Benci-Hofer-Rabinowitz [12]. E.g. in [12] it is shown that:

PROPOSITION 6. If M bounds a compact neighborhood of 0 in $\mathbf{R}^{2n}$ and $p \cdot H_p > 0$ if $p \neq 0$ on M, then there exists an $\alpha > 0$ such that $T^{-1}A(z) \geq \alpha$ for any T-periodic solution z of (2) on M.

Recently we have obtained a slight generalization of Theorem 4 which shows that (2) has a richer structure of periodic solutions near M:

THEOREM 7 [13]. <u>Under the hypotheses of Theorem 4, either (i) there are uncountably many values of</u> ε <u>near</u> 0 <u>such that</u> $H^{-1}(1 + \varepsilon)$ <u>contains a periodic solution of (2) or (ii) there is a sequence</u> $\varepsilon_m \to 0$ <u>as</u> $m \to \infty$ <u>such that</u> $H^{-1}(1 + \varepsilon_m)$ <u>contains uncountably many distinct periodic solutions of (2).</u>

While providing more distinct solutions of (2) than Theorem 4, Theorem 7 sheds no additional light on whether M itself contains a periodic solution of (2). Indeed this remains the major open question in this field. If such an M does not possess a periodic solution of (2), then Theorem 7 shows the structure of the set of periodic solutions of (2) near M must be exceedingly complicated.

Also in [13], a variant of the proof of Theorem 7 shows that:

THEOREM 8. <u>Under the hypotheses of Theorem 7, if</u> H <u>is also even in</u> p, <u>the alternatives of Theorem 7 hold for a family of periodic solutions</u> (p,q) <u>where</u> p <u>is odd about</u> 0 <u>and</u> $\frac{T}{2}$ <u>and</u> q <u>even about</u> 0 <u>and</u> $\frac{T}{2}$, T <u>being the period of</u> (p,q).

We will sketch the proofs of Theorems 4 and 7. Before doing so a few remarks about the state of the theory concerning the multiplicity of periodic solutions on M is in order. Here much less is known. The first major result in this direction is due to Ekeland and Lasry [14] who proved:

THEOREM 9. *If M bounds a compact convex neighborhood of 0 in $\mathbf{R}^{2n}$ and*

$$B_r(0) \equiv \{x \in R^{2n} \mid |x| < r\} \subset H^{-1}((-\infty,1]) \subset B_R(0)$$

where

(10) $$1 < \frac{R}{r} < \sqrt{2}\ ,$$

Then (2) possesses at least n geometrically distinct periodic solutions on M.

There have been some variants of Theorem 9. See e.g. Berestycki, Lasry, Mancini, Ruf [15], Hayashi [16], van Groesen [17], Girardi [18]. However, all of them require a piercing condition like (10). Whether such conditions are essential for a multiplicity result is not known. The basic difficulty is that all current proofs of Theorem 9 involve some sort of comparison argument for which (10) is required.

In another interesting result, Ekeland [19] uses index arguments analogous to those encountered in the theory of closed geodesics to show that in a generic setting if M bounds a compact convex neighborhood of 0 in $\mathbf{R}^{2n}$, then M contains infinitely many distinct periodic solutions of (2). This is a puzzling result. A simple possible explanation of it is the following: By results from [2] or [3] mentioned above, we know M contains at least one periodic solution of (2). Suppose we further knew M always contains one such solution, $z(t)$, of elliptic type. Then the Birkhoff-Lewis Theorem [20] suggests that generically, $z(t)$ is the limit of subharmonic solutions of (2). Thus we are led to conjecture that the arguments that yield Theorems 4 and 7 always give an elliptic periodic solution of (2).

One final multiplicity result that has been announced recently by Ekeland and Hofer [21] is:

THEOREM 11. *If M bounds a compact convex neighborhood of 0 in $\mathbf{R}^{2n}$ with $n > 2$, it contains at least two geometrically distinct periodic solutions of (2).*

See also [22-23] for the case $n > 3$. Optimists conjecture that for a more general class of such surfaces, M contains at least n geometrically distinct periodic solutions of (2).

Now we turn to a sketch of the proofs of Theorems 4 and 7. We begin with the former and show how small modifications lead to the latter. The proof consists of four main parts:

(A) *Equivalence to a new Hamiltonian system*: Suppose that H and $\bar{H}$ belong to $C^1(\mathbf{R}^{2n},\mathbf{R})$, $H^{-1}(a) = \bar{H}^{-1}(b) \equiv S$, a compact hypersurface with $H_z \neq 0 \neq \bar{H}_z$ on S. Then an easy calculus argument shows that any solution of

(12) $\dot{\zeta} = J\bar{H}_z(\zeta)$

on S is a reparametrization of a solution of (2) on S and conversely. In particular, the existence of a periodic solution of (12) on S yields a periodic solution of (2) on S. Thus if a new Hamiltonian, $\bar{H}$, can be found for which the existence of a periodic solution can be established, we also get one for (2).

(B) <u>A variational formulation:</u> Let $W_T^{k,2}(\mathbf{R},\mathbf{R}^{2n})$ denote the Sobolev space of T-periodic functions on $\mathbf{R}$ with values in $\mathbf{R}^{2n}$ which possesses square integrable derivatives of order k. We are interested in particular in $E = W_1^{1/2,2}(\mathbf{R},\mathbf{R}^{2n})$. This space is perhaps simplest to describe in terms of Fourier series where if

$$z = \sum_{j\in\mathbf{Z}} a_j e^{2\pi ijt} \quad (\text{and} \quad a_j = \bar{a}_{-j} \in \mathbf{C}^{2n}) ,$$

the norm on E can be taken to be

$$\|z\|^2 = \sum_{j\in\mathbf{Z}} (1 + 2\pi|j|)|a_j|^2 .$$

Suppose that $\bar{H} \in C^1(\mathbf{R}^{2n},\mathbf{R})$ and grows at most polynomially in z. Consider the functional

$$I(z) \equiv \int_0^1 [p \cdot \dot{q} - \bar{H}(z)]dt . \tag{13}$$

It is easy to verify that $I \in C^1(E,\mathbf{R})$ and any critical point of I is a classical solution of (12) (see e.g. [24, Chapter 6]). Thus, to find periodic solutions of (12), it suffices to find critical points of I on E. Of course, this is not a simple matter since the functional I is highly indefinite and in fact is unbounded from above and below.

(C) <u>Construction of the new Hamiltonian:</u> This is a key step. Hofer and Zehnder show for each $\delta > 0$, there is a choice of $\bar{H} > 0$ such that any 1-periodic solution z of (12) for which $I(z) > 0$ lies in $H^{-1}(1 - \delta, 1 + \delta)$ on a common energy level of H and $\bar{H}$. We refer to [10] for the details. The proof is elementary but insightful and generalizes and extends the construction of Viterbo [9].

Note that step (C) together with (A) and (B) reduces the problem of finding solutions of <u>prescribed energy</u> of (2) to that of finding critical points of <u>prescribed period</u> of (13). Such a trick in a much simpler setting where M bounds a starshaped neighborhood of 0 was used to give a simple proof of the result of [3] mentioned earlier. See e.g. [24, Chapter 6].

(D) <u>Finding a critical point of</u> I: By using a direct minimax argument, Hofer and Zehnder now show I has a positive critical value. (Viterbo for his setting transformed to a dual variational problem, used a

Lyapunov-Schmidt type argument to reduce to a finite dimensional functional, and made a critical point analysis of this simplified problem.) One can interpret the argument of [10] as reproving a special case of an abstract critical point theorem of Benci and Rabinowitz - see e.g. [24, Theorem 5.29] - which could be used as an alternate existence tool.

Combining (A)-(D) and observing that δ is arbitrary yields Theorem 4.

Next we will indicate the modifications that are required in the above argument to get Theorem 7. Step (A) remains the same and (B) changes only to the extent that we introduce an additional parameter λ and study the functional

$$(14) \qquad I_\lambda(z) = \int_0^1 [p \cdot \dot{q} - \lambda\bar{H}(z)]dt$$

on E. Critical points of $I_\lambda(\cdot)$ are classical 1-periodic solutions of

$$(15) \qquad \dot{\zeta} = \lambda J\bar{H}_z(\zeta) .$$

In (C), choosing $\bar{H}$ as earlier, we need only observe that the previous assertions hold for any $\lambda > 0$ and 1-periodic solution z_λ of (15) for which $I_\lambda(z_\lambda) > 0$. Step (D) requires more restrictions relative to λ. In particular using the argument of [10] or Theorem 5.29 of [24] it follows that for all λ near 1, e.g. $\lambda \in [\frac{3}{4}, \frac{5}{4}]$, I_λ has a positive critical value, c_λ. Moreover, we get a uniform positive lower bound for these numbers, i.e. there exists an $\alpha > 0$ such that $c_\lambda > \alpha$ for all $\lambda \in [\frac{3}{4}, \frac{5}{4}]$. For each such λ, let z_λ be a critical point of I_λ corresponding to c_λ. Now an additional step is required:

(E) $\{z_\lambda | \lambda$ near $1\}$ consists of geometrically distinct trajectories.

Assuming (E) for the moment, Theorem 7 follows: By (D), (E), and (B), for each λ near 1, we get geometrically distinct 1-periodic solutions of (15). Then by (C) and (A) reparametrizations of these solutions, which of course are geometrically distinct, are solutions of (2) in $H^{-1}(1 - \delta, 1 + \delta)$. Since there are uncountably many such distinct solutions and δ is arbitrary, the theorem follows.

It remains to indicate why (E) holds. This requires three steps:

(i) For $\lambda \in [\frac{3}{4}, \frac{5}{4}]$, let $j(\lambda)^{-1}$ denote the minimal period of z_λ. Therefore $j(\lambda) \in \mathbb{N}$. The key fact required of $j(\lambda)$ is the existence of $M > 0$ such that $j(\lambda) \le M$ for all $\lambda \in [\frac{3}{4}, \frac{5}{4}]$. Assume this for now together with:

(ii) If $z_\lambda(t)$ and $z_\mu(t)$ are not geometrically distinct,

$$\frac{\lambda}{\mu} = \frac{j(\lambda)}{j(\mu)} \, . \tag{16}$$

Combining (i) and (ii) yields

(iii) If $\rho = \min(\frac{1}{4}, \frac{1}{2M+1})$ and $\lambda, \mu \in \{\sigma \mid |\sigma - 1| < \rho\}$, then z_λ and z_μ are geometrically distinct. Indeed if not, without loss of generality assume $\lambda > \mu$. Then by (i), (ii) and the choice of λ and μ,

$$1 + \frac{2\rho}{1-\rho} = \frac{1+\rho}{1-\rho} > \frac{\lambda}{\mu} = \frac{j(\lambda)}{j(\mu)} \geq \frac{j(\mu)+1}{j(\mu)} \geq 1 + \frac{1}{M} \tag{17}$$

which implies that $(2M+1)\rho > 1$, contrary to the definition of ρ.

To see why (i) and (ii) hold, note first that if $\{j(\lambda) \mid \lambda \in [\frac{3}{4}, \frac{5}{4}]\}$ were unbounded, along a sequence λ_m of λ's such that $j(\lambda_m) \to \infty$, we have $\lambda_m \to \bar{\lambda} \in [\frac{3}{4}, \frac{5}{4}]$. By (B), $z_{\lambda_m}(t) \subset H^{-1}(1-\delta, 1+\delta)$. Hence the functions z_{λ_m} are uniformly bounded in the L^∞ norm. Thus (15) gives uniform bounds for $\dot{z}_{\lambda_m}$ in L^∞. Consequently by the Argela-Ascoli Theorem and (15), z_{λ_m} converges in C^1 to a 1-periodic function w satisfying

$$\dot{w} = \bar{\lambda} J \bar{H}_z(w) \, . \tag{18}$$

Moreover by (D), $I_{\lambda_m}(z_{\lambda_m}) \geq \alpha > 0$ so $I_{\bar{\lambda}}(w) \geq \alpha$. On the other hand since $j(\lambda_m) \to \infty$, w has minimal period 0, i.e. $w \equiv$ constant. Therefore, recalling that $\bar{H} \geq 0$, we have $I_{\bar{\lambda}}(w) \leq 0$, a contradiction.

Finally, to verify (ii), suppose that $z_\mu(t)$ represents the same trajectory as $z_\lambda(t)$. Then there exists a C^1 function $r(t)$ such that $z_\lambda(t) = z_\mu(r(t))$. Hence

$$\dot{z}_\lambda = \lambda J \bar{H}_z(z_\mu(r(t))) = \mu J \bar{H}_z(z_\mu(r(t)))\dot{r} \tag{19}$$

so

$$\dot{r} = \frac{\lambda}{\mu} \tag{20}$$

or

$$r(t) = \frac{\lambda}{\mu} t + \gamma \, . \tag{21}$$

Consequently

$$z_\lambda(t) = z_\mu(\frac{\lambda}{\mu} t + \gamma) = z_\lambda(t + j(\lambda)^{-1}) = z_\mu(\frac{\lambda}{\mu}(t + j(\lambda)^{-1}) + \gamma) \tag{22}$$

which implies that

$$\frac{\lambda}{\mu} j(\lambda)^{-1} \in j(\mu)^{-1}\mathbb{N} \, . \tag{23}$$

Similarly

$$\begin{aligned} z_\mu(\frac{\lambda}{\mu} t + \gamma) &= z_\lambda(t) = z_\mu(\frac{\lambda}{\mu} t + j(\mu)^{-1} + \gamma) = \\ &= z_\mu(\frac{\lambda}{\mu}(t + \frac{\mu}{\lambda} j(\mu)^{-1}) + \gamma) = z_\lambda(t + \frac{\mu}{\lambda} j(\mu)^{-1}) \end{aligned} \tag{24}$$

so

$$\frac{\mu}{\lambda} j(\mu)^{-1} \in j(\lambda)^{-1} \blacksquare . \tag{25}$$

Combining (23) and (25) yields (16) and completes the sketch of the proof of Theorem 7.

BIBLIOGRAPHY

[1] Seifert, H., Periodische Bewegungen mechanischen Systeme, Math. Z. **51** (1948), 197-216.

[2] Weinstein, A., Periodic orbits for convex Hamiltonian systems, Ann. of Math. **108** (1978), 507-518.

[3] Rabinowitz, P. H., Periodic solutions of Hamiltonian systems, Comm. Pure Appl. Math. **31** (1978), 157-184.

[4] Rabinowitz, P. H., Periodic solutions of a Hamiltonian system on a prescribed energy surface, J. Differential Equations **33** (1979), 336-352.

[5] Weinstein, A., On the hypotheses of Rabinowitz's periodic orbit theorems, J. Differential Equations **33** (1979), 353-358.

[6] Gluck, H. and W. Ziller, Existence of periodic solutions of conservative systems, Seminar on Minimal Submanifolds, Princeton Univ. Press (1983), 65-98.

[7] Hayashi, K., Periodic solutions of classical Hamiltonian systems, Tokyo J. Math. **6** (1983), 473-486.

[8] Benci, V., Closed geodesics for the Jacobi metric and periodic solutions of prescribed energy of natural Hamiltonian systems, Ann. Inst. H. Poincaré. Anal. Non Linéaire **1** (1984), 401-412.

[9] Viterbo, C., A proof of the Weinstein conjecture in R^{2n}, preprint.

[10] Hofer, H. and E. Zehnder, Periodic solutions on hypersurfaces and a result by C. Viterbo, to appear Invent. Math.

[11] Benci, V. and P. H. Rabinowitz, A priori bounds for periodic solutions of Hamiltonian systems, to appear in Ergodic Theory and Dynamical Systems.

[12] Benci, V., H. Hofer, and P. H. Rabinowitz, A remark on a priori bounds and existence for periodic solutions of Hamiltonian systems, to appear in Periodic Solutions of Hamiltonian Systems and Related Topics, A. Ambrosetti, I. Ekeland, P. H. Rabinowitz, and E. Zehnder, eds., Reidel Publishers.

[13] Rabinowitz, P. H., On a theorem of Hofer and Zehnder, to appear in Periodic Solutions of Hamiltonian Systems and Related Topics, A. Ambrosetti, I. Ekeland, P. H. Rabinowitz, and E. Zehnder, eds., Reidel Publishers.

[14] Ekeland, I. and J. M. Lasry, On the number of periodic trajectories for a Hamiltonian flow on a convex energy surface, Ann. of Math. **112** (1980), 283-319.

[15] Berestycki, H., J. M. Lasry, G. Mancini, and B. Ruf, Existence of multiple periodic orbits on starshaped Hamiltonian surfaces, Comm. Pure Appl. Math. **38** (1985), 253-289.

[16] Hayashi, K., Periodic solutions on a convex energy surface of a Hamiltonian system, II, Tokyo J. Math. **9** (1986), 439-452.

[17] van Groesen, E. W. C., Existence of multiple normal mode trajectories of even classical Hamiltonian systems, J. Differential Equations **57** (1985), 70-89.

[18] Girardi, M., Multiple orbits for Hamiltonian systems on starshaped surfaces with symmetries, Ann. Inst. H. Poincaré. Anal. Non Linéaire **1** (1984), 285-294.

[19] Ekeland, I., Une théorie de Morse pour les systémes hamiltonien convexes, Ann. Inst. H. Poincaré. Anal. Non Linéaire **1** (1984), 19-78.

[20] Moser, J., Proof of a generalized form of a fixed point theorem due to G. D. Birkhoff, Springer Lec. Notes in Math. **597** Geometry and Topology (1977), 464-494.

[21] Ekeland, I. and H. Hofer, Sur les hypersurfaces convexes et leurs caracteristiques fermeés, C. R. Acad. Sci. Paris **304** (1987), 237-240.

[22] Ekeland, I. and L. Lassoued, Un flot hamiltonien a an moins deux trajectories fermeés sur toute surface d'energie convexe et bornée, C. R. Acad. Sci. Paris **301** (1985), 161-164.

[23] Szulkin, A., Morse theory and existence of periodic solutions of convex Hamiltonian systems, preprint.

[24] Rabinowitz, P. H., Minimax Methods in Critical Point Theory with Applications to Differential Equations, CBMS Regional Conf. Ser. in Math. **65** Amer. Math. Soc. (1986).

MATHEMATICS DEPARTMENT
UNIVERISTY OF WISCONSIN-MADISON
MADISON, WI 53706

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986266

Contemporary Mathematics
Volume **81**, 1988

HOMOCLINIC AND HETEROCLINIC PHENOMENA IN SOME HAMILTONIAN SYSTEMS

Carles Simó[1]

ABSTRACT. Two examples of 2 degrees of freedom Hamiltonian Systems are studied. The first one is the family of truncations of the periodic 3 equal particles Toda lattice. The second one is a harmonic oscillator plus the terms $-xy^2+\underline{a}y^4$, $\underline{a}$ being a parameter, which includes a system numerically studied by Barbanis. In both cases one shows how homoclinic and heteroclinic orbits give relevant information. For the first system the following items are studied: integrability, families of simple periodic orbits (using normal forms), splitting of the separatrices between hyperbolic orbits and periodic orbits ending on a homoclinic orbit to a saddle-center. For the second one, the characteristic curve of a family of symmetric triple periodic orbits is studied. It is shown that for values of the parameter $\underline{a}$ in a given range the characteristic curve spirals to a finite curve obtained from the invariant manifolds of Lyapunov orbits for some range of values of the energy. The behavior of the characteristic curve with respect to $\underline{a}$ is discussed.

1. THE FIRST EXAMPLE.

a) Statement of the problem. We consider the periodic 3 equal masses Toda lattice with Hamiltonian given by (24)

$$H = \frac{1}{2}\sum_{i=1}^{3} P_i^2 + \exp(Q_1-Q_2) + \exp(Q_2-Q_3) + \exp(Q_3-Q_1) .$$

After using the momentum first integral and some scaling the Hamiltonian is reduced to the form to be used through the paper

$$H = \frac{1}{2}(p_1^2+p_2^2) + \frac{1}{24}\left[\exp(-2\sqrt{3}q_1-2q_2) + \exp(2\sqrt{3}q_1-2q_2) + \exp(4q_2) - 3\right] .$$

This system is known to be integrable (15). Taking positive energy, $h>0$, and performing a Poincaré section of the flow through $q_1=0$, the behavior, in the q_2,p_2 variables, is shown in Fig.1. Given a point on the (q_2,p_2) plane it determines uniquely an orbit because $q_1=0$ and p_1 can be recovered from the energy integral $H=h$ (and $p_1\geq 0$, say).

We recall that a fixed point, P, of a conservative map of the plane, T, is said elliptic, hyperbolic or parabolic if the eigenvalues of DT(P) belong

---

1980 Mathematics Subject Classification (1985 Revision) 58F

[1]Supported by CAICYT Grant 3534/83C3 (Spain).

to $\{z\in\mathbb{C}, |z|=1\}\setminus\{\pm1\}$, $\mathbb{R}\setminus\{\pm1\}$ or $\{\pm1\}$, respectively. Fixed points of the Poincaré map are related to simple periodic orbits of the flow. Periodic points of period k of the Poincaré map are related to the so called k-ple periodic orbits of the flow.

The boundary of Fig.1 is a periodic orbit of H of parabolic type. There is also a full line of parabolic fixed points. The remaining of Fig.1 contains two elliptic fixed points. The space around them, until the boundary or the line of parabolic points, is foliated by invariant curves corresponding to tori in the energy level. A nicer picture is obtained by identification of the boundary periodic orbit to one point, giving a two dimensional sphere (see Fig.2).

As a useful fact we remark that H is invariant under rotation of angle $2\pi/3$.

Despite H is integrable the truncation of the power series expansion around the origin at order 3 gives the well known Hénon-Heiles system (16,4,19) which is known to be non integrable (26). Consider in general, the n^{th} order truncations

$$H_{(n)}=\frac{1}{2}(p_1^2+p_2^2)+\frac{1}{24}T_n[\exp(-2\sqrt{3}q_1-2q_2)+\exp(2\sqrt{3}q_1-2q_2)+\exp(4q_2)-3],$$

where T_n stands for the n^{th}-order Taylor operator, $n\geq2$. Recent interest in that problem has lead to papers by Contopoulos-Polymilis (7) and Yoshida (25). We note that $H_{(n)}$ is still invariant under $2\pi/3$ rotation.

Our objective is to understand the behavior for $n\geq3$ and small positive energy, as well as to study a simple family of periodic orbits when the energy increases.

b) <u>Integrability</u>. It follows immediately, from the Hamilton equations obtained from $H_{(n)}$, that for any $n\geq2$ and small positive h there is a periodic orbit $q_1=p_1=0$ (the boundary of the allowed region in the Poincaré section) projecting on a straight line on the (q_1,q_2) plane. To see what happens when the energy increases we need the

<u>Lemma 1.1</u>. Let $V_{(n)}(0,q_2)$ the potential energy in $H_{(n)}$ restricted to the line $q_1=0$. Then for n even $V_{(n)}(0,q_2)$ is a convex function with minimum at $q_2=0$. For n odd, $n\geq3$, $V_{(n)}(0,q_2)$ has exactly one minimum at $q_2=0$ and one maximum at $q_2 = q_{2,max} = -A_n/2$ with

$$A_n = v_0 n + \left(\frac{2}{v_0}-2\right)^{-1}\ln n + \left(\frac{1}{v_0}-1\right)^{-1}\ln\left(\sqrt{2\pi}(1+2v_0)\right) + o(1), \quad n\to\infty,$$

where v_0 is the first positive solution of $\exp(v)=2ev$ $(v_0\simeq 0.231961)$. Then $V_{(n)}(0,q_{2,max}) = \frac{1}{12}\exp(A_n)(1-v_0+O(1/n))$ for n odd. Furthermore the point $(0,q_{2,max},0,0)$ is a saddle-center of $H_{(n)}$ for n odd with eigenvalues $\pm i\,\exp(-q_{2,max})(1+o(1))$, $\pm\exp(-q_{2,max})((1-v_0)/(3v_0))^{1/2}(1+o(1))$, for $n\to\infty$.

<u>Proof</u>: It is immediately seen that for all $n\geq2$ $V_{(n)}(0,q_2)$ has a minimum at the origin. First we remark that $T_m(e^z)>0$ for all $z\geq0$ and all $m\geq0$. For negative z

as we have $T_m(e^z)=e^z-z^{m+1}e^w/(m+1)!$, $w \in (z,0)$, we have $T_m(e^z)>0$ for even m. For n even one should prove $\partial^2 V_{(n)}/\partial q_2^2|_{(0,q_2)}>0$. But this amounts to prove positiveness for $T_{n-2}(8\exp(-2q_2)+16\exp(4q_2))$ which follows from the remark.

For n odd the same reasoning shows $\partial^2 V_{(n)}/\partial q_2^2|_{(0,q_2)}>0$ for $q_2>0$. Let $z=-2q_2$. We wish to study the first derivative of $f(z)=T_n(2e^z+e^{-2z}-3)$ for $z>0$. Let $g(z)=\frac{1}{2}f'(z)=T_{n-1}(e^z-e^{-2z})=3z-3z^2/2+\ldots+(2^{n-2}+1)z^{n-2}/(n-2)!-(2^{n-1})z^{n-1}/(n-1)!$. Letting aside the case n=3, for which the uniqueness of the maximum is trivial, we check easily $g(z)<0$ for $z\geq n/2$ by comparing in g each couple of consecutive terms, the first of odd degree and the second of even degree. The same technique proves $g(2)>0$. Hence it remains to show the uniqueness of the zero of g in $(2,n/2)$ and to compute an asymptotic expression for its value.

To this end we introduce $v=z/n$ (it is enough to consider $0<v<1/2$) and express $T_{n-1}(e^z)$ as $e^{vn}-v^n n^n A(v,n)/n!$ and $T_{n-1}(e^{-2z})$ as $e^{-2vn}+v^n n^n 2^n B(v,n)/n!$, where

$$A(v,n)=1+\frac{vn}{n+1}+\frac{(vn)^2}{(n+1)(n+2)}+\ldots \quad \text{and} \quad B(v,n)=1-\frac{2vn}{n+1}+\frac{(2vn)^2}{(n+1)(n+2)}-\ldots .$$

We recall that $n!=(n/e)^n C(n)$ with $\sqrt{2\pi n}<C(n)<\sqrt{2\pi n}\exp(1/12n)$. The equation $g(z)=0$ is written as $e^{vn}(1-e^{-3vn})=(ve)^n(2^nB+A)/C$ or as $e^v=2veD(v,n)$, where $D(v,n)=((B+2^{-n}A)/(C(1-e^{-3vn})))^{1/n}$.

One easily obtains the bounds $e^{-3vn}<e^{-6}$, $1<A<2$ and $B_1<B<1$, where

$$B_1=\frac{1}{n+1}+\frac{3n^2}{(n+1)\ldots(n+3)}+\frac{5n^4}{(n+1)\ldots(n+5)},$$

using the fact that the sum of the first 6 terms in B decreases for $v\in(0,1/2)$.

We skip the cases n=5,7,9,11, for which a numerical computation or an adhoc proof can be used. For instance, for n=7, $g(z)=0$ can be written as

$$g_1(z)=\frac{3}{z}+\frac{z^3}{40}-\frac{z^4}{40}=\frac{3}{2}(1-\frac{z}{2})^2+\frac{1}{4}z^2(1-\frac{z}{2})^2=g_2(z).$$

One has $g_1(2)>0=g_2(2)$ and $g_1'(z)<0$, $g_2'(z)>0$ for $z>2$.

We claim that $D_1<D<D_2$ where D_2 can be taken trivially equal to 1 and D_1 can be taken equal to 0.775 if $n\geq 13$. To prove the lower bound we should check $E(n)/(n+1)>0.775^n\sqrt{n}F(n)$, where

$$E(n)=1+\frac{3n^2}{(n+2)(n+3)}+\frac{5n^4}{(n+2)\ldots(n+5)} \quad \text{and} \quad F(n)=\sqrt{2\pi}\exp(1/12n).$$

As $E(n)$ (resp. $F(n)$) increases (resp. decreases) with n, it is enough to see $Q(n)<E(13)/F(13)$ for all $n\geq 13$, where $Q(x)=0.775^x\sqrt{x}(1+x)$. But $Q'(x)=Q(x)\cdot(\ln 0.775+1/(2x)+1/(1+x))$ and the second factor is decreasing. Hence it remains only to check numerically $Q(13)<E(13)/F(13)$ and $Q'(13)/Q(13)<0$, which turns out to be true.

Therefore any solution of $e^v=2veD(v,n)$ should be in (v_2,v_1) where $\exp(v_j)$ coincides with $2v_jeD_j$, j=1,2. One has $v_1<0.331$, $v_2>0.231$.

To prove uniqueness we write $g(z)=0$ as $G(v,n)=2veD(v,n)e^{-v}=1$. The only thing to check is $G'(v,n)>0$ at the solution, where $'$ denotes $\partial/\partial v$. But

$$G'(v,n)\Big|_{G(v,n)=1}=(\ln D)'+\frac{1}{v}-1>\frac{1}{n}\frac{B'+2^{-n}A'}{B+2^{-n}A}-\frac{3e^{-3vn}}{1-e^{-3vn}}+\frac{1}{v_1}-1>$$

$$\frac{1}{n}\frac{B'+2^{-n}A'}{B+2^{-n}A}+2.021 \quad .$$

We know that B, A, A' are positive in (v_2,v_1). We claim that even in the larger interval $(0,1/3)$ the following is true: i) $B'<0$; ii) $B'>b=-2$: iii) $B>a=1/3$.

Then to have $G'(v,n)>0$ for $G(v,n)=1$ it is enough to have $b>-2.021\,n\,a$ for $n\geqslant 13$. Let us prove the claim.

Introducing $w=2v$ it follows $B'<0$ for $w\in(0,1/2)$ because B' is an alternate series with decreasing terms in absolute value (a.s.d.t.a.v.) and the first term is negative. Splitting B' in the first 6 terms and the remainder we have $B'/2=np(w)/((n+1)\dots(n+6))+\bar{B}$. $\bar{B}$ turns out to be an a.s.d.t.a.v. if $w\in(0,2/3)$ with negative first term. Furthermore $p(2/3)<0$ for all positive n and $p'(w)/n=\sum_{j=0}^{4}c_j(w)n^j$. For $w\in(0,2/3)$ one has $c_0=360$, $c_1\geqslant 102$, $c_2\geqslant 1011/20$, $c_3\geqslant -4/9$ and $c_4\geqslant 11/6$ showing $p'(w)>0$ for $n\geqslant 1$ and hence $B'<0$.

To prove ii) we split B' as

$$B' = 2\left(-\frac{n}{n+1}+\frac{2wn^2}{(n+1)(n+2)}-\frac{3w^2n^3}{(n+1)\dots(n+3)}\right)+\bar{\bar{B}} ,$$

$\bar{\bar{B}}$ being positive for $w\in(0,2/3)$. The minimum value of the first part is reached at 0 and therefore we can take $b=-2$.

Finally iii) follows from $B(v)>B(1/3)=\frac{1}{3}+\frac{2}{3}-\frac{2n}{3(n+1)}+$ a.s.d.t.a.v. with positive initial term and then $B(v)>1/3$. Uniqueness follows because $-2>-2.021\,n/3$ for $n\geqslant 3$.

To compute the asymptotic expression of $q_{2,max}$ we remark that for large n the term $2^{-n}A$ can be neglected in front of B in D, as well as e^{-3vn}. Furthermore $B=1/(1+2v)+O(1/n)$. Then $D=((1/(1+2v)+O(1/n))/\sqrt{2\pi n})^{1/n}$ tends to 1 when $n\to\infty$ giving for v the value v_2, equal to the value v_0 in the statement. One step of differential correction produces $z=vn=A_n$.

The value of the potential at $q_{2,max}$ is computed expressing again $T(e^z)$ as the exponential minus the remainder. This gives

$$\frac{1}{12}e^{vn}-\frac{1}{24}\,2^{n+1}\,\frac{v^{n+1}}{1+2v}\,\frac{n^{n+1}}{(n+1)!}\left(1+O\left(\frac{1}{n}\right)\right) ,$$

and using the equation which determines v we obtain $\frac{1}{12}e^{vn}(1-v+O(1/n))$.

The characteristic equations at $(0,q_{2,max},0,0)$ are given by $\lambda^2+V_{q_1,q_1}=0$, $\lambda^2+V_{q_2,q_2}=0$. If we introduce $r=T_{n-2}(\exp(-2q_{2,max}))$, $s=T_{n-2}(\exp(4q_{2,max}))$ one has $V_{q_1,q_1}=r$, $V_{q_2,q_2}=(r+2s)/3$. From $r=e^{vn}(1+o(1))$, $s=-e^{vn}(1+o(1))/(2v)$ the eigenvalues follow. ∎

Numerically computed values of $q_{2,max}$, h_{max}, a and b ($\pm a, \pm ib$ being the eigenvalues at the saddle-center) are given in Table 1.

Those rectilinear solutions are used by H.Yoshida (25) that has obtained a Theorem (derived from Ziglin's one (26)) to study $\frac{1}{2}(p_1^2+p_2^2)+\Sigma_{j=k}^{n} V_j(q_1,q_2)$, where V_j is a homogeneous polynomial of degree j (see also related work of Churchill and Rod (5)). Analyzing the behavior for $h \searrow 0$ and $h \nearrow \infty$ Yoshida gives criteria ensuring non integrability. In particular

Theorem 1.2.(Yoshida(25)).$H_{(n)}$ is non integrable unless n=2.

We remark that for n=4 the system obtained taking only into account V_2 and V_4 is integrable. The role of V_3 is essential. The behavior close to integrable for $h \searrow 0$ and $h \nearrow \infty$ of $H_{(4)}$ has been noticed in numerical simulations (7).

c) Normal forms and simple periodic orbits. To study $H_{(n)}$ for small positive h we use a normal form technique. The goal is here to obtain from $H_{(n)}$ a system which is integrable but displays many of the characteristic features of $H_{(n)}$ (except, of course, all the homoclinic and heteroclinic tangles and its consequences). As the quadratic part of $H_{(n)}$ for $n \geqslant 2$ is $\frac{1}{2}(q_1^2+p_1^2+q_2^2+p_2^2)$ we are in an 1 to 1 resonance. Hence the Gustavson normal form should be used (13,21). We adopt the notation from (12).

Let $x_j=(q_j-ip_j)/\sqrt{2}$, $y_j=(-iq_j+p_j)/\sqrt{2}$, j=1,2 new canonical variables. Then the normal form up to order 2m of an analytical 1 to 1 resonant Hamiltonian is

$$G^{(2m)} = \sum_{2<|l+k|<2m,\ |l|=|k|} g(l,k)X^l Y^k \quad , \quad k,l \in (\mathbb{N}\cup\{0\})^2,\ X^l = X_1^{l_1}X_2^{l_2},\ |l|=l_1+l_2,$$

where X_j, Y_j, j=1,2 denote the new variables. Of course the transformed Hamiltonian obtained from the initial one, H, is $G=G^{(2m)}+R^{(2m)}$, where the remainder contains terms of degree greater than 2m. $G^{(2m)}$ is integrable, $X_1Y_1+X_2Y_2$ being an additional first integral. We go back to real canonical variables Q_j, P_j, j=1,2 such that $X_j=(Q_j-iP_j)/\sqrt{2}$, $Y_j=(-iQ_j+P_j)/\sqrt{2}$ and obtain $G^{(2m)}(Q,P)$ whose coefficients are real.

We introduce new variables, denoted again by q,p supposing that no confusion with the original ones arises, by $Q_j=\varepsilon q_j$, $P_j=\varepsilon p_j$, j=1,2. Hence

$$G^{(2m)} = \varepsilon^2 G_2(q,p) + \varepsilon^4 G_4(q,p) + \ldots + \varepsilon^{2m} G_{2m}(q,p) \ ,$$

where G_{2k} is a homogeneous polynomial of degree 2k. Furthermore $R^{(2m)}=O(\varepsilon^{2m+1})$. Taking $\varepsilon=\sqrt{h}$ and scaling the time t to a new time $s=\varepsilon^{-2}t$ we can divide G by ε^2 and consider the level of energy equal to 1. Then $\varepsilon=0$ is exactly a harmonic oscillator.

Following Braun (3) we perform the canonical change $q_j=\sqrt{2r_j}\sin a_j$, $p_j=\sqrt{2r_j}\cos a_j$ and then, as a new change, either

(*) $\quad R_1 = r_1+r_2,\ R_2 = r_2,\ b_1 = a_1,\ b_2 = a_2 - a_1,$ or

(**) $\quad R_1 = r_1+r_2,\ R_2 = r_1,\ b_1 = a_2,\ b_2 = a_1 - a_2$.

The change (*) is suitable in the full energy level G=1 except in a neighborhood of the periodic orbit $q_1=p_1=0$. In this case (**) is the right change to study the vicinity of the periodic orbit.

With either (*) or (**) one obtains (3) that $G^{(2m)}$ is independent of b_1. Hence R_1 is constant and equal to $1+O(\varepsilon^2)$. It can be skipped from $G^{(2m)}$ which only depends on R_2 and b_2, i.e., it is a one dimensional Hamiltonian. Finally the change $u=\sqrt{2R_2}\sin b_2$, $v=\sqrt{2R_2}\cos b_2$ leads to a Hamiltonian in cartesian coordinates: $K^{(2m)}(u,v)=\varepsilon^2K_4+\varepsilon^4K_6+\ldots+\varepsilon^{2m-2}K_{2m}$. We remark that K_{2j} is no longer a homogeneous polynomial of degree 2j in u,v but the highest degree is 2j.

<u>Proposition 1.3</u>. Let m be such that $K^{(2m)}$ is of the form $u^2(\varepsilon^2P_1(v)+O(\varepsilon^3)) + O(u^3\varepsilon^2) + \varepsilon^{2m-2}P_2(v)$ with P_1,P_2 polynomials satisfying $P_1(0)\neq 0$, $P_2(0)=P_2'(0)=0$, $P_2''(0)\neq 0$. Then $K^{(2m)}$ has a non parabolic fixed point at the origin. Furthermore there exists ε_0 such that for all ε with $|\varepsilon|<\varepsilon_0$ the Poincaré map of the original system has a non parabolic fixed point close to the origin.

<u>Proof</u>: At the origin one has $K_u^{(2m)}=K_v^{(2m)}=0$, $K_{uu}^{(2m)}=\varepsilon^2P_1(0)+O(\varepsilon^3)$, $K_{uv}^{(2m)}=0$, $K_{vv}^{(2m)}=2\varepsilon^{2m-2}P_2''(0)$ proving the first part. The second one follows from the Implicit Function Theorem after a scaling $u=\varepsilon^{2m-4}\bar{u}$ is done. ■

Now we describe the results of carrying out all the computations described above for the problem under consideration up to n=15. As starting Hamiltonian we take $H_{(n)}$, and we apply 1.3. to obtain the next result.

<u>Theorem 1.4</u>. Let $3\leq n\leq 15$. Then the smallest value of m such that for $K_{(n)}^{(2m)}$ all the fixed points are non parabolic is given by $m=E[(n+2+\delta_{n,3})/2]$, where $E[\cdot]$ denotes the integer part and $\delta_{i,j}$ the Kronecker index. In all the cases there are 8 fixed points. They correspond to periodic orbits of the original system of period going to 2π when $h\searrow 0$. Two of them are elliptic and they already appear in $K_{(n)}^{(4)}$. For $2\leq j<m$, $K_{(n)}^{(2j)}$ has a line of parabolic fixed points which is broken, giving rise to 3 elliptic fixed points and 3 hyperbolic ones, for j=m. The pattern is given in Fig.3, where the change (*) has been used for odd n and (**) for even n. After compactification of the boundary periodic orbit to one point we obtain Fig.4 .

As an example of the final one dimensional Hamiltonian from which 1.4. shows up, we display $K_{(11)}^{(12)}$ in Table 2. The columns give the exponents of ε,u and v and the related coefficient, respectively.

We remark that for the Toda Hamiltonian the line of parabolic points should be present at all orders of normalization. Hence, unless n=3, the qualitative behavior of $H_{(n)}$ differs from Toda's one at the first opportunity.

We also remark that thanks to the symmetry it is clear that, generically, the number of fixed points should be equal to 2 plus a multiple of 6.

<u>Conjecture 1.5</u>. For all $n\geq 3$ the truncated Toda lattice has 8 simple periodic orbits, 5 of them elliptic and 3 hyperbolic, with period going to 2π when $h\searrow 0$.

Proposition 1.6. Any other periodic orbit of $H_{(n)}$ has a period which is $O(\varepsilon^{-2})$.
Proof: The speed of the flow on the (u,v) plane is $O(\varepsilon^2)$ and hence the time required to return to an initial point far from the fixed points is $O(\varepsilon^{-2})$. Also the elliptic fixed points of the Poincaré map have eigenvalues $1\pm iO(\varepsilon^s)$ with s at least equal to 2. Therefore the number of iterates to return to the initial point is also $O(\varepsilon^{-2})$. ∎

d) Analysis of the separatrices (heteroclinic connections). To study the behavior of the separatrices connecting the hyperbolic points for the one degree of freedom Hamiltonian system $K^{(2m)}_{(n)}$ obtained from the normal form we perform a new scaling of variables. The line of parabolic fixed points corresponding to the Toda lattice is broken in a more complicate pattern. The following result follows from the expressions of $K^{(2m)}_{(n)}$ using scaling.
Theorem 1.7. Let m as in 1.4. and $\overline{K}^{(2m)}_{(n)}$ the terms of lower degree in ε (which is equal to 2m-2) in $K^{(2m)}_{(n)}(\varepsilon^r x,y)$, where r=m-2. Then in a strip near the separatrices the Poincaré map of $H_{(n)}$ is $O(\varepsilon^{m+1})$ close to the $2\pi\varepsilon^m$ time map of $\overline{K}^{(2m)}_{(n)}$, at least for $3\leq n\leq 15$. Furthermore

$$\overline{K}^{(2m)}_{(n)} = a_n x^2(1-\frac{y^2}{2}) - \sum_{j\geq 1} b_{2j,n}(3y-2y^3)^{2j}$$

with $b_{2j,n}=0$ if $n<6j-2$. The values of a_n, $b_{2j,n}$ are given in Table 3.

The role played by $3y-2y^3$ is certainly related to the symmetry of the problem. In Fig.5 we display the separatrices for $\overline{K}^{(2m)}_{(n)}$ (using (*) for odd n, (**) for even n.
Corollary 1.8. Let $3\leq n\leq 15$. Then the separatrix of the origin (on $\overline{K}^{(2m)}_{(n)}=0$) is given by

$$3y-2y^3 = \sqrt{2}\left[\mathrm{Cosh}^2(6\sqrt{a_n b_{2,n}}\,u) + (2b_{4,n}/b_{2,n})\mathrm{Sinh}^2(6\sqrt{a_n b_{2,n}}\,u)\right]^{-1/2} ,$$

where u denotes here the time for $\overline{K}^{(2m)}_{(n)}$.
Proof: The origin being hyperbolic one has $b_{2,n}>0$. We skip the index n in a, b_2, b_4 and both indices in K. On K=0 one has

$$ax^2(1-\frac{y^2}{2}) = b_2(3y-2y^3)^2 + b_4(3y-2y^3)^4 \quad\text{and}\quad y'=-2ax(1-\frac{y^2}{2}),\ ('=\frac{d}{du}) .$$

Introducing $w=\sqrt{2}/(3y-2y^3)$ one obtains $ww'=6\sqrt{a(b_2w^2+2b_4)}\sqrt{w^2-1}$ after some manipulation. If $v=w^2$ one has $v'=12\sqrt{ab_2}\sqrt{(v-1)(v+2b_2/b_4)}$ with solution $v=\mathrm{Cosh}^2(6\sqrt{ab_2}u)+(2b_4/b_2)\mathrm{Sinh}^2(6\sqrt{ab_2}u)$, from which the result follows.
We wish to point out that in the general case (several b_{2j}) the differential equation for $z=w^{-2}$ is $z'=-12\sqrt{a}\,z\,\sqrt{1-z}\sqrt{\sum_{j\geq 1} b_{2j}(2z)^{j-1}}$. ∎

It follows easily from 1.7. that for $\overline{K}^{(2m)}_{(n)}$ the elliptic points are $(0,\pm\sqrt{1/2})$ with eigenvalues $\pm i6\sqrt{a_n(b_{2,n}+4b_{4,n})}$ (which implies eigenvalues equal to $\exp(\pm 2\pi i\varepsilon^m 6\sqrt{a_n(b_{2,n}+4b_{4,n})})(1+O(\varepsilon^{m+1}))$ for the Poincaré map). The hyperbolic ones are (0,0) and $(0,\pm\sqrt{3/2})$ with eigenvales $\pm 6\sqrt{a_n b_{2,n}}$ (and, therefore

$\exp(\pm 2\pi\varepsilon^m 6\sqrt{a_n b_{2,n}})(1+O(\varepsilon^{m+1}))$ for the Poincaré map).

When the remainder of the Poincaré map is introduced it appears a heteroclinic (and, by chain connection, also homoclinic) tangle as displayed in Fig6.

Concerning the splitting of the separatrices (see(11) and also (18)) we have, as a consequence of 1.8. the following result.

Proposition 1.9. For $3\le n\le 15$ and m depending on n as in 1.4., given $c>0$ the area between the loops of the heteroclinic tangle (or the distance between the left branches of the manifolds $W^u_{(0,\sqrt{3/2})}$ and $W^s_{(0,0)}$ at a given y, $0<y<\sqrt{3/2}$, or the angle d at some symmetric heteroclinic point as in Fig.7) is bounded by

$$N(c)\exp(-(|\mathrm{Im}(\mathrm{Pole})|-c)/(6\sqrt{a_n b_{2,n}}\varepsilon^m),$$

for $|\varepsilon|<\varepsilon_0(c)$, where $N(c)$ is a constant depending only on c and Pole is given by $\frac{1}{2}\ln((1+\sqrt{-2b_{4,n}/b_{2,n}})/(-1+\sqrt{-2b_{4,n}/b_{2,n}}))$.

Proof: According to Theorem A in (11) the distance is bounded by $N(c)\exp(-2\pi(f-c)/\ln g)$, where g is the eigenvalue greater than 1 of the Poincaré map and f is the minimum distance from the poles of the separatrix of $\overline{K}^{(2m)}_{(n)}$ to the real axis when the time is scaled in such a way that the eigenvalues, for the differential equation obtained from the Hamiltonian $\overline{K}^{(2m)}_{(n)}$, are ± 1. The only thing to do is to compute the poles, i.e., the zeros of v (see the proof of 1.8.). Let $q=-2b_4/b_2$. Introducing $z=6\sqrt{a_n b_{2,n}}u$ as suitable independent variable to achieve eigenvalues ± 1, we have $z=\frac{1}{2}\ln((\sqrt{q}+1)/(\sqrt{q}-1))$, the required pole. If $q\le 0$ then $z=\pm i\,\mathrm{arccot}(\sqrt{-q})$. For $q\in(0,1)$ (note that $q\ge 1$ is not allowed) one has $z = \pm i\,\frac{\pi}{2} + \frac{1}{2}\ln\frac{1+\sqrt{q}}{1-\sqrt{q}}$. ∎

We remark that for n=10,12,14 the poles of the separatrix have real part different from zero. Based on numerical computations carried out for a generalization of the standard map (23) one should expect for the angle d (see Fig.7) a behavior like

$$A\varepsilon^z\exp\left(-\frac{|\mathrm{Im}(\mathrm{Pole})|}{6\sqrt{a_n b_{2,n}}\varepsilon^m}\right)\left[\cos\left(\frac{\mathrm{Re}(\mathrm{Pole})}{6\sqrt{a_n b_{2,n}}\varepsilon^m}\right)+o(1)\right] \text{ for } \varepsilon \text{ going to zero and}$$

suitable A and z. This would imply that the angle d changes sign an infinity of times when $\varepsilon\searrow 0$ (for values of the type $\varepsilon_k=(C/(k+1/2))^{1/m}$ with k positive integer and C a suitable constant). Each time that d=0 the manifolds $W^u_{(0,-\sqrt{3/2})}$ and $W^s_{(0,\sqrt{3/2})}$ have, generically, a cubic tangency. Pairs of new heteroclinic points appear or dissappear, alternatively, in a fundamental domain for such values ε_k.

Conjecture 1.10. The behavior given by 1.7. is true for all $n\ge 3$. An extension of 1.8. and 1.9. is possible using elliptic functions for $n<28$ and hyperelliptic ones if $n\ge 28$.

e) The trace for the rectilinear periodic orbits. As stated in b) it is immediately seen that for every $n\geq 2$, $q_1=p_1=0$ is a (rectilinear) periodic orbit for $H_{(n)}$ whose differential equation is $\dot{q}_2=p_2$, $\dot{p}_2=-\partial V_{(n)}/\partial q_2(0,q_2)$. Furthermore $p_2^2/2+V_{(n)}(0,q_2)=h$ and the period depends only on h. Of course, for n odd the periodic orbits exist only for $h<h_{max}=V_{(n)}(0,q_{2,max})$, with $q_{2,max}$ defined in 1.1. For $h=h_{max}$ a homoclinic orbit to the saddle-center $(0,q_{2,max},0,0)$ is found. This is the natural termination of the family of periodic orbits for n odd. Hence the period of the periodic orbit goes to infinity, if n is odd, for $h\nearrow h_{max}$. Conversely, for n even the period goes to zero when $h\nearrow\infty$.

The normal variational equations are $(\Delta q_1)^{\cdot} = \Delta p_1$, $(\Delta p_1)^{\cdot} = -\partial^2 V_{(n)}/\partial q_1^2(0,q_2(t))\cdot\Delta q_1$. The character of those periodic orbits is obtained from the trace, Tr, of the monodromy matrix of the normal variational equations. Elliptic (resp. hyperbolic) orbits are found for $|Tr|<2$ (resp. $|Tr|>2$), while $|Tr|=2$ gives parabolic orbits. From the comments following 1.8. one has that for small h (and for, at least, $3\leq n\leq 15$) $Tr=2+144\pi^2 a_n b_{2,n} h^m+O(h^{m+1})$ for the hyperbolic orbits and $Tr=2-144\pi^2 a_n(b_{2,n}+4b_{4,n})h^m+O(h^{m+1})$ for the elliptic ones, where m depends on n as stated in 1.4. For small positive energy the rectiline periodic orbits of $H_{(n)}$ are hyperbolic for n even and elliptic for n odd.

We can ask about the behavior of those periodic orbits for n even and increasing h. It has been proved by Yoshida (25) that if n=2k the value of Tr, when $h\to\infty$, tends to the limit

$$\frac{4}{\sin^2\frac{\pi}{2k}}\cos^2\left[\frac{\pi}{2k}\sqrt{(k-1)^2+\frac{24k(2k-1)}{2+4^k}}\right]-2\ .$$

For n=4 this gives the value 2 as it should be, because if in $V_{(4)}$ we only keep the fourth degree terms (the dominant ones for $h\to\infty$) the system is integrable. For big values of k one has the approximation

$$\lim_{h\to\infty} Tr = 2-\frac{96k(2k-1)}{(2+4^k)(k-1)}\,\frac{\pi/2k}{\tan(\pi/2k)}\,(1+O(k\cdot 4^{-k}))\ .$$

The results of a numerical computation for h up to 10^{16} and n=4,6,... are shown in Fig.8. Table 4 presents extrema values of Tr and the related values of the energy. The following comments are in order:

i) Except for n=4 the pattern is quite similar: There are 4 values (2 for n=4) of h for which the orbit becomes parabolic, the curve Tr versus h crossing twice Tr=2 and twice Tr=-2.

ii) The extreme values of Tr, Tr_{max} and Tr_{min}, show, roughly, a linear dependence with respect to n, with slopes 0.217 and 0.277, respectively.

iii) The values of the logarithms of the energy at the extrema, $\ln h_{max}$ and $\ln h_{min}$, also show a linear dependence in n as dominant term in its behavior.

iv) The quotient h_{min}/h_{max} is slightly decreasing and it seems to converge to some finite number close to 8.

v) Let $q_{2,max}$, $q_{2,min}<0$ and $p_2=0$ be the values of the initial conditions for those h_{max}, h_{min}. Then $q_{2,max}$ and $q_{2,min}$ are close (they differ by less than 4% for n=100) and also close to the value given by the expression for $q_{2,max}$ in 1.1. for the odd case.

Now we pass to the case n odd. Fig.9 shows some examples of Tr vs h from h=0 to $h=h_{max}$. The successive extrema of Tr tend to some values $\pm g$. The observed behavior is explained by the next result.

Theorem 1.11. Consider a family of periodic orbits of a 2 degrees of freedom Hamiltonian ending on a homoclinic orbit to a saddle-center lying on the level h_{max} as the ones in the example. Let $\pm a$, $\pm ib$ the eigenvalues at the saddle-center. Let $\{h_k\}$ a monotone sequence of values of h for which Tr=2 (or any other value in (-g,g)) of the same type, i.e., with a given sign of dTr/dh. Then $\lim_{k\to\infty} (h_{max}-h_k)/(h_{max}-h_{k+1}) = \exp(2\pi a/b)$. Furthermore let $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ be the differential of the Poincaré map from S_1 to S_2 (see Fig.10) restricted to the variables normal to the orbit. Then the value of g is given by
$$g = \lim_{q\to 0} (A^2+B^2+C^2+D^2+2)^{1/2}.$$

Proof: We can suppose the saddle-center located at the origin. Using a result of Moser (20) there is an analytical change of coordinates giving as transformed Hamiltonian $H=axy+b(u^2+v^2)/2+\ldots$ where (x,y) and (u,v) are pairs of canonical conjugate variables. Through the proof we shall skip the non dominant terms because we are only interested in the linear behavior. Let S_1 (S_2) be given by y=q (x=q). The map going from S_1 to S_2 has linear part $(x,u,v)^T \to (x,M(u,v)^T)^T$ where $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ preserves area. Furthermore, going from S_2 to S_1, the time needed is $\ln(q/x)(1+O(x))/a$ or, skipping additive constants $-\ln h(1+O(h))/a$. Hence the passage from S_2 to S_1 in the (u,v) plane is a rotation R_d of angle $d=-b\ln h(1+O(h))/a$. Then $Tr(R_dM)=(A+D)\cos d+(B-C)\sin d+o(1)$. The scaling is found because an increment of d in 2π requires to divide h by $\exp(2\pi a/b)$. The maximum of Tr is given by $Tr_m^2=(A+D)^2+(B-C)^2=A^2+B^2+C^2+D^2+2$. Changing q to q_1 produces essentially the conjugation of M by a rotation (of angle $\frac{b}{a}\ln(q/q_1)$) and hence $A_1^2+B_1^2+C_1^2+D_1^2=A^2+B^2+C^2+D^2+o(1)$, showing the existence of the limit. The value g can be called the "Trace of the homoclinic orbit". We remark that g=2 requires M to be exactly a rotation, a highly exceptional case.

In a more general case, when the periodic orbit is not supposed to be on u=v=0 as it happens in the case studied, the linear part of the map from S_1 to S_2 is $(x,u,v)^T \to (x,x(E,F)^T+M(u,v)^T)^T$ with suitable constants E,F. The persistence of x as first component of the image is due to conservation of the energy. The existence of the family of periodic orbits requires $MR_d(u,v)^T+x(E,F)^T=(u,v)^T$, which needs a careful analysis of higher order terms when $Tr(MR_d)$ is close to 2. Outside a neighborhood of Tr=2 the same scaling phenomenon is obtained. ∎

In our example the scaling factor $\exp(2\pi a/b)$ depends on n. However, using 1.1. one has $\lim_{n\to\infty} \exp(2\pi a/b) = \exp(2\pi\sqrt{(1-v_0)/(3v_0)})$, where v_0 is also given in 1.1. With the numerical value given for v_0 the scaling factor goes to 735.76 when $n\to\infty$.

The result given in 1.11. has been already stated in (19). A proof of the existence of infinite bifurcations and the geometric accumulation of the related energies is also given in (14) for the Contopoulos potential.

2. THE SECOND EXAMPLE.

a) Statement of the problem. We consider a Hamiltonian introduced by Barbanis (1) (a modification of a classical Hamiltonian of Contopoulos and Moutsoulas (6)) given by $H=\frac{1}{2}(p_x^2+p_y^2+x^2+y^2)-xy^2+\frac{1}{2}y^4$. We summarize some results concerning this Hamiltonian:

i) All the zero velocity curves (z.v.c.) given by $p_x=p_y=0$ are closed. They only exist for positive energy.

ii) There is a symmetry with respect to the x axis.

iii) There is only one fixed point, located at the origin, with double eigenvalues $\pm i$.

iv) There is a family of symmetric periodic orbits with projection on (x,y) as the one given in Fig.11. As those orbits cut $y=0$ three times with $\dot{y}>0$ (or with $\dot{y}<0$) we call them triple periodic orbits. Each one of the symmetric triple periodic orbits cuts $y=0$ in a point as A with $\dot{x}=0$. Then we recover the orbit from the values of the energy, h, and the x coordinate of the point A. The curve displaying x versus h for this family of periodic orbits is called, as usual, characteristic curve. Fig.12 shows, qualitatively, the characteristic curve obtained by Barbanis.

Our objective will be to understand the origin of this curve. We remark that Devaney (8) and also Henrard (17) have found infinite spiral characteristic curves going to one point which represents a homoclinic orbit to a complex-saddle fixed point. However in our system there is not such point and, in the present example, the spiral is "finite". This means that it spirals inwards for some 8 revolutions and then it spirals outwards. Furthermore some bubbles appear in the inner part (the so called characteristic curves of irregular families of periodic orbits (2), which seem to be not connected to the main family of triple periodic orbits).

What we shall present here is the numerical evidence (in pictorial form) that similar characteristic curves of a closely related family of Hamiltonians can be explained using the invariant manifolds of a family of Lyapunov periodic orbits. The result of Barbanis appears to be the remnant of the behavior for close systems, for which an infinite spiral appears as characteristic curve.

b) <u>A family of Hamiltonians</u>. We consider instead the Hamiltonian family $H=\frac{1}{2}(p_x^2+p_y^2+x^2+y^2)-xy^2+\underline{a}y^4$, $\underline{a}<1/2$. They have additional fixed points $L_\pm$ at $p_x=p_y=0$, $x=(2-4\underline{a})^{-1}$, $y=\pm(2-4\underline{a})^{-1/2}$. They are of saddle-center type and go to infinity when $\underline{a}\nearrow 1/2$. Let h_{crit} be the value of H at those points. For $H=h>h_{crit}$ the component of the Hill region containing the origin opens and reaches infinity. This region is bounded by the z.v.c. (see also Fig.15).

There is a critical value $\underline{a}=\underline{a}^*\simeq 0.4918863722$ such that there exist a double heteroclinic connection (see Fig.13). For this value of $\underline{a}$ the bounded branches of the stable and unstable manifolds of the upper saddle-center L_+ (which have the same projection on the (x,y) plane due to the symmetry) coincide with the unstable and stable ones, respectively, of the lower saddle-center L_-.

J.Font and M.Grau (9) looked for the characteristic curve of the family of symmetric triple periodic orbits for $\underline{a}=\underline{a}^*$. The result, displayed in Fig.14, shows that the characteristic curve consists of two spirals going to a (big) finite curve. Hence now we are faced to a new problem: Try to understand what happens for $\underline{a}=\underline{a}^*$.

c) <u>The Lyapunov family of periodic orbits and the related invariant manifolds</u>. For $\underline{a}$ 1/2 but close to 1/2 we look at h slightly greater than h_{crit}. There appear Lyapunov periodic orbits, P.O.±, (see, for instance, (20)) close to $L_\pm$ which project on the (x,y) plane on a curve with end points on the z.v.c. The orbits of this family, parameterized by the energy, are, at least locally, of hyperbolic type. They have stable and unstable manifolds $W^{s,u}_{P.O.\pm}$ with a strong symmetry (see Fig.15). The manifolds are cylinders and the projections on (x,y), of both the stable and the unstable manifolds of the same periodic orbit, coincide. The left branches of $W^{s,u}_{P.O.\pm}$ intersect y=0 (with a suitable sign of $\dot{y}$) in curves diffeomorphic to S^1 (see Fig.16). Let us call them $S^{s,u}_{P.O.\pm}$.

Take a segment U ending on M in the line $\dot{x}=0$ of the Poincaré section y=0 (see Fig.16). Following the flow downwards close to $W^s_{P.O.-}$ and returning to y=0 close to $S^u_{P.O.-}$ we obtain, as image of the segment, a curve, TU, spiraling around $S^u_{P.O.-}$. This is similar to the behavior in the Sitnikov problem (see (22)) and related problems.The points in $U\cap TU$ cut twice y=0 with vertical velocity, giving rise to symmetric triple periodic orbits (see Fig.17 a). Fig. 17 b) displays the same behavior at a suitable section, Z, of P.O.- (for instance through some constant value of y). Let us call $\bar{S}^{s,u}_{P.O.-}$ the local invariant manifolds of P.O.- on Z. The image of $S^s_{P.O.-}$ under forward flow is a fundamental domain on $\bar{S}^s_{P.O.-}$ which contains a point $\bar{M}$ image of M and an arc of curve $\bar{U}$, image of U, with one end point in $\bar{M}$. The image of $\bar{U}$ under the (local) Poincaré map $\bar{T}$ associated to Z produces an infinity of arcs $\bar{U}_n$, n=1,2,..., which accumulate on $\bar{S}^u_{P.O.-}$. Take a fundamental domain of $\bar{S}^u_{P.O.-}$ and two transversal arcs, A1, A2, the one being the image of the other under $\overline{T}$. The infinite arcs of the $\overline{U}_n$ between A1 and A2 are mapped by the forward flow on TU. If A1 is taken such that it is mapped by the forward flow on U, then each one of the points $Q_n = A1 \cap \overline{U}_n$, $n \geqslant n_1$, for some n_1, represents a symmetric triple periodic orbit on that level of energy.

A similar thing happens with the segment ending on N in the line $\dot{x}=0$ of the Poincaré section through y=0. Furthermore the spiral TU cuts the segment ending on N giving rise to new periodic orbits that we shall not consider here. The related characteristic curves for $\underline{a}=1/2$ can be seen in (1).

d) Evolution with h. For $\underline{a}=\underline{a}^*$ increasing h from h_{crit} on, the intersections of $W^u_{P.0.+}(h)$ with y=0 give a set of curves $S^u_{P.0.+}(h)$ diffeomorphic to S^1. There is a range (h_0,h_f), $h_0=h_{crit}$, such that, for every one of those values of h, the related $S^u_{P.0.+}(h)$ cuts $\dot{x}=0$ in two points giving a behavior like the one described in c). For $h=h_j$ let them be M_j,N_j (Fig.18). Putting all of them together on the (h,x) plane (Fig.19) we obtain an infinity of leaves of the characteristic curve of the symmetric triple periodic orbits (continuity of the leaves is ensured by transversality). An infinity of leaves (essentially vertical) is also obtained to the left of h_0 (see (10) for an analysis showing this fact).

For h greater than h_f, but close to it, Fig.20 shows the related behavior. Fig.20 a) and b) are the ones equivalent to Fig.17 a) and b), respectively. The infinite spiral obtained for $h \in (h_0,h_f)$ is now converted to a "finite" one, giving rise to a finite number of periodic orbits for any $h>h_f$, as it happens for $h<h_0$.

The standing problem is to see how the upper and lower leaves connect to produce the picture given in Fig.14.

e) Evolution with the parameter. Now we let $\underline{a}$ change. A decrease (increase) of $\underline{a}$ simply pushes up (down) the set of intersections $S^u_{P.0.+}(h)$, but for values of $\underline{a}$ close to $\underline{a}^*$ still some values $h_0(\underline{a})$, $h_f(\underline{a})$ are found. For $h_0(\underline{a})$ and $\underline{a}<\underline{a}^*$ (resp. $\underline{a}>\underline{a}^*$) it is found that $S^u_{P.0.+}(h_0(\underline{a}))$ has a tangency with $\dot{x}=0$ leaving $S^u_{P.0.+}(h_0(\underline{a}))$ on the $\dot{x}\geqslant 0$ region (resp. on the $\dot{x}\leqslant 0$ region). For $h_f(\underline{a})$ it is found that $S^u_{P.0.+}(h_f(\underline{a}))$ has a tangency with $\dot{x}=0$ leaving $S^u_{P.0.+}(h_f(\underline{a}))$ on the $\dot{x}\leqslant 0$ region. For $h \in (h_0(\underline{a}),h_f(\underline{a}))$ the curve $S^u_{P.0.+}(h)$ has two transversal intersections with $\dot{x}=0$.

In any case the characteristic curves of the symmetric triple periodic orbits for $\underline{a}$ close to $\underline{a}^*$ look similar to the one found for $\underline{a}=\underline{a}^*$.

Further increase of $\underline{a}$ let us reach a value $\underline{a}=\underline{a}^{**}\simeq 0.494644$ for which $h_0(\underline{a}^{**})=h_f(\underline{a}^{**})$. For $\underline{a}$ greater than $\underline{a}^{**}$ none of the curves $S^u_{P.0.+}(h)$ cuts the $\dot{x}=0$ axis and the finite curve around which the spiral characteristic curves accumulate has dissappeared. For a=a** it has shrinked to one point which is related to the tangency point of the x=0 axis with $S^u_{P.O.+}(h_o(a^{**}))$. This point represents a double heteroclinic connection between the Lyapunov orbits P.O.+ and P.O.-.

For a>a** a behavior similar to the one obtained by Barbanis (1) (see Fig. 12) is found. This looks as the remnant of the previous infinite spirals. The spiral becomes "finite", the outer arcs being only slightly distorted, and some of the innest ones, too weak to resist perturbation , become a bubble or diss-appear. Furthermore the irregular families of periodic orbits can be seen as connected with the regular ones through variation of the parameter a.

The standing problem is to obtain quantitative information on this phenomenon, specially for a close to a**.

ACKNOWLEDGEMENTS

The author is deeply grateful to H.Yoshida for some comments concerning the first example, and to J.Font and M.Grau for their help in the numerical computations involved in the second one.

n	$-q_{2,max}$	h_{max}	b	a
3	1.000000	1.666667 E-1	1.732051 E 0	1.000000 E 0
5	1.136568	3.586468 E-1	2.795409 E 0	1.819461 E 0
7	1.328456	6.490098 E-1	3.673161 E 0	2.651046 E 0
9	1.542562	1.113666 E 0	4.644058 E 0	3.610452 E 0
11	1.766625	1.861677 E 0	5.840712 E 0	4.782442 E 0
21	2.927121	2.107867 E 1	1.867376 E 1	1.713708 E 1
31	4.098879	2.242300 E 2	6.027270 E 1	5.759759 E 1
41	5.270401	2.354747 E 3	1.944940 E 2	1.898220 E 2
51	6.440702	2.458134 E 4	6.268469 E 2	6.198345 E 2
61	7.609835	2.555974 E 5	2.017945 E 3	2.013291 E 3
71	8.777980	2.650049 E 6	6.489751 E 3	6.517192 E 3
81	9.945305	2.741508 E 7	2.085409 E 4	2.104701 E 4
91	11.111950	2.831158 E 8	6.696664 E 4	6.785324 E 4
101	12.278022	2.919585 E 9	2.149202 E 5	2.184640 E 5
121	14.608777	3.094453 E11	2.210606 E 6	2.258074 E 6
141	16.938076	3.268644 E13	2.270454 E 7	2.327504 E 7
161	19.266253	3.443824 E15	2.329310 E 8	2.394360 E 8
181	21.593541	3.621175 E17	2.387568 E 9	2.459541 E 9
201	23.920108	3.801591 E19	2.445517 E10	2.523648 E10

Table 1

2	2	0	0.666666667 E0
	4	0	-0.333333333 E0
	2	2	-0.333333333 E0
4	2	0	-0.155555556 E1
	4	0	0.777777778 E0
	2	2	0.777777778 E0
6	2	0	0.372839506 E1
	4	0	0.711934156 E0
	2	2	-0.186419753 E1
	6	0	-0.257613169 E1
	4	2	-0.257613169 E1
	8	0	0.644032922 E0
	6	2	0.128806584 E1
	4	4	0.644032922 E0
8	2	0	-0.362277092 E1
	4	0	-0.258278464 E2
	2	2	-0.233676269 E1
	6	0	0.283305898 E2
	4	2	0.359355281 E2
	2	4	0.760493827 E1
	8	0	-0.771639232 E1
	6	2	-0.200418381 E2
	4	4	-0.169344993 E2
	2	6	-0.460905350 E1
	10	0	0.230452675 E0
	8	2	0.161316872 E1
	6	4	0.345679012 E1
	4	6	0.299588477 E1
	2	8	0.921810700 E0

10	2	0	-0.310951966 E 2
	0	2	-0.600000000 E-1
	4	0	0.185113700 E 3
	2	2	0.633466107 E 2
	0	4	0.750000000 E-1
	6	0	-0.134839285 E 3
	4	2	-0.265012460 E 3
	2	4	-0.875115226 E 2
	0	6	-0.133333333 E-1
	8	0	-0.123468393 E 2
	6	2	0.923254801 E 2
	4	4	0.157665666 E 3
	2	6	0.529800137 E 2
	0	8	-0.133333333 E-1
	10	0	0.293559236 E 2
	8	2	0.454617513 E 2
	6	4	-0.773812300 E 1
	4	6	-0.344318793 E 2
	2	8	-0.105820027 E 2
	0	10	0.592592593 E-2
	12	0	-0.533443492 E 1
	10	2	-0.160038603 E 2
	8	4	-0.160064529 E 2
	6	6	-0.534153368 E 1
	4	8	-0.796296296 E-2
	2	10	-0.444444444 E-2
	0	12	-0.987654321 E-3

Table 2

n	3	4	5	6	7	8	9	10	11	12	13	14	15
a_n	$\frac{7}{6}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$
$b_{2,n}$	$\frac{7}{18}$	$\frac{1}{3}$	$\frac{1}{18}$	$\frac{4}{9}$	$\frac{1}{18}$	$\frac{73}{270}$	$\frac{1}{40}$	$\frac{163}{1620}$	$\frac{1}{150}$	$\frac{1447}{56700}$	$\frac{1}{840}$	$\frac{319}{68040}$	$\frac{3}{19600}$
$b_{4,n}$	0	0	0	0	0	0	0	$\frac{-1}{648}$	$\frac{1}{16200}$	$\frac{-29}{28350}$	$\frac{1}{28350}$	$\frac{-31}{97200}$	$\frac{1}{105840}$

Table 3

n	h_{max}	Tr_{max}	h_{min}	Tr_{min}
10	0.203 E 1	3.33	0.567 E 2	-6.68
20	0.256 E 2	5.40	0.305 E 3	-9.00
30	0.284 E 3	7.54	0.283 E 4	-11.66
40	0.304 E 4	9.69	0.280 E 5	-14.39
50	0.321 E 5	11.86	0.284 E 6	-17.14
60	0.336 E 6	14.02	0.290 E 7	-19.91
70	0.351 E 7	16.19	0.298 E 8	-22.68
80	0.365 E 8	18.36	0.305 E 9	-25.45
90	0.378 E 9	20.53	0.313 E10	-28.22
100	0.390 E10	22.70	0.323 E11	-30.99

Table 4

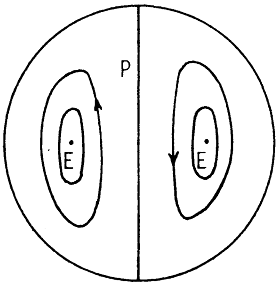


Figure 1

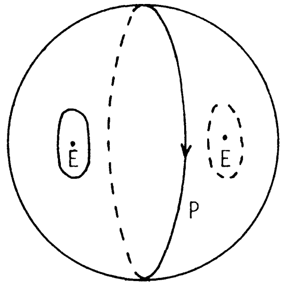


Figure 2

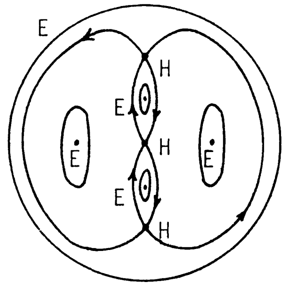


Figure 3

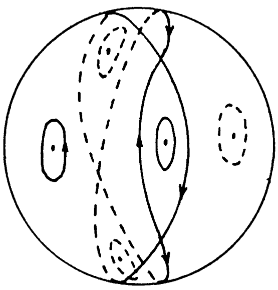


Figure 4

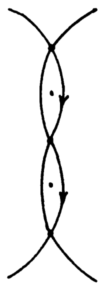


Figure 5

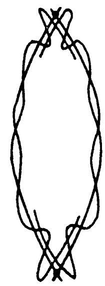


Figure 6

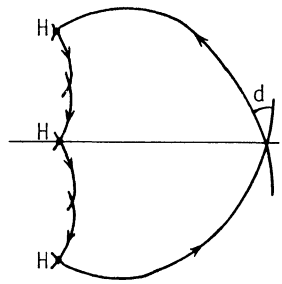


Figure 7

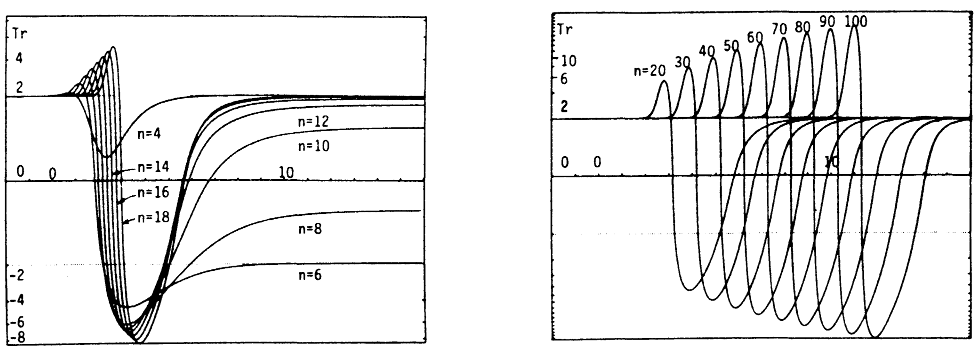


Figure 8. As abscisa it is taken $\log_{10}(24h)$.

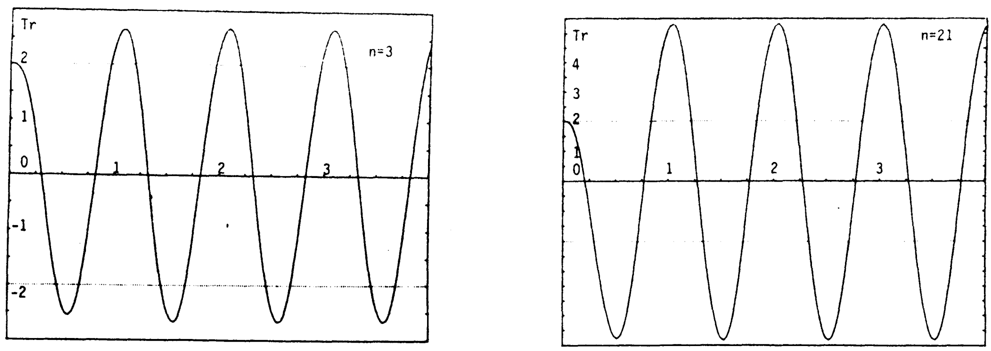


Figure 9. As abscisa it is taken $-\frac{b}{2\,a}\ln((h_{max} - h)/h_{max})$.

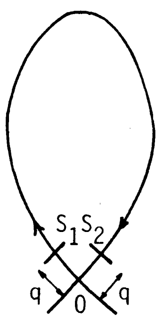


Figure 10

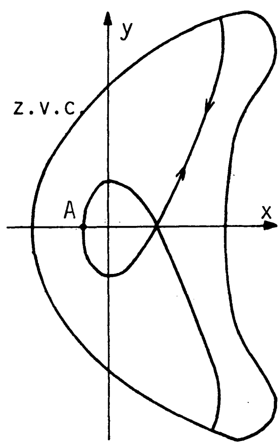


Figure 11

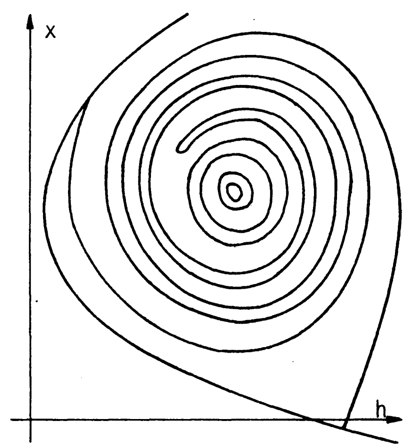


Figure 12

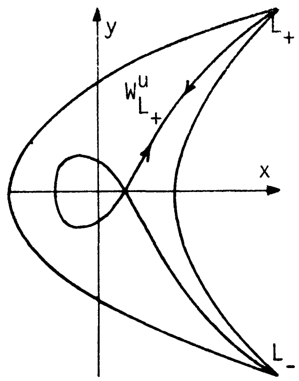


Figure 13

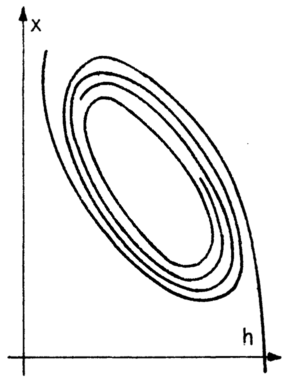


Figure 14

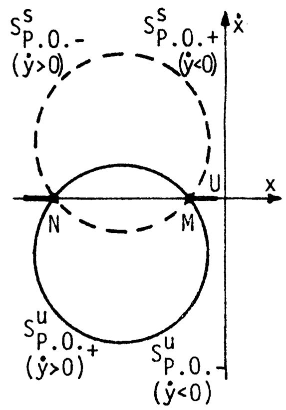


Figure 16

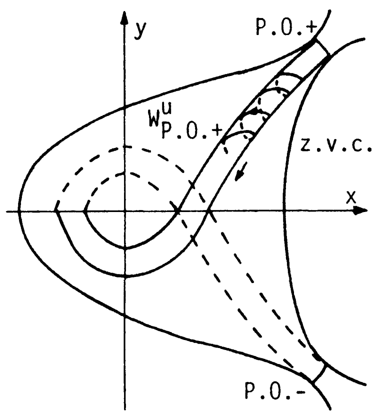


Figure 15

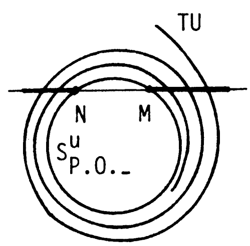


Figure 17 a)

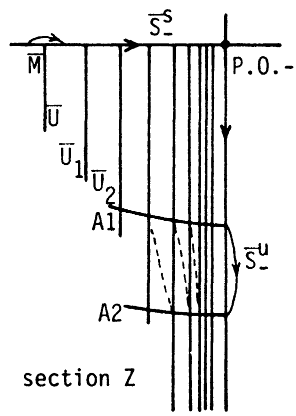


Figure 17 b)

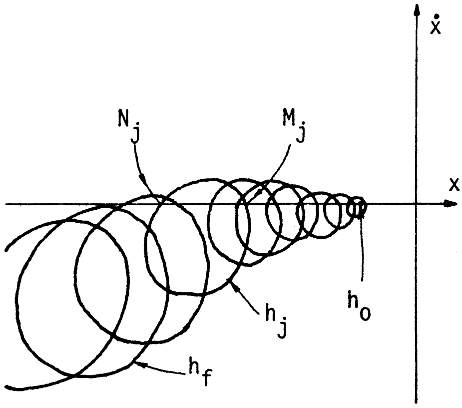


Figure 18

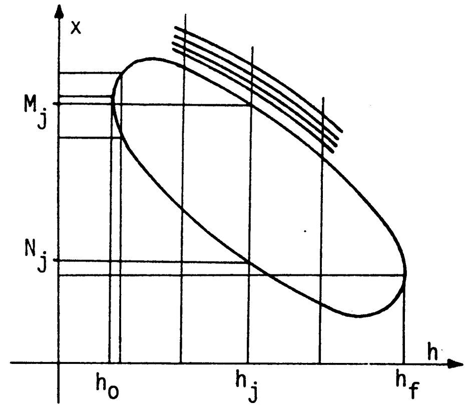


Figure 19

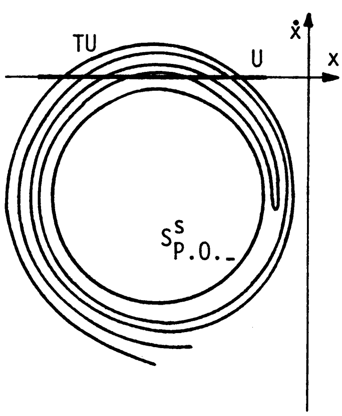


Figure 20 a)

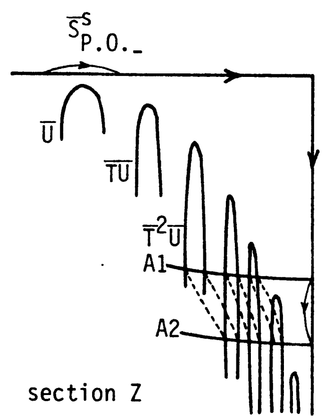


Figure 20 b)

BIBLIOGRAPHY

(1) Barbanis,B., Celes. Mech. 36 (1985) 257.

(2) Barbanis,B., Celes. Mech. 39 (1986) 345.

(3) Braun,M., J. Diff. Eq. 13 (1973) 300.

(4) Churchill,R.C., Pecelli,G. and Rod,D.L., in Lect. Not. in Physics 93 (1979) 76.

(5) Churchill,R.C. and Rod,D.L., Geometrical Aspects of Ziglin's Non-Integrability Theorem for Complex Hamiltonian Systems, preprint, 1987.

(6) Contopoulos,G. and Moutsoulas,M., Astron. J. 70 (1965) 817.

(7) Contopoulos,G. and Polymilis,C., Physica 24D (1987) 328.

(8) Devaney,R., J. Diff. Eq. 21 (1976) 431.

(9) Font,J. and Grau,M., Non-local spiraling characteristic curves, to appear in Proceed. NATO ASI on Long-Term Behavior of Natural and Artificial n-Body Systems, Reidel, 1988.

(10) Font,J. and Simó,C., Families of periodic orbits which come close to a fixed point in 2 degrees of freedom Hamiltonian systems, to appear in Actas VIII Congreso Ec. Dif. y Aplic., Univ. Santander, 1985.

(11) Fontich,E. and Simó,C., Splitting of Separatrices II: The Analytical Case, preprint, 1987.

(12) Giorgilli,A., Delshams,A., Fontich,E., Galgani,L. and Simó,C., Effective stability for a Hamiltonian system near an elliptic point, with an application to the restricted three body problem, preprint, 1987.

(13) Gustavson,F., Astron. J. 71 (1966) 670.

(14) Heggie,D.C., Celes. Mech. 29 (1983) 207.

(15) Hénon,M., Phys. Rev. B9 (1974) 1921.

(16) Hénon,M. and Heiles,C., Astron. J. 69 (1964) 73.

(17) Henrard,J., Celes. Mech. 7 (1973) 449.

(18) Holmes,P., Marsden,J. and Scheurle,J., Exponentially Small Splitting of Separatrices, preprint, 1987.

(19) Llibre,J. and Simó,C., in Actas III Congreso Ec. Dif. y Aplic., Univ. Santiago, 1980, 183.

(20) Moser,J.K., Comm. Pure Appl. Math. XI (1958) 257.

(21) Moser,J.K., Lectures on Hamiltonian Systems, Mem. Amer. Math. Soc. 81 (1968).

(22) Moser,J.K., Stable and Random Motions in Dynamical Systems, Princeton Univ. Press, 1973.

(23) Simó,C. and Fontich,E., in Singularités, feuilletages et mécanique hamiltonienne, Hermann 1985, 41.

(24) Toda,M., J. Phys. Soc. Jap. 22 (1967) 431 ; 23 (1967) 501.

(25) Yoshida,H., Non-integrability of the truncated Toda lattice Hamiltonian at any order, preprint, 1987.

(26) Ziglin,S.L. Funct. Anal. Appl. 16 (1982) 181 ; 17 (1983) 6.

DEPARTAMENT DE MATEMATICA APLICADA I ANALISI
UNIVERSITAT DE BARCELONA
BARCELONA 08007, SPAIN

http://dx.doi.org/10.1090/conm/081/986267

Contemporary Mathematics
Volume **81**, 1988

Exponentially Small Splittings of Separatrices with applications to KAM Theory and Degenerate Bifurcations

Philip Holmes, Jerrold Marsden, and Jurgen Scheurle

Abstract

Both upper and lower estimates are established for the separatrix splitting of rapidly forced systems with a homoclinic orbit. The general theory is applied to the equation

$$\ddot{\varphi} + \sin\varphi = \delta \sin\left(\frac{t}{\varepsilon}\right)$$

for illustration. There are two types of results. First, fix $\eta > 0$ *and let* $0 < \varepsilon \leq 1$ *and* $0 \leq \delta \leq \delta_0$ *where* δ_0 *is sufficiently small. If the separatrices split, they do so by an amount that is no more than*

$$C\,\delta \exp\left(-\frac{1}{\varepsilon}\left(\frac{\pi}{2} - \eta\right)\right)$$

where $C = C(\delta_0)$ *is a constant depending on* δ_0 *but is uniform in* ε *and* δ. *Second, if we replace* δ *by* $\varepsilon^p\delta$, $p \geq 8$, *then we have the sharper estimate*

$$C_2\,\varepsilon^p\,\delta\, e^{-\pi/2\varepsilon} \leq \text{ splitting distance } \leq C_1\varepsilon^p\,\delta\, e^{-\pi/2\varepsilon}$$

for positive constants C_1 *and* C_2 *depending on* δ_0 *alone. In particular, in this second case, the Melnikov criterion correctly predicts exponentially small splitting and transversal intersection of the separatrices. After developing this theory we discuss some of its applications, concentrating on a 2:1 resonance that occurs in a KAM (Kolmogorov, Arnold, and Moser) situation and in the forced saddle node bifurcation described by*

$$\ddot{x} + \mu x + x^2 + x^3 = \delta f(t)\ .$$

Introduction

In Poincaré's celebrated memoir [1890] on the 3-body problem, he introduced the mechanism of transversal intersection of separatrices which obstructs the integrability of the equations and the attendant convergence of series expansions for the solutions. This idea has been developed by Birkhoff and Smale using the horseshoe construction to describe the resulting chaotic dynamics. However, in the region of phase space studied by Poincaré, it has never been proved (except in some generic sense that is not easy to interpret in specific cases) that the equations really are nonintegrable. In fact Poincaré himself traced the difficulty to the presence of terms in the separatrix splitting which are exponentially small. A crucial component of the measure of the splitting is given by the following formula of Poincaré [1890, page 223]:

$$J = \frac{-8\pi i}{\exp\left(\frac{\pi}{\sqrt{2\mu}}\right) + \exp\left(-\frac{\pi}{\sqrt{2\mu}}\right)}$$

which is exponentially small (or beyond all orders) in μ. Poincaré was well aware of the difficulties that this exponentially small behavior causes; on page 224 of his article, he comments that "En d'autres termes, si on regarde μ comme un infiniment petit du premier ordre, la distance BB', sans être nulle, est un infiniment petit d'ordre infini. C'est ainsi que la fonction $e^{-1/\mu}$ est un infiniment petit d'ordre infini sans être nulle .... Dans l'example particulier que nous avons traité plus haut, la distance BB' est du même ordre de grandeur que l'integral J, c'est à dire que $\exp(-\pi/\sqrt{2\mu})$."

In this paper we overcome some of the essential difficulties that are encountered in this type of problem, in KAM theory, and in chaotic motions occurring in the unfoldings of degenerate singularities. Based on numerical evidence and formal calculations, it is known that one **should** get exponentially fine splittings and exponentially long escape times for problems of this type. Some rigorous but rough upper bounds for this phenomena have been given by Nekhoroshev [1971,77] and Neishtadt [1984]; see also the discussion in Arnold [1978], p.395ff and 407, Chirikov [1979] and Simo and Fontich [1985]. The analyticity argument of Cushman [1978] and Kozlov [1984] (and reference therein) uses the Poincaré-Melnikov method to prove that the separatrices do split for most parameter values. However, it is not easy to prove from these arguments that splittings really do occur for specific parameter values and what the sharp upper and lower estimates for the splitting distances are. The seriousness and significance of this difficulty was further emphasized by Sanders [1982].

In KAM theory one also finds that the splitting of separatrices is governed by systems of the form considered here, and so would be formally beyond all orders if a power series in the perturbation parameter were developed. Indeed, a formal calculation based on the Melnikov method shows that the splitting of separatrices is probably of exponentially small order, a phenomenon discussed in Arnold's book (see especially page 397). Zehnder [1973] also shows that there are transverse homoclinic orbits for *generic* nonlinearities in KAM theory. In a similar fashion, the same type of behavior arises in the unfolding of degenerate singularities, such as the interaction of the Hopf and the pitchfork or transcritical bifurcation (see Guckenheimer and Holmes [1983] and Scheurle and Marsden [1984] for discussions of this bifurcation and for further references). See also the paper of Dangelmayr and Knobloch [1987] for the case of symmetry breaking bifurcations and the work of Golubitsky and Stuart [1986] for the application of unfolding techniques to the Taylor Couette problem, where it is expected that similar phenomena will occur. Since these splittings are exponentially small, standard methods for detecting them based on averaging, normal forms, or perturbation expansions using power series in ε, will not succeed. This is also behind the fact that one has, in general, divergence of the Birkhoff series.

In this paper, we give a new method that overcomes many of these difficulties. We give sharp upper bounds, with the constant in the exponential being the distance of the nearest pole in the complex t-plane of the unperturbed homoclinic orbit to the real axis. If a high enough power of ε is present in front of the forcing term then there is a lower bound for the splitting, which is also exponentially small with the **same** exponential factor. In the latter case, the Melnikov integral is sufficient to predict the transversality of, and to estimate the magnitude of the splitting. In general, however, it appears that one must go to higher orders to obtain a predictable criterion, in which case one has to revert to an intricate calculation, or else use the Cushman-Koslov analyticity argument, which only gives a generic result.

Our approach is based on a ***convergent*** iteration scheme using the Liapunov-Perron method and a special extension of the scheme to the complex t plane that enables us to estimate the splitting distance. A naive extension will run into difficulties since the forcing term $\sin(t/\varepsilon)$ is exponentially big for t in a complex strip. As mentioned above, these estimates relate the singularities in the complex plane and the factor in the exponential [the separatrix for the homoclinic orbit in the pendulum case has one component given by sech t, which has simple poles at $t = \pm i\pi/2$, and the corresponding exponential factor is $\exp(-\pi/2\varepsilon)$.] Because of this, one can conjecture a connection between the results here and the Painlévé property. The work of Ziglin [1982], van Moerbeke [1983] and Bountis et.al.[1986] may be helpful in this regard. The key to our method is that the special iteration scheme preserves the exponentially small structure, with the same factor in the exponent at each stage, controls the possible accumulation in the pole behavior, and exhibits the cancellation of terms that move each of the stable and unstable manifolds (and the hyperpolic fixed point) an amount that is algebraic in ε, even though the difference between them is exponentially small. The key points of the proof are given here; a more detailed paper is in preparation.

There have been other approaches to exponentially small phenomena based on asymptotic methods. For example, the works of Meyer [1976] on adiabatic variation, Meyer [1982] on wave reflection and quasiresonance, Segur and Kruskal [1987] on breathers in the φ^4 model, and Kruskal and Segur [1987] on dendritic crystals, use this technique. While there seem to be some points in common with our approach, it appears that additional work would be needed to apply and justify the estimates that we obtain for separatrix splitting.

Acknowledgements We thank B. Birnir, J. Carr, B. Chirikov, W. Eckhaus, M. Golubitsky, J. Guckenheimer, K. Kirchgassner, M. Kummer, R. Meyer, and H. Segur for useful comments. The research of P. Holmes was partially supported by NSF grant MSM 84-02069 and AFOSR contract 84-0051, and that of J. Marsden and J. Scheurle was partially supported by NSF grant DMS 87-02502 and DOE contract DE-AT03-85ER12097.

§1 *Preliminaries*

We begin by recalling a few basic facts about the standard Poincaré - Melnikov method. The phase portrait of the simple pendulum

$$\ddot{\phi} + \sin\phi = 0 \tag{1.1}$$

is as shown in Figure 1 in the (ϕ, v) plane, where $v = d\phi/dt$. The homoclinic orbits shown there are explicitly given by the solutions

$$\left.\begin{aligned} \bar{\phi}(t) &= \pm 2\tan^{-1}(\sinh t) \\ \bar{v}(t) &= \pm 2\,\mathrm{sech}(t) \end{aligned}\right\} . \tag{1.2}$$

We observe for later use that sech t has poles in the complex t-plane at $t = \pm i\pi/2$.

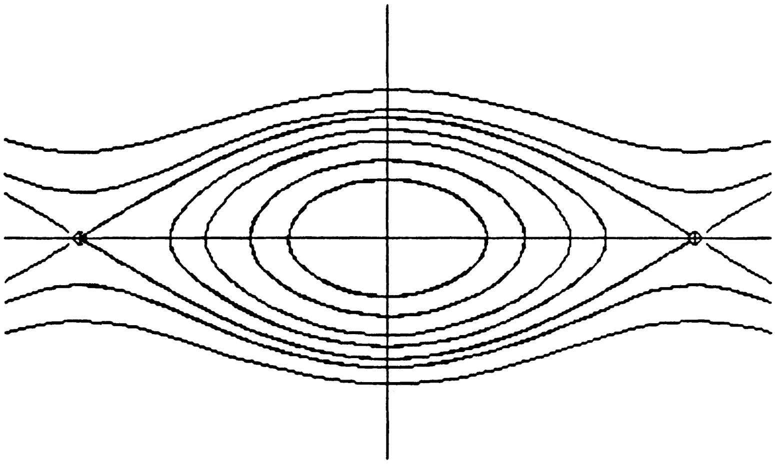


Figure 1. Phase portrait of the simple pendulum.

If we modify (1) by including a T-periodic forcing, we get the equation

$$\ddot{\phi} + \sin \phi = \varepsilon f(t) \ , \tag{1.3}$$

for which the dynamics is conveniently described by the Poincaré map $P(t_0) : \mathbb{R}^2 \to \mathbb{R}^2$ defined by mapping initial conditions (ϕ_0, v_0) at time t_0 to the solution after one period, at time $t_0 + T$. For small ε, the hyperbolic fixed points for (1.1) get perturbed to fixed points for $P(t_0)$ (i.e., periodic orbits for (1.3)) and $P(t_0)$ has stable and unstable manifolds at these fixed points which, in general, intersect. This leads one to define the ***splitting distance***

$$d = \max_{t_0} d(t_0) \tag{1.4}$$

and the ***splitting angle***

$$\alpha = \max_{t_0} \alpha(t_0) \tag{1.5}$$

where, for any t_0, $d(t_0)$ and $\alpha(t_0)$ are shown in Figure 2.

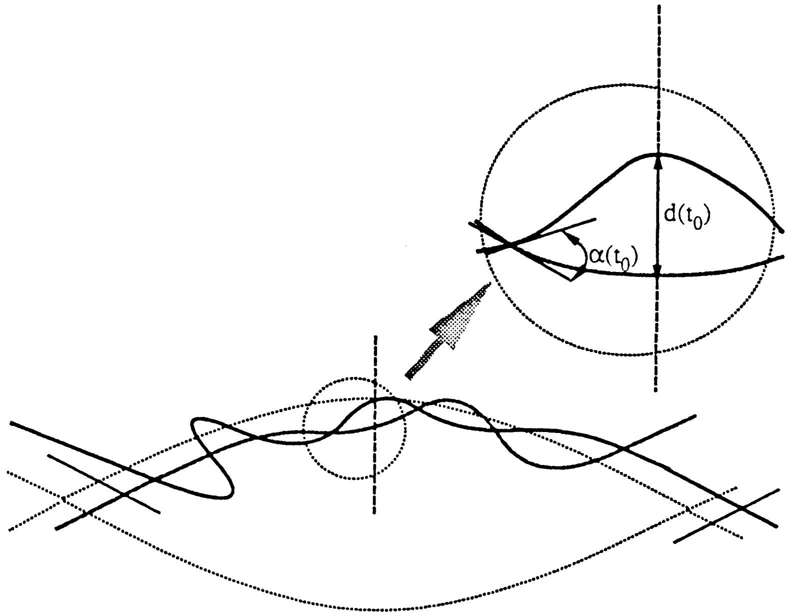


Figure 2. The splitting distance and angle.

This splitting distance and angle are correlated with the thickness of the stochastic layer; the trajectories of some sample points are shown in Figure 3 for illustration. One should be cautious, however, that there is little analytic work on the precise relation between the splitting distance and angle and the thickness of the stochastic layer. However, the celebrated horseshoe construction of Poincaré, Birkhoff, and Smale does establish that a transversal intersections ($\alpha \neq 0$) implies the existence of complicated orbits (and periodic orbits with arbitrarily high period) and thus warrants using the word "chaotic" to describe the dynamics.

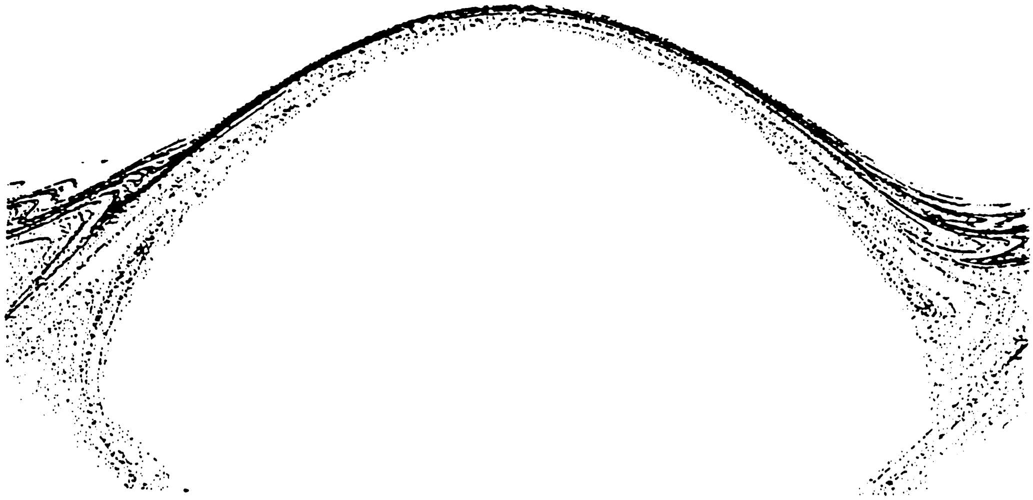


Figure 3. Orbits of points under the Poincaré map of (1.3) for $f(t) = (0.1) \sin t$ (plot courtesy of B. Birnir).

The splitting distance is typically measured by a ***Poincaré-Melnikov function***. For a planar Hamiltonian system

$$\left.\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} + \varepsilon \frac{\partial K}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} - \varepsilon \frac{\partial K}{\partial q} \end{aligned}\right\}, \tag{1.6}$$

where $K = K(q, p, t)$ is a perturbing T-periodic Hamiltonian, the Poincaré - Melnikov function is the T-periodic function

$$M(t_0) = \int_{-\infty}^{\infty} \{H, K\}(\bar{q}(t), \bar{p}(t), t + t_0)\, dt \ , \tag{1.7}$$

the splitting distance is proportional to

$$d = \varepsilon \max_{t_0} | M(t_0) | + O(\varepsilon^2) , \tag{1.8}$$

and the angle is proportional to

$$\alpha = \varepsilon \max_{t_0} | M'(t_0) | + O(\varepsilon^2) . \tag{1.9}$$

This follows readily from an analysis of the first variation equation. See, for instance, Holmes and Marsden [1982] and Guckenheimer and Holmes [1983] for discussions and proofs. For example, for

$$\ddot{\phi} + \sin \phi = \varepsilon \sin \omega t , \tag{1.10}$$

one finds

$$M(t_0) = 2\pi \operatorname{sech}\left(\frac{\pi\omega}{2}\right) \cos(\omega t_0) \tag{1.11}$$

by evaluating (1.7) using residues, noting the pole of $\operatorname{sech} t$ at $i\pi/2$. Thus,

$$d \approx 2\pi\varepsilon \operatorname{sech}\left(\frac{\pi\omega}{2}\right) \quad \text{and} \quad \alpha \approx 2\pi\omega\varepsilon \operatorname{sech}\left(\frac{\pi\omega}{2}\right) \tag{1.12}$$

§2 *Exponentially Small Splittings*

To illustrate the main idea, first consider the rapidly forced pendulum

$$\ddot{\phi} + \sin \phi = \varepsilon \sin(t/\varepsilon) . \tag{2.1}$$

If one applies equation (1.12), one finds the splitting distance should be of the order

$$d \approx 2\pi\varepsilon e^{-\pi/2\varepsilon} . \tag{2.2}$$

(The constant 2π is not important -- it depends on the units of measure; for example, it may be convenient to use the unperturbed energy as a distance measure.) However, (2.2) is not easy to justify; for one thing, the errors in (1.8) are $O(\varepsilon^2)$, while (2.2) is already smaller than any power of ε.

There are two main results for problems of this sort as follows:

UPPER ESTIMATE *Consider*

$$\ddot{\phi} + \sin\phi = \delta \sin(t/\varepsilon) . \tag{2.3}$$

For any $\eta > 0$ *there is a* $\delta_0 > 0$ *and a constant* $C = C(\eta, \delta_0)$ *such that, for all* ε *and* δ *satisfying* $0 < \varepsilon \le 1$ *and* $0 < \delta \le \delta_0$, *we have*

$$\text{splitting distance} \le C\delta \exp\left[-\left(\frac{\pi}{2} - \eta\right)\frac{1}{\varepsilon}\right] . \tag{2.4}$$

There is a similar estimate for the *splitting angle.*

LOWER ESTIMATE AND SHARP UPPER ESTIMATE *Consider*

$$\ddot{\phi} + \sin\phi = \varepsilon^p \delta \sin(t/\varepsilon) . \tag{2.5}$$

If $p \ge 8$, *then there is a* $\delta_0 > 0$ *and (absolute) constants* C_1 *and* C_2 *such that, for all* ε, δ *satisfying* $0 < \varepsilon \le 1$ *and* $0 < \delta \le \delta_0$, *we have*

$$C_2 \varepsilon^p \delta e^{-\pi/2\varepsilon} \le \text{Splitting Distance} \le C_1 \varepsilon^p \delta e^{-\pi/2\varepsilon} . \tag{2.6}$$

Observe that $\pi/2$, which appears in the exponent in both estimates, is the distance from the real axis to the closest pole of sech t; see Figure 4.

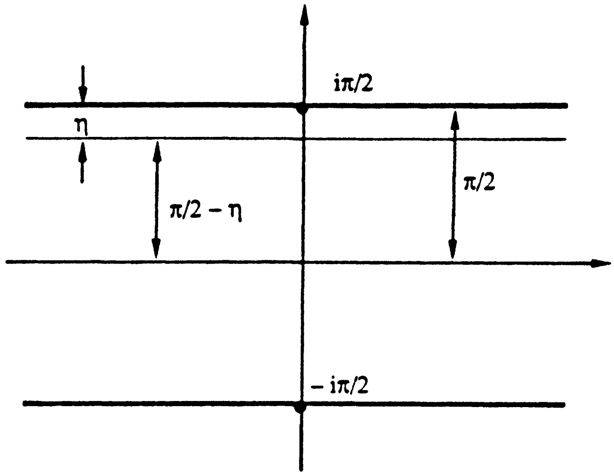


Figure 4. The exponent in the exponential estimate is the distance to the nearest pole of the homoclinic orbit in the complex t-plane.

These estimates are special cases of estimates for a planar system

$$\dot{u} = g(u, \varepsilon) + \varepsilon^p \delta h\left(u, \varepsilon, \frac{t}{\varepsilon}\right), \tag{2.7}$$

where one assumes:

- g and h are entire in u and ε;
- h is of Sobolev class H^1 (for the splitting distance results) or H^2 (for the splitting angle results) and T-periodic in the variable $\theta = t/\varepsilon$;
- $\dot{u} = g(u, \varepsilon)$ has a homoclinic orbit $\bar{u}(\varepsilon, t)$ analytic in t on a strip in the complex $t-$plane, with width r.

under additional assumptions on the fundamental solution of the first variation equation

$$\dot{v} = D_u g(\bar{u}, \varepsilon) \cdot v$$

which can be checked to hold in the pendulum example, there are analogues of the upper and lower estimates above for this general situation, with $\pi/2$ replaced by r. We shall give additional details in the subsequent sections. The proofs depend on detailed estimates of the terms in an iterative process in the complex strip that are used to define the invariant manifolds. It is important to extend these iterates to the complex strip in the proper way; as we have mentioned, $\sin(t/\varepsilon)$ becomes very large for complex t and naively extended iteration procedures for the stable and unstable manifolds will lead to unbounded sequences of functions.

§3 *The Hypotheses and Set-Up.*

We recall some of the general theory and the ideas involved in the proofs from Holmes, Marsden and Scheurle [1988] for the convenience of the reader. We consider a differential equation of the following form

$$\dot{u} = g(u, \varepsilon) + \varepsilon^p \delta\, h\left(u, \varepsilon, \frac{t}{\varepsilon}\right) \tag{3.1}$$

where p is a positive integer (one can think of the term ε^p as being part of h or as being divided between δ and h as is appropriate), $u = (x, y) \in \mathbb{R}^2$, $\varepsilon > 0$, $g(u, \varepsilon) : \mathbb{C}^2 \times \mathbb{C} \to \mathbb{C}^2$ is entire, h is entire in (u, ε) and is 2π-periodic and C^1 (or of Sobolev class H^1) in its third arguement t / ε. Both g and h are assumed to be real for real values of their arguments. (The H^1 assumption on h is needed below to get bounds on the ***splitting distance***; for exponentially small bounds on the ***angle*** at a transversal intersection, we need to assume that h is of class H^2 in t / ε - see Remark 1 at the end of section 5.)

Although it is not really needed, we shall introduce a symmetry condition for simplicity. (A more general case without this condition is discussed at the end of this paper.) Namely, we assume that the system (2.1) is ***reversible*** in the sense that there is a real linear reflection operator $R : \mathbb{R}^2 \to \mathbb{R}^2$ i.e. a 2×2 matrix satisfying $R^2 = \text{Identity}$, with eigenvalues ± 1, and satisfying the following conditions:

$$g(Ru, \varepsilon) = -Rg(u, \varepsilon) \quad \text{and} \quad h(Ru, \varepsilon, -t/\varepsilon) = -Rh(u, \varepsilon, t/\varepsilon). \tag{3.2}$$

For instance, for the example given in the preceding section, we take

$$g(x, y) = (y, -\sin x), \quad h((x,y), \varepsilon, t/\varepsilon) = (0, \sin(t/\varepsilon)) \quad \text{and} \quad R(x, y) = (-x, y).$$

Assume that the homogeneous equation $\dot{u} = g(u, \varepsilon)$ has a homoclinic orbit Γ_ε which is asymptotic to a hyperbolic fixed point; we write the homoclinic orbit as $u = \bar{u}_\varepsilon(t)$. We shall assume that $\bar{u}_\varepsilon(t)$ has an analytic continuation in the complex t plane into a complex strip S_ε defined to be the set of complex numbers z such that $|\operatorname{Im} z| \leq r_\varepsilon$, where r_ε is some positive real number. Typically, $\bar{u}_\varepsilon(t)$ will be an analytic function in t, and will be analytic in a strip, with r_ε smaller than the smallest distance of the poles to the real axis. We assume that the initial condition of the homoclinic orbit satisfies $R\bar{u}_\varepsilon(0) = \bar{u}_\varepsilon(0)$, so that $R\,\bar{u}_\varepsilon(-t) = \bar{u}_\varepsilon(t)$. In the example, the homoclinic orbit is given by

$$(\bar{\varphi}(t), \bar{\zeta}(t)) = 2(\tan^{-1}(\sinh t), \operatorname{sech} t)$$

where $\zeta = d\varphi / dt$, so this assumption is clear.

As indicated in section 2, there are two cases to consider. In the first, we choose $p = 0$ and $r_\varepsilon = (\pi/2) - \eta$ for a fixed $\eta > 0$ and in the second, we choose $r_\varepsilon = (\pi/2) - \varepsilon$ (and later we will require $p \geq 8$).

The first variation equation

$$\dot{v} = D_u g(\bar{u}_\varepsilon(t), \varepsilon) \cdot v \tag{3.3}$$

has exponential dichotomies corresponding to t in $\mathbb{R}^+$ and $\mathbb{R}^-$ (see for example, Hartman [1982], Ch. 13). That is, the plane splits into two subbundles

$$\mathbb{R}^2 = X_{t,\varepsilon} \oplus Y_{t,\varepsilon}. \tag{3.4}$$

that are invariant under the evolution of the first variation equation, such that the components φ_1 and φ_2 of the fundamental solution matrix $\varphi(t, \tau)$, which are defined by restriction of φ to $X_{\tau,\varepsilon}$ and $Y_{\tau,\varepsilon}$ respectively, satisfy the inequalities:

$$|\varphi_1(t,\tau)\xi| \le K e^{-\alpha(\tau - t)} |\xi| \qquad (\text{if } \tau \ge t) \tag{3.5a}$$

$$|\varphi_2(t,\tau)\eta| \le K e^{-\alpha(t - \tau)} |\eta| \qquad (\text{if } \tau \le t) \tag{3.5b}$$

for t, τ satisfying $-\infty < t, \tau \le 0$, and where $\xi \in X_{\tau,\varepsilon}$, and $\eta \in Y_{\tau,\varepsilon}$. For simplicity of notation, we have suppressed the possible ε dependence of φ, φ_1, and φ_2. Similarly,

$$|\varphi_1(t,\tau)\xi| \le K e^{\alpha(\tau - t)} |\xi| \qquad (\text{if } \tau \le t) \tag{3.6a}$$

$$|\varphi_2(t,\tau)\eta| \le K e^{\alpha(t - \tau)} |\eta| \qquad (\text{if } \tau \ge t) \tag{3.6b}$$

for t, τ satisfying $0 \le \tau,\ t < \infty$, and where $\xi \in X_{\tau,\varepsilon}$ and $\eta \in Y_{\tau,\varepsilon}$. In these equations α and K are positive constants. The constant α is related to the eigenvalues of the linearization of the equation at the hyperbolic fixed point. We choose the dichotomies amongst all possible ones by the requirement that at $t = 0$, the bundles satisfy

$X_{0,\varepsilon}$ is the eigenspace of R corresponding to the eigenvalue -1
$Y_{0,\varepsilon}$ is the eigenspace of R corresponding to the eigenvalue 1

In the example, we take the bundles to be the tangential space to the homoclinic orbit and the normal direction at the point $t = 0$ swept out by the flow of the first variation equation. The first variation solution is explicitly found in this case to be as follows

$$\varphi_1(t,\tau)\xi = \frac{1}{2}\,[\{\cosh\tau + \operatorname{sech}\tau - \tau\operatorname{sech}\tau\tanh\tau\}\,\psi - \{\sinh\tau + \tau\operatorname{sech}\tau\}\upsilon]\begin{pmatrix} \operatorname{sech} t \\ -\operatorname{sech} t \tanh t \end{pmatrix} \tag{3.7a}$$

$$\varphi_2(t,\tau)\eta = \frac{1}{2}[\{\operatorname{sech}\tau\tanh\tau\}\psi - \{\operatorname{sech}\tau\}\upsilon\,]\begin{pmatrix} \sinh t + t\operatorname{sech} t \\ \cosh t + \operatorname{sech} t - t\operatorname{sech} t\tanh t \end{pmatrix} \tag{3.7b}$$

where v has components (ψ, υ) in the original coordinate system and where ξ and η are the projections of v onto the spaces $X_{\tau,\varepsilon}$ and $Y_{\tau,\varepsilon}$ respectively.

§4 *The Iteration Method.*

We shall locate the stable and unstable manifolds of the perturbed equation using a special Liapunov-Perron type iteration scheme that is coupled with a Fourier expansion and a certain extension to the complex t-plane. It will be important to keep track of the estimates during the iteration process itself. We write the perturbed stable and unstable manifold of the hyperbolic fixed point as follows:

$$u^{\pm}(t, t_0, \varepsilon, \delta) = \overline{u}_{\varepsilon}(t - t_0) + \delta v^{\pm}(t - t_0, t_0, \varepsilon, \delta) \tag{4.1}$$

where $t \geq t_0$ in the + case, and $t \leq t_0$ in the – case. Dropping the $\pm$ and writing $s = t - t_0$, we get an equation for v, regarded as a function of s, t_0, ε and δ by substituting (3.1) into (2.1); the stable and unstable manifolds will later be picked out by looking for bounded solutions of the resulting fixed point problem. We first compute (suppressing the ε and δ dependence for the moment):

$$\dot{v} - A(s)v = F(v, s, (s + t_0)/\varepsilon) \tag{4.2}$$

where

$$A(s) = D_u\, g(\overline{u}_{\varepsilon}(s), \varepsilon) \tag{4.3}$$

and

$$F\left(v, s, \frac{s + t_0}{\varepsilon}\right) =$$

$$\frac{1}{\delta}\left[g(\bar{u}_{\varepsilon}(s) + \delta v, \varepsilon) - g(\bar{u}_{\varepsilon}(s), \varepsilon) - A(s)\, \delta v \right] + \varepsilon^{p} h\left(\bar{u}_{\varepsilon}(s) + \delta v, \varepsilon, \frac{s + t_0}{\varepsilon}\right) \tag{4.4}$$

Here the ε^p is grouped with h, but we also could group appropriate powers with δ; this freedom is important later. We look for solutions of (4.2) that are uniformly bounded in the $\pm$ cases by reformulating it as a fixed point problem for the following integral equation

$$v = K^{\pm} F\left(v, s, \frac{s + t_0}{\varepsilon}\right) \tag{4.5}$$

where $K^{\pm}$ are the linear operators (again with the ε dependence supressed) that are given by

$$(K^+ f)(s) = \int_0^s \varphi_1(s, \sigma)\, f_1(\sigma)\, d\sigma - \int_s^\infty \varphi_2(s, \sigma)\, f_2(\sigma)\, d\sigma \quad \text{for all } s \geq 0 \tag{4.6a}$$

$$(K^- f)(s) = \int_0^s \varphi_1(s,\sigma)\, f_1(\sigma)\, d\sigma + \int_{-\infty}^s \varphi_2(s, \sigma)\, f_2(\sigma)\, d\sigma \quad \text{for all } s \leq 0 \tag{4.6b}$$

Here the projections of f onto the dichotomy subspaces (2.4) are denoted f_1 and f_2. With $K^\pm$ thought of as operating on the space of bounded continuous (vector valued) functions on $\mathbb{R}^\pm$, one gets a contraction mapping for sufficiently small δ_0 and a unique fixed point that can be solved for iteratively (see, for example, Hartman [1982], Chapter 12, part III). The stable and unstable manifolds are determined by the fixed points, which we denote by $v^+(s, t_0)$ and $v^-(s, t_0)$, where we again suppress the dependence on ε and δ. We start the iteration scheme in each case with the zero solution and then define inductively

$$^{(n+1)}v^\pm(s, t_0) = K^\pm F\left({}^{(n)}v^\pm(s, t_0),\ s,\ \frac{s + t_0}{\varepsilon} \right) \tag{4.7}$$

so that $^{(n)}v^\pm(s, t_0)$ converges to $v^\pm(s, t_0)$ as $n \to \infty$.

§5 *Estimates for the Splitting Distance*

The next step is to estimate the splitting distance between the stable and unstable manifolds. To do this we estimate the following quantity:

$$\Delta(t_0, \varepsilon, \delta) = \left| u^+(t_0, t_0, \varepsilon, \delta) - u^-(t_0, t_0, \varepsilon, \delta) \right| = \delta \left| v^+(0, t_0, \varepsilon, \delta) - v^-(0, t_0, \varepsilon, \delta) \right| \tag{5.1}$$

where $u^\pm(s + t_0, t_0, \varepsilon, \delta) = \bar{u}_\varepsilon(s) + \delta v^\pm(s, t_0, \varepsilon, \delta)$ are the stable and unstable manifolds of the perturbed equation. The ***splitting distance*** is defined to be the maximum of $\Delta(t_0, \varepsilon, \delta)$ over one $2\pi\varepsilon$ period in t_0. We extend each of the solutions of the iteration scheme (4.7) to strips in the complex t plane. We do this in a way that makes the iterates uniformly bounded in the appropriate (ε dependent) half strips

$$S^+ = \left\{ z \in \mathbb{C} \mid |\operatorname{Im} z| \leq r_\varepsilon \text{ and } \operatorname{Re} z \geq 0 \right\}$$

$$S^- = \left\{ z \in \mathbb{C} \mid |\operatorname{Im} z| \le r_\varepsilon \text{ and } \operatorname{Re} z \le 0 \right\}$$

and the vertical segment

$$S^0 = \left\{ z \in \mathbb{C} \mid |\operatorname{Im} z| \le r_\varepsilon \text{ and } \operatorname{Re} z = 0 \right\}.$$

Thus, a uniform exponential estimate on the distance between the iterates

$$^{(n)}\Delta(t_0, \varepsilon, \delta) = \delta \left| {}^{(n)}v^+(0, t_0, \varepsilon, \delta) - {}^{(n)}v^-(0, t_0, \varepsilon, \delta) \right| \tag{5.2}$$

produces the corresponding result for the limiting solution as $n \to \infty$.

Let $\theta = (s + t_0) / \varepsilon$ and consider the following iteration scheme for an (ε and δ dependent) function $w(s, \theta)$:

$$^{(n+1)}w^{\pm}(s, \theta) = L^{\pm} F\left({}^{(n)}w^{\pm}(s, \theta), s, \theta \right) \tag{5.3}$$

where we start with $w = 0$, and define $L^{\pm}$ as follows. For any vector valued function

$$f(s, \theta) = \sum_{k=-\infty}^{\infty} f_k(s)\, e^{ik\theta}, \text{ where } s \in S^{\pm} \text{ and } \theta \in \mathbb{R},$$

we set

$$L^{\pm} f = \sum_{k=-\infty}^{\infty} a_k^{\pm}(s)\, e^{ik\theta}, \tag{5.4}$$

where we define

$$a_0^+(s) = \int_0^s \varphi_1(s,\sigma)\, f_{0,1}(\sigma) d\sigma - \int_s^\infty \varphi_2(s,\sigma)\, f_{0,2}(\sigma)\, d\sigma \tag{5.5a}$$

$$a_0^-(s) = \int_0^s \varphi_1(s,\sigma)\, f_{0,1}(\sigma) d\sigma + \int_{-\infty}^s \varphi_2(s,\sigma)\, f_{0,1}(\sigma)\, d\sigma \tag{5.5b}$$

$$a_k^+(s) = \int_{\pm ir_\varepsilon}^{s} \varphi_1(s,\sigma)\, e^{ik(\sigma-s)/\varepsilon} f_{k,1}(\sigma)d\sigma - \int_{s}^{\infty} \varphi_2(s,\sigma)\, e^{ik(\sigma-s)/\varepsilon} f_{k,2}(\sigma)d\sigma \qquad (5.6a)$$

and

$$a_k^-(s) = \int_{\pm ir_\varepsilon}^{s} \varphi_1(s,\sigma)\, e^{ik(\sigma-s)/\varepsilon} f_{k,1}(\sigma)d\sigma + \int_{-\infty}^{s} \varphi_2(s,\sigma)\, e^{ik(\sigma-s)/\varepsilon} f_{k,2}(\sigma)d\sigma \qquad (5.6b)$$

Here the projections of the f_k on the dichotomy subspaces (2.4) are denoted $f_{k,1}$ and $f_{k,2}$, and in (5.6) we choose "+" if $k > 0$ and "−" if $k < 0$.

One now introduces the function spaces $X^\pm$ of f's with the Sobolev H^1 norm in the variable θ and the sup norm over $S^\pm$ in the variable s. Also, let X^0 be the space of f's endowed with the H^1 norm in θ and the sup norm with weight $\exp[(r_\varepsilon - |s|)/\varepsilon]$ in s over S^0. We make the assumption that *the fundamental solution* $\varphi(t, \tau)$ *of the linear equation (3.3) has an analytic continuation into the complex strip* S_ε *in both* t *and* τ *such that the estimates (3.5) and (3.6) hold with* t *and* τ *on the right hand sides replaced by* Re t *and* Re τ. This is verified in our example using the representation (3.7).

Fact 1 *Define the (ε and δ dependent) maps*

$$G^\pm : w(s, \theta) \mapsto L^\pm F(w(s, \theta), s, \theta) \qquad (5.7)$$

where F *and* $L^\pm$ *are defined in (4.4) and (5.4). Then for each* ε *and* δ, $G^\pm$ *is a bounded map (maps bounded sets to bounded sets) of* $X^\pm$ *to itself.*

This is proved using Sobolev type estimates; in fact it is useful to break the argument into the two steps of consideration of the maps $w \mapsto F(w, s, \theta)$, to which a standard composition (or Ω lemma) argument can be applied and a study of the operators $L^\pm$ using explicit Fourier series methods. In general, the bound on the image set depends on ε; it could grow as $\varepsilon \to 0$. To prevent this one needs to balance the growth of the norm of $L^\pm$ and the accumulation of poles in w with the powers of ε in front of h. It is at this stage that some powers of ε in front of h are needed to get uniformity in ε; this is required for the lower and upper estimates that have the exact distance to the pole in the exponent, and not a smaller one.

Fact 2 *With the assumptions as above, let* $B^{\pm}$ *be a bounded subset in* $X^{\pm}$. *Then there exist constants* C_1 *and* C_2 *depending only on the bounds of the set* $B^{\pm}$ *such that for any pair of functions* w^+ *and* w^- *in* $B^{\pm}$, *we have*

$$\| G^-(w^-)_{|S^0} - G^+(w^+)_{|S^0} \|^0 \le C_1 \| w^-_{|S^0} - w^+_{|S^0} \|^0 + C_2 \tag{5.8}$$

where $\| \cdot \|^0$ *is the norm on the space* X^0.

This is proved by an analysis of the formulas explicitly representing the maps G^+ and G^-. For example, to estimate the difference of the terms coming from the first terms of (5.6 a,b), we use a Lipschitz property of the composition map $w \mapsto F(w, s, \varepsilon, \theta)$ and this contributes to the first term on the right hand side of (5.8). To estimate the second terms in (5.6a,b), for $k > 0$, one uses Cauchy's theorem to shift the path of integration in complex σ–plane along the real axis from s to ∞ to a path up the imaginary axis to the point $i\, r_\varepsilon$ and then along the line $\operatorname{Im} \sigma = r_\varepsilon$ to ∞. Because of the way the extensions to the complex plane have been chosen and the bounds obtained in **Fact 1**, the integral along the line $\operatorname{Im} \sigma = r_\varepsilon$ contributes to the second term on the right hand side (5.8). After subtraction with the corresponding terms in (5.6b), the other terms contribute to the first term on the right side of (5.8).

In the preceding argument, the case $k = 0$ requires special attention. These terms would contribute algebraic, not exponentially small terms, were it not for a crucial cancellation. As above, one first reduces to the case $s = 0$ by noting that the difference of the terms contributes to the first term on the right hand side of (5.8). Then we are left to estimate the difference between the terms

$$\Delta_1 = \int_0^\infty \varphi_2(0, \sigma) f_{0,2}(\sigma) d\sigma \quad \text{and} \quad \Delta_2 = - \int_{-\infty}^0 \varphi_2(0, \sigma) f_{0,2}(\sigma) d\sigma \tag{5.9}$$

But one checks that we have the symmetry $\Delta_1 = R\, \Delta_2 = \Delta_2$, and so these terms cancel. (See the remark below regarding this symmetry assumption).

Now assume that there are bounded neighborhoods $B^{\pm}$ in $X^{\pm}$ of 0 which are independent of ε and δ and which are mapped into themselves by the ε and δ dependent mappings $G^{\pm}$ and so our iterates remain in $B^{\pm}$ for all n. This requires an estimate on the poles that occur in the mapping $G^{\pm}$ and the balance between this behaviour and the factors of ε in front of the nonlinear inhomogeneous term h. By choosing δ_0 sufficiently small, we can arrange that C_1 in **Fact 2** is less than 1/2. By the contraction mapping principle, the iterates ${}^{(n)}w^{\pm}$ converge to $w^{\pm}$ in $X^{\pm}$. For real s, $w^{\pm}$ are related to the stable and unstable manifolds in the following sense:

Fact 3 *For each real* t_0, *there are* $t_0^\pm$ *with* $|t - t_0^\pm|$ *of order* δ *such that for all real* t,

$$u^\pm(t, t_0, \varepsilon, \delta) = \overline{u}_\varepsilon(t - t_0^\pm) + \delta\, w^\pm(t - t_0^\pm, t/\varepsilon)$$

We prove the required estimates as follows. Making the inductive assumption that $\| w^+ - w^- \|^0$ is bounded by $2C_2$, the estimate (5.8) shows that the next iterate obeys the same inequality. Passing to the limit, using **Fact 3** and rescaling back to the original variables then gives the desired result that the splitting is bounded above by $2\, C_2\, \delta \exp(-r_\varepsilon / \varepsilon)$. To get a lower bound, one needs to show that the higher iterates are of lower order than the first iterate. The first term in the iteration is the same as one would get from the Melnikov method, which, in the example can be evaluated explicitly. To estimate the higher order terms requires one to show that the power of ε in front of h can be used to control the growing norm of the operators L^+ and L^- as $\varepsilon \to 0$, and still produce an overall power that increases with each iteration. This is how the condition $p \geq 8$ arises in the example; in that case, we divide ε^8 into ε^3 to go with h and ε^5 to go with δ. These specific powers are chosen to (i) balance the growth in the norm of $L^\pm$ as $\varepsilon \to 0$ and (ii) to ensurethat the difference between the first and the higher iterates will be small compared to the first iterate. Notice that this analysis is not based on an asymptotic series argument, but rather on a comparison between the first term in the iteration scheme and the subsequent iterates.

Remark 1 From the reversibility assumption, it follows that $\Delta(0, \varepsilon, \delta) = 0$ for all ε and δ. Therefore, the separatrices obviously intersect in this case. However, the proof of the transversality of the intersection requires additional estimates for

$$\left| \frac{\partial}{\partial t_0} \Delta(t_0, \varepsilon, \delta) \,\Big|_{t_0=0} \right|$$

which measures the angle of intersection of the separatrices. Estimates for this again come in two cases, namely upper and lower estimates. These estimates are of the same exponentially small form as those for the splitting distance, with an additional factor of $1/\varepsilon$. Estimates for the t_0 derivatives of the iterates in (5.2) can be obtained by the same techniques as for the iterates themselves using the space H^2 instead of H^1 in the above setting. In the example, again the assumption that $p \geq 8$ implies that the separatrices do have a transversal intersection with an exponentially small angle of intersection. ♦

Remark 2 If we consider a one parameter unfolding

$$\dot{u} = g(u, \varepsilon, \lambda) + \varepsilon^p \delta h\left(u, \varepsilon, \frac{t}{\varepsilon}, \lambda\right), \quad \text{where } \lambda \in \mathbb{R}, \tag{5.10}$$

which agrees with the problem (3.1) for $\lambda = 0$, then under a certain non-degeneracy condition with respect to the parameter λ, a slight modification of our method yields the following result without the reversibility assumption. For sufficiently small δ_0, all $t_0 \in \mathbb{R}$, and all ε and δ satisfying $0 < \varepsilon \leq 1$ and $0 \leq \delta \leq \delta_0$, there exists a value of λ given by an expression of the form

$$\lambda = \lambda_0(\varepsilon, \delta) + \lambda_1(t_0, \varepsilon, \delta)$$

such that for this λ-value, (5.10) has a unique solution which is δ_0-close to $\bar{u}_\varepsilon(t - t_0)$ (that is, the difference in the sup norm is $\leq$ Const $\cdot\ \delta_0$) for all $t \in \mathbb{R}$. In fact, one can find sucessive approximations λ_n of this λ-value such that in (4.7)

$${}^{(n)}v^+(0, t_0) = {}^{(n)}v^-(0, t_0)$$

holds for all n. Thus it follows that $v^+(0, t_0) = v^-(0, t_0)$, and the desired solution is given by

$$u(s + t_0, t_0, \varepsilon, \delta) = \bar{u}_\varepsilon(s) + \delta v^-(s, t_0, \varepsilon, \delta) \quad \text{for all } s \leq 0$$

$$u(s + t_0, t_0, \varepsilon, \delta) = \bar{u}_\varepsilon(s) + \delta v^+(s, t_0, \varepsilon, \delta) \quad \text{for all } s \leq 0 \tag{5.12}$$

Here $\lambda_0(\varepsilon, 0) = 0$, and we have the estimate

$$|\lambda_1(t_0, \varepsilon, \delta)| \leq C\, \delta \exp(-r_\varepsilon / \varepsilon). \tag{5.13}$$

Moreover, near $\lambda = 0$, there are no other λ-values such that (4.10) has a solution which is δ_0-close to $\bar{u}_\varepsilon(t - t_0)$ for some t_0. Thus, if we replace δ by ε for example, then there is an exponentially thin wedge-like zone in (ε, λ) - space such that the local stable and unstable manifolds of the perturbed hyperbolic fixed point intersect if and only if (ε, λ) is contained in this zone. Also, the splitting distance is bounded above by $C\, \varepsilon \exp(-r_\varepsilon/\varepsilon)$ for such values of ε and λ. This zone of (ε, λ) - values corresponds to the Arnold tongues of perturbed periodic solutions (cf. Scheurle [1986]).

Our techniques also show that $\Delta(t_0, \varepsilon, \delta)$ in (5.1) is bounded above by $C\,\delta \exp(-r_\varepsilon/\varepsilon)$ as $\varepsilon \to 0$ whenever the stable and the unstable manifolds of the perturbed hyperbolic fixed point of (3.1) intersect in a solution which is δ_0-close to $\overline{u}_\varepsilon(t - t_0)$ with some $t_0 = t_0^*(\varepsilon, \delta)$ for all small ε. Besides reversible problems, where $t_0^*(\varepsilon, \delta) = 0$ for all ε, Hamiltonian systems also have this property (cf. Arnold [1965]). We point out, however, that the equation that we have considered as an example is locally, but not globally Hamiltonian. Our theory requires a homoclinic orbit, so we have chosen the phase space to be the cylinder. ♦

§6 *A 2:1 Resonance and KAM Theory*

In KAM theory, arguments based on numerical evidence and formal calculations lead to the conjecture that one has exponentially fine splittings and exponentially long escape times. Some rigorous but rough upper bounds for this phenomena have been given by Nekhoroshev [1971, 77] and Neishtadt [1984]; see also the discussion in Arnold [1978, pp. 395ff and 407], Chirikov [1979], and Simo and Fontich [1985]. The analyticity argument of Cushman [1978] and Koslov [1984] (and reference therein) uses the Poincaré-Melnikov method to prove that the separatrices do split for most parameter values. However, it is not easy to prove from these arguments that splittings really do occur for specific parameter values and what the sharp upper and lower estimates for the splitting distances are. The seriousness and significance of this difficulty was emphasized by Sanders [1982].

Exponentially fine phenomena appear to be prevalent in a number of situations beyond those discussed here and in the next section. For example:

1 The action appears to change by an exponentially small amount in adiabatic theory (see, for example, Lenard [1959], Meyer [1976], and Berry [1985] - see also Marsden, Montgomery and Ratiu [1988]). We expect that our techniques will be relevant for these problems.

2 The existence of breathers in the ϕ^4 model involves exponentially small phenomena (see Segur and Kruskal [1987]).

3 The growth of dendritic crystals also involves exponentially small phenomena (see Kruskal and Segur [1987]).

4 Various problems in critical phenomena in water waves also seem to involve these issues; cf. Hunter and Scheurle [1987].

5 Exponentially small phenomena are known to occur in the study of relaxation oscillations; cf., Eckhaus [1982]

6 Finally, it has been suspected for some time that these problems also arise in the unfolding of degenerate singularities; see for example, Takens [1974]. We shall illustrate the basic ideas in §7.

Here we consider a simple illustration of why these problems come up in KAM theory. Consider the dynamics of two coupled oscillators with Hamiltonian, written in action angle variables, of the form

$$\mathcal{H}(\theta, I, \phi, J, \varepsilon) = F(I) + J + \varepsilon K(\theta, I, \phi) . \tag{6.1}$$

We have taken the second oscillator to be a harmonic oscillator and the coupling independent of J purely for simplicity. If we set $H = \text{constant}$, (3.1) determines J. We can also let $\phi = t$ be the new time, so (3.1) becomes equivalent to a forced one degree of freedom system with Hamiltonian

$$H(\theta, I, t, \varepsilon) = F(I) + \varepsilon K(\theta, I, t) . \tag{6.2}$$

For example, choose $K(\theta, I, \phi) = I \sin^2\theta \cos \phi$ and $F(I) = I - I^2/2$. Then one sees that the circle $I = 1/4$ resonates with the forcing in a 2:1 resonance. To study it, we make the change of variables

$$I = \frac{1}{4} - \sqrt{\varepsilon}\, p , \quad \theta = \frac{t}{2} + \psi \tag{6.3}$$

to get

$$\begin{aligned} \dot{\psi} &= \sqrt{\varepsilon}\,(2p) + \varepsilon \left[\frac{1}{4}\cos 2\psi + \cos 2(\psi + t) - \frac{1}{2}\cos t\right] , \\ \dot{p} &= \sqrt{\varepsilon}\left[-\frac{1}{8}\sin 2\psi + \sin 2(\psi + t)\right] + \varepsilon\left[\frac{p}{2}\sin 2\psi + \sin 2(\psi + t)\right] \end{aligned} \tag{6.4}$$

Now one removes the t-dependence at order $\sqrt{\varepsilon}$ by the averaging transformation

$$\psi = \psi' , \quad p = p' - \frac{\sqrt{\varepsilon}}{16} \cos 2(\psi' + t) . \tag{6.5}$$

Dropping the primes, the new system becomes Hamiltonian with

$$H = \sqrt{\varepsilon}F(\psi, p) + \varepsilon G(\psi, p, t) \ , \tag{6.6}$$

where

$$F(\psi, p) = p^2 - \frac{1}{16}\cos 2\psi \ ,$$

and

$$G(\psi, p, t, \varepsilon) = \frac{p}{8}\,[2\cos 2\psi + 3\cos 2(\psi + t) - 4\cos t] + O(\varepsilon^{1/2}) \ .$$

Rescaling time to $\tau = \sqrt{\varepsilon}\ t$, (6.6) transforms to

$$H = F(\psi, p) + \sqrt{\varepsilon}G\left(\psi, p, \frac{\tau}{\sqrt{\varepsilon}}\right) \tag{6.7}$$

which has our form of a rapidly forced perturabation of the Hamiltonian F, which has homoclinic orbits with

$$p = \pm\frac{1}{2\sqrt{2}}\,\text{sech}\,(4\sqrt{2}\tau) \ . \tag{6.8}$$

The situation is shown in Figure 5, represented in (I, θ) coordinates, viewed as polar coordinates. With the addition of $\sqrt{\varepsilon}\,G(\psi, p, t/\sqrt{\varepsilon})$, one develops stochastic layers around the homoclinic orbits shown in figure 5.

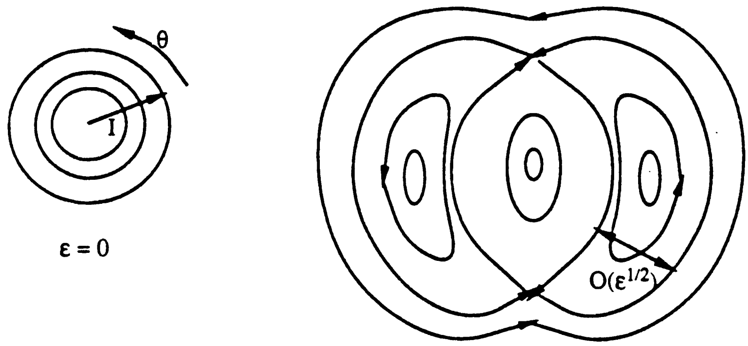


Figure 5

A *formal* Melnikov calculation suggests the splitting distance is of order

$$\frac{\pi}{256\sqrt{\varepsilon}}\,\mathrm{sech}\left(\frac{\pi}{8\sqrt{2\varepsilon}}\right), \tag{6.9}$$

which is exponentially small. Our upper estimate shows that

$$\text{splitting} \le C\sqrt{\varepsilon}\,\exp\left(-\left(\frac{\pi}{8\sqrt{2}}-\eta\right)\frac{1}{\sqrt{\varepsilon}}\right) \tag{6.10}$$

with a similar estimate for the splitting angle. Note that (6.10) is compatible with (6.9), although (6.9) suggests a sharper result. Our lower estimates do not apply to (6.7) since the *same* power of ε appears as an amplitude coefficient in front of G and also as the denominator of $\tau/\sqrt{\varepsilon}$. Our analysis of the estimates suggests that it may be very difficult to rigorously establish an estimate above and below by an expression like (6.9). However, one can show the following: Consider the same system with an additional term:

$$H = I - \frac{I^2}{2} + J + \varepsilon I \sin^2\theta \cos\phi + \varepsilon^2 H_2 . \tag{6.11}$$

We choose H_2 such that after averaging, terms of lower order in ε cancel, leaving only higher order terms in ε and our lower bound now *does* apply. The algebra involved to get H_2 is a little involved, so we illustrate the result with a simpler explicit example. Consider

$$\ddot{\phi} + \sin\phi = \varepsilon \sin\left(\frac{t}{\varepsilon}\right) . \tag{6.12}$$

Again, the upper bound $C\varepsilon e^{-(\pi/2 - \eta)/\varepsilon}$ is valid, but an optimal upper and lower estimate are not known. However, we can modify (6.12) a bit to

$$\ddot{\phi} + \sin\phi = \varepsilon \sin\left(\frac{t}{\varepsilon}\right) + \varepsilon^2 h\left(\frac{t}{\varepsilon}, \varepsilon, \phi\right) , \tag{6.13}$$

where

$$h(\tau, \varepsilon, \varphi) = \varepsilon \cos\varphi \sin\tau + \frac{\varepsilon^4}{2} \sin\varphi \sin^2\tau - \frac{\varepsilon^7}{3} \cos\varphi \sin^3\tau \tag{6.14}$$

so that (6.13) satisfies

$$C_2 \varepsilon^{12} e^{-\pi/2\varepsilon} \le \text{splitting} \le C_1 \varepsilon^{12} \varepsilon^{-\pi/2\varepsilon} . \tag{6.15}$$

This is done by choosing K so that after averaging, the system has the form required for both our upper and lower estimates.

Thus, while we cannot prove the upper and lower estimates for (6.12), there is a nearby system (6.13) for which they are valid. We conclude that while the upper estimates are fairly robust, the lower estimates appear to be very delicate and in fact one can perturb a given system slightly to get a splitting distance (and angle) ***much smaller*** than one might have expected -- see the extra power of ε^{12} in (6.14). Even more extreme, one can sometimes add a term which completely cancels all the higher order terms and the perturbed system becomes completely integrable! For instance, a trivial example of this sort is the completely integrable system

$$\left.\begin{aligned} \dot{x} &= y - \varepsilon^2 \sin\left(\frac{t}{\varepsilon}\right) \\ \dot{y} &= \sin x + \varepsilon \cos\left(\frac{t}{\varepsilon}\right) \end{aligned}\right\} \tag{6.16}$$

which is simply a complicated way of writing the pendulum equation. This general behavior appears to be rather common and shows that an asymptotic estimate good enough to give lower bounds independent of (or robust with respect to) all higher order terms is *not possible*.

§7 *Exponentially Small Splittings in a Bifurcation Problem*

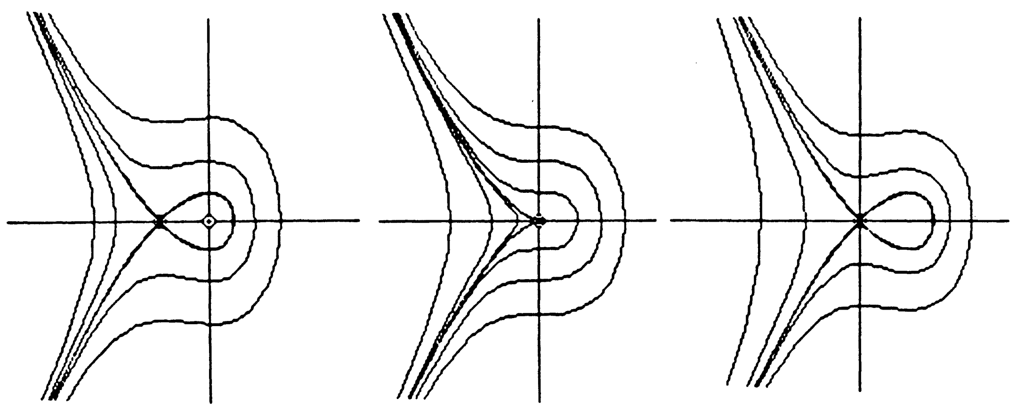
We consider the problem of a Hamiltonian saddle node bifurcation

$$\ddot{x} + \mu x + x^2 = 0 \tag{7.1}$$

with the addition of higher order terms and forcing:

$$\ddot{x} + \mu x + x^2 + \text{h.o.t.} = \delta f(t) \ . \tag{7.2}$$

The phase portrait of (7.1) is shown in Figure 6.



$\mu < 0$ $\mu = 0$ $\mu > 0$

Figure 6

The system (7.1) is Hamiltonian with

$$H(x, p) = \frac{1}{2} p^2 + \frac{1}{2} \mu x^2 + \frac{1}{3} x^3 . \tag{7.3}$$

Let us first consider the system without higher order terms:

$$\ddot{x} + \mu x + x^2 = \delta f(t) . \tag{7.4}$$

To study it, we rescale to blow up the singularity:

$$x(t) = \lambda \xi(\tau) \tag{7.5}$$

where $\lambda = |\mu|$ and $\tau = t\sqrt{\lambda}$. We get

$$\ddot{\xi} - \xi + \xi^2 = \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{-\mu}}\right) , \quad \mu < 0 , \tag{7.6a}$$

$$\ddot{\xi} + \xi + \xi^2 = \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{\mu}}\right) , \quad \mu > 0 . \tag{7.6b}$$

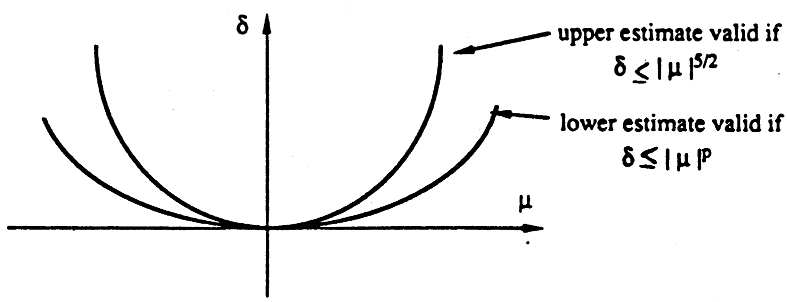


Figure 7

Our exponentially small estimates apply to (7.6a and b). We will get upper and lower estimates in algebraic sectors of the δ-μ plane, as in Figure 7. The power p depends on the nature of f.

Now we consider

$$\ddot{x} + \mu x + x^2 + x^3 = \delta f(t) . \tag{7.7}$$

With $\delta = 0$, there are equilibria at

$$\left.\begin{aligned} x &= 0, -r, \text{ or } -\frac{\mu}{r} \\ \dot{x} &= 0 \end{aligned}\right\} \tag{7.8a}$$

where

$$r = \frac{1 + \sqrt{1 - 4\mu}}{2} , \tag{7.8b}$$

which is approximately 1 when $\mu \approx 0$. The phase portrait of equation (7.7) with $\delta = 0$ and $\mu = -\frac{1}{2}$ is shown in Figure 8. As μ passes through 0, the small lobe in Figure 8 undergoes the same bifurcation as in Figure 5, with the large lobe changing only slightly.

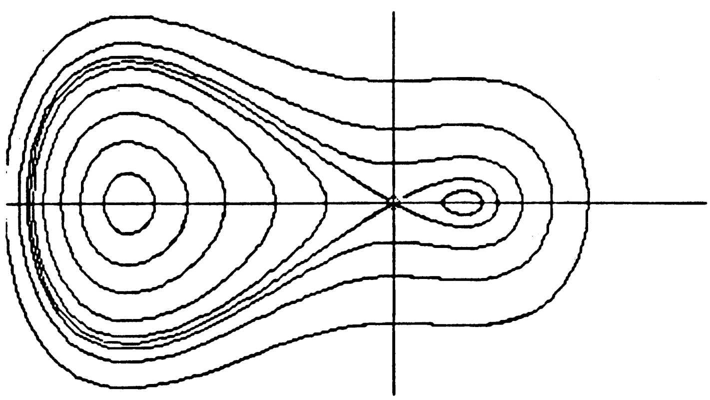


Figure 8

Again we rescale by (7.5) to give

$$\ddot{\xi} - \xi + \xi^2 - \mu\xi^3 = \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{-\mu}}\right) , \quad \mu < 0 , \tag{7.9a}$$

$$\ddot{\xi} + \xi + \xi^2 + \mu\xi^3 = \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{\mu}}\right) , \quad \mu > 0 . \tag{7.9b}$$

Notice that for $\delta = 0$, the phase portrait is μ-dependent. The homoclinic orbit surrounding the small lobe for $\mu < 0$ is given explicitly in terms of ξ by

$$\xi(\tau) = \frac{4e^{\tau}}{\left(e^{\tau} + \frac{2}{3}\right)^2 - 2\mu} , \tag{7.10}$$

which is μ-dependent. An interesting technicality is that without the cubic term, we get μ-independent double poles at $\tau = \pm i\pi + \log 2 - \log 3$ in the complex τ-plane, while (7.10) has a pair of simple poles that splits these double poles to the pairs of simple poles at

$$\tau = \pm i\pi + \log\left(\frac{2}{3} \pm i\sqrt{2\lambda}\right) \tag{7.11}$$

where again $\lambda = |\mu|$. (There is no particular significance to the real part, such as $\log 2 - \log 3$ in the case of no cubic term, since this can always be gotten rid of by a shift in the base point $\xi(0)$.)

If a quartic term x^4 is added, these pairs of simple poles will split into quartets of branch points and so on. Thus, while the analysis of higher order terms has this interesting μ-dependence, it seems that the basic exponential part of the estimates,

$$\exp\left(-\frac{\pi}{\sqrt{|\mu|}}\right) , \tag{7.12}$$

remains intact. ✍

Discussion and Conclusions We have given conditions under which one can obtain transversal intersection and both upper and lower estimates for the angle of intersection and for the splitting distance of separatrices in a rapidly forced system with a homoclinic orbit; the bounds obtained are exponentially small in the frequency parameter. Our main example is the rapidly forced pendulum equation, which is related to the pendulum suspended from a very stiff elastic rod. This example is a nonautonomous conservative system with a homoclinic orbit. With the addition of damping, exponentially small splitting and intersections of the separatrices typically occur only in an exponentially small wedge in parameter space (see Remark 2 above). Exponentially small splittings also occur in the unfolding of degenerate singularities and in KAM theory as was discussed in the last two sections. In future work we shall be applying these ideas to other problems, including sharp exponential estimates for adiabatic invariants of the sort that occur in Berry's phase. ♦

References

V. I. Arnold [1965] Sur une propriètè topologique des applications globelement canonique de la méchanique classique. *C. R. Acad. Sci. Paris.* **26**, 3719-3722.

V. I. Arnold [1978] *Mathematical Methods of Classical Mechanics,* Graduate Texts in Mathematics **60**, Springer-Verlag, NY.

M. Berry [1985] Classical Adiabatic Angles and Quantal Adiabatic Phase, *J. Phys. A: Math. Gen.* **18**, 15-27

T. Bountis, V. Papageorgiou, and M. Bier [1986] On the singularity analysis of intersecting separatrices in near-integrable dynamical systems, *Physica D (to appear).*

B. Chirikov [1979] A universal instability of many dimensional oscillator systems, *Phys. Reports,* **52**, 265.

R. Cushman [1978] Examples of non-integrable analytic Hamiltonian vector fields with no small divisors, *Trans. Am. Math. Soc.* **238**, 45-55.

G. Dangelmayr and E. Knobloch [1987] On the Hopf Bifurcation with Broken O(2) Symmetry, *preprint.*

W. Eckhaus [1982] Relaxation Oscillations, Including a Standard Chase on French Ducks, in *Asymptotic Analysis II*, ed. by F. Verhulst, Springer Lecture Notes in Mathematics, **985,** 449-494.

M. Golubitsky and I. Stuart [1986] Symmetry and stability in Taylor Couette flow, SIAM J. Math. Anal. **17**, 249-288.

J. Guckenheimer and P.J. Holmes [1983] Nonlinear oscillations, dynamical systems, and the bifurcation of vector fields, *Applied Mathematical Sciences, Springer-Verlag, N.Y.*

P. Hartman [1982] *Ordinary Differential Equations*, Wiley [1964], Second Edition, Birkhäuser.

P.J. Holmes and J.E. Marsden [1981] A partial differential equation with infinitely many periodic orbits: Chaotic oscillations of a forced beam, *Arch. for Rational Mech. and Analysis* **76**, 135-166.

P.J. Holmes and J.E. Marsden [1982a] Horseshoes in perturbations of Hamiltonian systems with two degrees of freedom, *Commun. Math. Physics* **82**, 523-544.

P.J. Holmes and J.E. Marsden [1982b] Melnikov's method and Arnold diffusion for perturbations of integrable Hamiltonian systems, *J. Math. Phys.* **23** , 669-675.

P.J. Holmes and J.E. Marsden [1983] Horseshoes and Arnold Diffusion for Hamiltonian Systems on Lie Groups, *Indiana Univ. Math. J.* **32**, 273-310.

P.J. Holmes, J.E. Marsden, and J. Scheurle [1988] Exponentially Small Splitting of Separatrices, *Proc. Nat. Acad. Sci.* (to appear).

J. Hunter and J. Scheurle [1987] *preprint*

V.V. Koslov [1984] Integrability and non-integrability in Hamiltonian mechanics, *Russian Math. Surveys*, **38**, 1-76.

M. Kruskal and H. Segur [1987] Asymptotics beyond all orders in a model of dendritic crystals, *preprint.*

V. Lazutkin, I.G. Shachmanski and M.B. Tabanov [1987], Separatrix splitting for the standard and semi-standard mappings, *preprint.*

A. Lenard [1959] Adiabatic Invariance to all Orders, *Ann. of Phys.* **6**, 261-276

J. Marsden, R. Montgomery and T. Ratiu [1988], Reduction, Symmetry, and Berry's Phase in Classical Machanics: Theory,*preprint.*

V.K. Melnikov [1963] On the Stability of the Center for Time Periodic Perturbations, *Trans. Moscow Math. Soc.* **12**, 1-57

R.E. Meyer [1976] Adiabatic Variation - Part V: Nonliear Near-Periodic Oscillator, J. Appl. Math. and Physics (ZAMP), **27**, 181-195.

R.E. Meyer [1982] Wave Reflection and Quasiresonance, in *Theory and Application of Singular Perturbations*, W. Eckhaus and E. M. de Jager, eds., Springer Lecture Notes in Mathematics, **942**, 84-112.

N.N. Nekhoroshev [1971] The behavior of hamiltonian systems that are close to integrable ones, *Funct. An. Appl.* **5**, 4-10.

N.N. Nekhoroshev [1977] An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems, *Russian Math. Surveys* **32**, 1-65.

A. Neishtadt [1984] The separation of motions in systems with rapidly rotating phase, *P.M.M. USSR*, **48**, 133-139.

H. Poincaré [1890] Sur la problème des trois corps et les équations de la dynamique, *Acta Math.* **13**, 1-271

P. van Moerbeke [1983] Algebraic complete integrability of Hamiltonian systems and Kac-Moody Lie algebras, *Proc. Intern. Cong. Mathem.* 881-899.

V. K. Melnikov [1963] On the stability of the center for time periodic perturbations, *Trans . Moscow Math. Soc.* **12**, 1-57.

J.A. Sanders [1982] Melnikov's method and averaging, *Celestial Mechanics*, **28**, 171-181.

J. Scheurle [1987] Asymptotic properties of Arnold tongues. *CMS Conference Series (AMS Publications)* **8**, 655-663

J. Scheurle and J. Marsden [1984] Bifurcation to quasi-periodic tori in the interaction of steady state and Hopf bifurcations, *SIAM J. Math. Anal.* **15**, 1055-1074.

H. Segur and M. Kruskal [1987] Non-existence of Small-Amplitude Breather Solutions in φ^4 Theory. *Phys. Rev. Lett.* **58**, 747-750

C. Simo and E. Fontich [1985] On the smallness of the angle between split separatrices, *Proc. Colloq. sur Geom. Symplectique et Mechanique, Montpellier.*

F. Takens [1974] Forced Oscillations and Bifurcations, *Comm. Math. Inst.., Rijkuniversiteit Utrecht*, **3**, 1-59.

E. Zehnder [1973] Homoclinic points near Elliptic fixed points, *Comm. Pure and Applied Math.* **26**, 131-182.

S. L. Ziglin [1982] Branching of solutions and nonexistence of first integrals in Hamiltonian mechanics, *Finct. Anal. Appl.* **16**, 30-41, **17**, 8-23.

Philip Holmes
Theoretical and Applied Mechanics and Mathematics
Cornell University
Ithaca, N.Y. 14853

Jerrold Marsden
Department of Mathematics
University of California
Berkeley, California, 94720

and

Department of Mathematics
Cornell Universtiy
Ithaca, N.Y. 14853-7901

Jurgen Scheurle
Institüt für Angew. Math.
Universitat Hamburg
Bundesstrasse 55
D-2000 Hamburg 13
West Germany

and

Department of Mathematics
Colorado State University
Fort Collins, CO 80523

http://dx.doi.org/10.1090/conm/081/986268

Contemporary Mathematics
Volume 81, 1988

THE THREE POINT VORTEX PROBLEM: COMMUTATIVE AND NON-COMMUTATIVE INTEGRABILITY

Malcolm Adams and Tudor Ratiu

ABSTRACT. The relationship between the usual concept of complete integrability (commutative integrability) and generic zero dimensional reduced phase space (non-commutative integrability) is investigated in the context of dual pairs introduced by Weinstein [1983]. The abstract theorem is illustrated by the example of the three point vortex motion.

1. INTRODUCTION

If S is a symplectic $2n$-manifold and $f_1, \ldots, f_n : S \to \boldsymbol{R}$ is a set of smooth functions which are in involution, i.e., $\{f_i, f_j\} = 0$ for all $i, j = 1, \ldots, n$, and independent, i.e., the covectors $df_1(x), \ldots, df_n(x)$ are linearly independent in T_x^*S for almost all $x \in S$, then $f_1, \ldots, f_n$ is *called a completely integrable* family of functions. If G is a Lie group acting on S in a Hamiltonian manner with momentum map $J : S \to \boldsymbol{g}^*$, $\boldsymbol{g}$ being the Lie algebra of G and $\boldsymbol{g}^*$ its dual, and if the reduced phase spaces $S_\mu := S^{-1}(\mu)/G_\mu$, $G_\mu = \{\mu \in \boldsymbol{g}^* \mid Ad_g^* \mu = \mu\}$, are zero dimensional for almost all $\mu \in \boldsymbol{g}^*$, the momentum map J is called *noncommutatively integrable.* If $\boldsymbol{F}$ is the ring of real valued functions generated by the components of J, then $\boldsymbol{F}$ is called *non-commutatively integrable* if J is non-commutatively integrable. The reason for this terminology is the following. Assume that the Hamiltonian vector fields $X_{f_1}, \ldots, X_{f_n}$ for $f_1, \ldots, f_n$ a completely integrable family of functions are complete, i.e., their flows $\phi_{t_1}^1, \ldots, \phi_{t_n}^n$ exist for all times $t_1, \ldots, t_n \in \boldsymbol{R}$. Then $\boldsymbol{R}^n$ acts on S by $(t_1, \ldots, t_n) \cdot x = (\phi_{t_1}^1 \circ \cdots \circ \phi_{t_n}^n)(x)$ and the order in which the flows are applied does not matter by involutivity. This action is Hamiltonian and has momentum map $J = f_1 \times \cdots \times f_n : S \to \boldsymbol{R}^n$. By generic independence of the differentials, the reduced manifolds S_μ, $\mu \in \boldsymbol{R}^n$, exist for almost all μ and their dimension is zero. Thus complete

1980 *Mathematics Subject Classification* (1985 *Revision*) 50 FO5, 58F10, 70K20

MA partially supported by an NSF grant. TR partially supported by an A.P. Sloan Foundation fellowship.

integrability is just non-commutative integrability for an $\mathbf{R}^n$-momentum map and is therefore also called *commutative integrability*. Mishchenko and Fomenko [1978b] have shown that under certain hypotheses on $\mathbf{g}^*$ non-commutative integrability implies commutative integrability, see also Guillemin and Sternberg [1983a,b].

The purpose of this paper is to generalize this result in the context of dual pairs introduced by Weinstein [1983] and to illustrate the theory with the classical example of the three point vortex motion. Section 2 reviews the N-point vortex motion and its four integrals from the point of view of symplectic geometry and momentum maps. It also shows, for the case $N=3$, that a certain combination of these four non-commuting integrals give a family of three commuting generically independent integrals, thus showing that the three point vortex motion is completely integrable; this recalls classical work as reviewed in e.g., Aref and Pomphrey [1982]. Section 3 analyzes the Poisson manifold which appears naturally as the reduction of the phase space of the three point vortex problem by the special Euclidean group in two dimensions. Section 4 discussess the dual pairs of Weinstein and proves a theorem about the pull-backs of integrable families of functions on Poisson manifolds. As a corollary it is shown that if the dual pair is $\mathbf{g}^* \leftarrow S \rightarrow S/G$, then non-commutative integrability implies commutative integrability if $\mathbf{g}^*$ admits an integrable system; this reproduces the result of Mishchenko and Fomenko [1978b]. Section 5 returns to the three point vortex problem, applies the theory of section 4, and deduces the classical commuting integrals.

2. THE N-POINT VORTEX PROBLEM

The motion of N point vortices for an ideal inviscid incompressible fluid in the plane is given by the equations

$$\left\{\begin{aligned} \frac{dx_j}{dt} &= -\frac{1}{2\pi}\sum_{\substack{i=1\\ i\neq j}}^{N}\Gamma_i(y_j-y_i)/r_{ij}^2 \\ \frac{dy_j}{dt} &= \frac{1}{2\pi}\sum_{\substack{i=1\\ i\neq j}}^{N}\Gamma_i(x_i-x_j)/r_{ij}^2 \end{aligned}\right. \tag{2.1}$$

where $r_{ij}^2=(x_i-x_j)^2+(y_i-y_j)^2$ and $\Gamma_1,\ldots,\Gamma_N$ are N non-zero constants, the circulations given by the corresponding point vortices; see Chorin and Marsden [1979]. Kirchhoff [1883] noted that (2.1) can be written in the form

$$\Gamma_j\frac{dx_j}{dt}=\frac{\partial H}{\partial y_j},\quad \Gamma_j\frac{dy_j}{dt}=-\frac{\partial H}{\partial x_j} \tag{2.2}$$

where

$$H(x_1, \ldots, x_N, y_1, \ldots, y_N) = -\frac{1}{4\pi} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \Gamma_i \Gamma_j \log r_{ij}, \tag{2.3}$$

which is a Hamiltonian system relative to the symplectic form

$$\Omega = \sum_{i=1}^{n} \Gamma_i \, \mathbf{d}x_i \wedge \mathbf{d}y_i . \tag{2.4}$$

To determine the constants of the motion of this system, note that the special Euclidean group $SE(2) = \{(\mathbf{A}, \mathbf{a}) \mid \mathbf{A} \in SO(2) = S^1,\ \mathbf{a} \in \boldsymbol{R}^2\}$ with multiplication

$$(\mathbf{A}, \mathbf{a})(\mathbf{B}, \mathbf{b}) = (\mathbf{AB}, \mathbf{Ab} + \mathbf{a}), \tag{2.5}$$

identity $(\mathbf{I}, \mathbf{0})$ and inverse $(\mathbf{A}, \mathbf{a})^{-1} = (\mathbf{A}^{-1}, -\mathbf{A}^{-1}\mathbf{a})$ acts on $\boldsymbol{R}^2$ by

$$(\mathbf{A}, \mathbf{a}) \cdot \mathbf{z} = \mathbf{Az} + \mathbf{a}. \tag{2.6}$$

The Lie algebra of $SE(2)$ is $se(2) = \{(\xi, \mathbf{u}) \mid \xi \in \boldsymbol{R}, \mathbf{u} \in \boldsymbol{R}^2\}$ with bracket

$$[(\xi, \mathbf{u}), (\eta, \mathbf{v})] = (0, (\eta u_2 - \xi v_2, \xi v_1 - \eta u_1)). \tag{2.7}$$

The action (2.6) is Hamiltonian relative to the symplectic form $\Gamma dx \wedge dy$, with momentum map given by

$$J(x, y) = (-\Gamma(x^2 + y^2)/2,\ \Gamma y,\ -\Gamma x) \tag{2.8}$$

where we identify $se(2)^*$ with $\boldsymbol{R}^3$ via the usual dot-product on $\boldsymbol{R}^3$. This momentum map is *not* equivariant. The $se(2)^*$-valued one-co-cycle of $SE(2)$ is given by (see Abraham and Marsden [1978], §4.2)

$$\sigma(\mathbf{A}, \mathbf{a}) := J((\mathbf{A}, \mathbf{a}) \cdot (x, y)) - Ad^*_{(\mathbf{A}, \mathbf{a})^{-1}}(J(x, y)) = \Gamma(-\|\mathbf{a}\|^2/2, a_2, -a_1), \tag{2.9}$$

and the $\boldsymbol{R}$-valued 2-cocycle of $se(2)$ is given by

$$\begin{aligned} \Sigma((\xi, \mathbf{u}), (\eta, \mathbf{v})) &:= \hat{J}([(\xi, \mathbf{u}), (\eta, \mathbf{v})]) - \{\hat{J}(\xi, \mathbf{u}), \hat{J}(\eta, \mathbf{v})\} \\ &= \Gamma(u_2 v_1 - u_1 v_2), \end{aligned} \tag{2.10}$$

where $\hat{J}(\xi, \mathbf{u}) : \boldsymbol{R}^2 \to \boldsymbol{R}$ is the function given by $(x, y) \mapsto J(x, y) \cdot (\xi, \mathbf{u})$.

Now let $SE(2)$ act diagonally on $(\boldsymbol{R}^2)^N$ so that the momentum map $\mathbf{J} : \boldsymbol{R}^{2N} \to \boldsymbol{R}^3$ relative to the symplectic form (2.4) is given by

$$\mathbf{J}(\mathbf{x}, \mathbf{y}) = \left(-\sum_{i=1}^{N} \Gamma_i (x_i^2 + y_i^2)/2,\ \ \sum_{i=1}^{N} \Gamma_i y_i,\ \ -\sum_{i=1}^{N} \Gamma_i x_i \right). \tag{2.11}$$

Since the Hamiltonian (2.3) is invariant under the $SE(2)$-action, $\mathbf{J}$ is a conserved quantity and thus the three components of **J** all Poisson commute with H . Let us denote as in Aref and Pomphrey [1982]

$$Q = \sum_{i=1}^{N} \Gamma_i x_i \tag{2.12}$$

$$P = \sum_{i=1}^{N} \Gamma_i y_i \tag{2.13}$$

$$L^2 = \sum_{i=1}^{N} \Gamma_i (x_i^2 + y_i^2) \tag{2.14}$$

so that the Poisson bracket relation between these integrals are

$$\{H, Q\} = \{H, P\} = \{H, L^2\} = 0$$

$$\{Q, P\} = \sum_{i=1}^{N} \Gamma_i,\ \{Q, L^2\} = 2P,\ \{P, L^2\} = -2Q,$$

which in turn imply

$$\{Q^2 + P^2, L^2\} = 0.$$

Consequently the three integrals H, L^2, $Q^2 + P^2$ are in involution. Since the gradients of these three functions are easily shown to be generically independent, it follows that for $N = 3$, the system of three point vortices is completely integrable.

The argument below will show that this system is also non-commutatively integrable. To begin with, note that the phase space of the three point vortex problem is not $\boldsymbol{R}^6$ but $S = \boldsymbol{R}^6 \setminus \Delta_{12} \cup \Delta_{13} \cup \Delta_{23}$, where $\Delta_{ij} = \{(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) \mid \mathbf{z}_i = \mathbf{z}_j\}$; this is because the self-interaction terms (which would give infinite energy) in the Hamiltonian (2.3) have been eliminated. Since the isotropy groups of the $SE(2)$-action (2.6) on $\boldsymbol{R}^2$ are

$$SE(2)_{\mathbf{0}} = SO(2), \quad SE(2)_{\mathbf{z}} = \{(\mathbf{I}, \mathbf{0}),\ (-\mathbf{I}, 2\mathbf{z})\}, \quad \text{for } \mathbf{z} \neq \mathbf{0},$$

this action is *not* free. However, the diagonal $SE(2)$-action on S is free. Properness of this action is easily verified so that $S/SE(2)$ is a smooth three dimensional Poisson manifold. The symplectic leaves, which are the reduced manifolds S_μ, are generically two-dimensional, so that the Hamiltonian (2.3) induces on them a real valued function. The resulting system is integrable having one degree of freedom. Thus the action given by the product of $SE(2)$ with the flow of H on S yields a non-commutatively integrable family given by H, Q, P, L^2. This approach to integrability of the three point vortex problem is symplectically more natural, whereas the combination $Q^2 + P^2$ making this problem commutatively integrable appears as an accident. We shall show in section 5 that this is not the case and that one is led naturally to such a combination.

3. THE REDUCED MANIFOLD $S/SE(2)$

In this section we will describe the Poisson geometry of the quotient manifold $S/SE(2)$ explicitly. Let us identify $\mathbb{R}^2$ with $\mathbb{C}$ by the usual map $(x, y) \mapsto x + iy = z$ so that the $SE(2)$-action on $\mathbb{C}$ is

$$(e^{i\theta}, a)\cdot z = e^{i\theta}z + a$$

for $\theta \in \mathbb{R}$, $a, z \in \mathbb{C}$. The map

$$(z_1, z_2, z_3) \in S \mapsto (z_2 - z_1, z_3 - z_1) \in \mathbb{C} \times \mathbb{C} \setminus Q,$$

where $Q = \{(u, 0) \mid u \in \mathbb{C}\} \cup \{(0, v) \mid v \in \mathbb{C}\} \cup \{(w, w) \mid w \in \mathbb{C}\}$, is onto and equivariant relative to the $SE(2)$-action on S and the diagonal S^1-action on $\mathbb{C} \times \mathbb{C} \setminus Q$. Now follow this map by the family of Hopf fibrations

$$(u, v) \in \mathbb{C} \times \mathbb{C} \setminus Q \mapsto (2u\bar{v}, |v|^2 - |u|^2) \in \mathbb{R}^3$$

giving the $SE(2)$-invariant surjective map $\phi : S \to T = \mathbb{R}^3 \setminus (\{(0, 0, c) \mid c \in \mathbb{R}\} \cup \{(a, 0, 0) \mid a \geq 0\})$

$$\phi(z_1, z_2, z_3) = ((z_2 - z_1)(\bar{z}_3 - \bar{z}_1), |z_3 - z_1|^2 - |z_2 - z_1|^2) \in T \tag{3.1}$$

whose fibers are the $SE(2)$-orbits on S. *Thus $S/SE(2)$ and T are diffeomorphic.*

The Poisson structure $\{\cdot,\cdot\}'$ of T induced by ϕ is determined by the relation

$$\phi^*\{F, H\}' = \{\phi^*F, \phi^*H\} \tag{3.2}$$

for any smooth functions $F, H : T \to \mathbb{R}$, where $\{\cdot,\cdot\}$ denotes the canonical Poisson bracket on S. Here is a sketch of the key steps in the computations. Denote by $\mathbf{a} \in T \subset \mathbb{R}^3$ the variable of F and H. Then if $z_j = x_j + iy_j$, $i = 1, 2, 3$, we have

$$a_1 = 2(x_2 - x_1)(x_3 - x_1) + 2(y_2 - y_1)(y_3 - y_1)$$

$$a_2 = 2(y_2 - y_1)(x_3 - x_1) - 2(x_2 - x_1)(y_3 - y_1)$$

$$a_3 = (x_3 - x_1)^2 + (y_3 - y_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2$$

so that the chain rule and the relations

$$(2x_1 - x_2 - x_3)(x_2 - x_3) - (2y_1 - y_2 - y_3)(y_3 - y_2)$$
$$+ (x_3 - x_1)^2 + (y_3 - y_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 = 2a_3$$

$$-(y_3 - y_2)^2 - (x_3 - x_2)^2 + (y_1 - y_3)(y_1 - y_2)$$
$$-(x_3 - x_1)(x_1 - x_2) + (y_2 - y_1)(y_3 - y_1) - (x_1 - x_2)(x_3 - x_1) = 2a_1 - \|\mathbf{a}\|^2$$

$$-(2x_1-x_2-x_3)(y_2-y_3)+(2y_1-y_2-y_3)(x_2-x_3)$$
$$-(x_3-x_1)(y_1-y_2)+(y_3-y_1)(x_1-x_2)-(x_2-x_1)(y_3-y_1)$$
$$+(y_2-y_1)(x_3-x_1)=2a_2$$

yield the expression

$$\begin{aligned}\{F,H\}'(\mathbf{a})=\phi^*\{F,H\}'(z_1,z_2,z_3)&=8a_3\left[\frac{\partial F}{\partial a_1}\frac{\partial H}{\partial a_2}-\frac{\partial F}{\partial a_2}\frac{\partial H}{\partial a_1}\right]\\&+8a_2\left[\frac{\partial F}{\partial a_3}\frac{\partial H}{\partial a_1}-\frac{\partial F}{\partial a_1}\frac{\partial H}{\partial a_3}\right]\\&+8a_1\left[\frac{\partial F}{\partial a_2}\frac{\partial H}{\partial a_3}-\frac{\partial F}{\partial a_3}\frac{\partial H}{\partial a_2}\right]\\&-4\|\mathbf{a}\|\left[\frac{\partial F}{\partial a_2}\frac{\partial H}{\partial a_3}-\frac{\partial F}{\partial a_3}\frac{\partial H}{\partial a_2}\right]\\&=8\mathbf{a}\cdot(\nabla F\times\nabla H)(\mathbf{a})-4\|\mathbf{a}\|(\nabla F\times\nabla H)_1.\end{aligned}\tag{3.3}$$

Thus the matrix of the Poisson bracket is

$$4\begin{bmatrix}0 & 2a_3 & -2a_2\\ -2a_3 & 0 & 2a_1-\|\mathbf{a}\|\\ 2a_2 & -2a_1+\|\mathbf{a}\| & 0\end{bmatrix}.\tag{3.4}$$

The zero dimensional leaves are given by the points where all (2×2)-minors of (3.4) vanish, which is easily seen to imply $\mathbf{a}=\mathbf{0}$. The Casimir function generating the center of the Poisson bracket $\{\cdot,\cdot\}'$ is given by

$$C(\mathbf{a})=\|\mathbf{a}\|-a_1/2\tag{3.5}$$

so that the two-dimensional leaves are the level sets of C, i.e., $\{\mathbf{a}\in T\mid C(\mathbf{a})=c,\ c\neq0\}$, or

$$\frac{\left[a_1-\dfrac{2c}{3}\right]^2}{16\dfrac{c^2}{9}}+\frac{a_2^2}{\dfrac{4c^2}{3}}+\frac{a_3^2}{\dfrac{4c^2}{3}}=1,\quad a_1+2c>0,\tag{3.6}$$

which are ellipsoids of revolution about the a_1-axis centered at $(2c/3,0,0)$. If $c>0$ and (a_1,a_2,a_3) is on the ellipsoid, then $-2c/3\le a_1\le 2c$ and the inequality $a_1+2c>0$ always holds. If $c<0$ and (a_1,a_2,a_3) is on the ellipsoid, then $2c\le a_1\le -2c/3$ so that the inequality $a_1+2c>0$ never holds. Thus the two-dimensional symplectic leaves are given by (3.6) for $c>0$ only. Out of these ellipsoids three points are missing: the intersection with the a_3-axis and with the positive a_1-axis, i.e., the points $(0, 0, \pm c)$ and $(2c, 0, 0)$.

Let us summarize the results of this section in the following.

PROPOSITION. *The map $\phi : S \to T$ given by (3.1), where T is R^3 minus the a_3-axis and the positive a_1-axis, is a surjective submersion whose fibers are the $SE(2)$-orbits on S and so $S/SE(2)$ is Poisson isomorphic to T with bracket (3.3). The Casimir functions of this Poisson bracket are generated by (3.5). The symplectic leaves are ellipsoids of revolution about the a_1-axis in R^3 minus the points $(0, 0, \pm c)$, $(2c, 0, 0)$, centered at $(2c/3, 0, 0)$ and having semiaxes equal to $4c/3$, $2c/\sqrt{3}$, $2c/\sqrt{3}$ for all $c > 0$.*

4. DUAL PAIRS AND INTEGRABILITY

We begin with the concept of integrability on Poisson manifolds. Let P be a Poisson manifold, $\dim P = 2n + k$, where $2n$ is the dimension of the maximal (generic) symplectic leaf.

DEFINITION *A ring of functions F on P is said to be* ***integrable****, if F is generated as a ring by $n + k$ functions and all functions in F Poisson commute.*

Next, let us recall the definition and a few key facts about dual pairs as introduced by Weinstein [1983].

DEFINITION. *Let (S, Ω) be a symplectic manifold, P_1, P_2 Poisson manifolds, and $J_i : S \to P_i$, $i = 1, 2$ Poisson maps. If for almost all $x \in S$, $(\ker T_xJ_1)^\Omega = \ker T_xJ_2$, the diagram $P_1 \xleftarrow{J_1} S \xrightarrow{J_2} P_2$ is called a* ***dual pair****; here $T_xJ_i : T_xS \to T_{J_i(x)}P_i$, $i = 1, 2$, denotes the derivative (tangent map) of J_i and $(\ker T_xJ_1)^\Omega = \{v \in T_xS \mid \Omega(x)(v, \ker T_xJ_1) \equiv 0\}$ is the Ω-orthogonal complement of $\ker T_xJ_1$ in T_xS. The dual pair is called* ***full****, if J_1, J_2 are surjective submersions.*

The following proposition is due to Weinstein [1983].

PROPOSITION. *Let $P_1 \xleftarrow{J_1} (S, \Omega) \xrightarrow{J_2} P_2$ be a full dual pair. Denote by $Cas(P_i)$, $i = 1, 2$, the space of Casimir functions on P_i and by Φ_i the space of functions on S constant on the fibers of J_i. Let $\dim P_i = 2n_i + k_i$, $i = 1, 2$, where $2n_i$ is the dimension of the maximal (generic) leaf of P_i. Then the following hold:*

(i) $\Phi_1 \cap \Phi_2 = Cas(P_1) \circ J_1 = Cas(P_2) \circ J_2$, *i.e., the spaces of Casimir functions on P_1 and P_2 are in bijective correspondence with $\Phi_1 \cap \Phi_2$.*

(ii) *For $p_1 \in P_1$ each connected component of $J_1^{-1}(p_1)$ maps under J_2 into the same symplectic leaf of P_2 and vice versa.*

PROOF. (i) If $H \in \Phi_1 \cap \Phi_2$, then $H = h_i \circ J_i$, $i = 1, 2$, for some $h_i : P_i \to R$. If Y is tangent to the fibers of J_2, i.e., Y is a section of $ker\ TJ_2$, then

$$\Omega(X_H, Y) = \langle dH, Y \rangle = 0, \tag{4.1}$$

which says that $X_H \in (ker\ TJ_2)^{\Omega} = ker\ TJ_1$ by the dual pair hypothesis. J_1 being a Poisson map we get

$$0 = TJ_1 \circ X_H = TJ_1 \circ X_{h_1 \circ J_1} = X^1_{h_1} \circ J_1, \tag{4.2}$$

where $X^1_{h_1}$ denotes the Hamiltonian vector field on P_1 defined by h_1. Since J_1 is surjective, it follows that $X^1_{h_1} = 0$, i.e., $h_1 \in Cas(P_1)$. We have shown that $\Phi_1 \cap \Phi_2 \subseteq Cas(P_1) \circ J_1$.

Conversely, if $h_1 \in Cas(P_1)$, then $H = h_1 \circ J_1 \in \Phi_1$ satisfies $TJ_1 \circ X_H = 0$ by (3.1), i.e., X_H is a section of $ker\ TJ_1 = (ker\ TJ_2)^{\Omega}$ and so for any section Y of $ker\ TJ_2$ we have $\langle dH, Y \rangle = 0$ by (3.1). This says that H is constant on the fibers of J_2, i.e., $H \in \Phi_1 \cap \Phi_2$ and we proved that $Cas(P_1) \circ J_1 \subseteq \Phi_1 \cap \Phi_2$. An identical argument proves the equality $Cas(P_2) \circ J_2 = \Phi_1 \cap \Phi_2$.

(ii) Let N be a connected component of $J_1^{-1}(p_1)$. We shall prove that the connected set $J_2(N)$ is a Poisson submanifold of P_2 which is symplectic. Since the symplectic leaves of P_2 are the maximal connected manifolds characterized by this property, it follows that $J_2(N)$ is contained in a symplectic leaf.

The following argument shows that $J_2|N$ has constant rank. For $n \in N$, $ker(T_nJ_2|T_nN) = ker\ T_nJ_2 \cap T_nN = ker\ T_nJ_2 \cap ker\ T_nJ_1$, since $T_nN = T_n(J_2^{-1}(p_1)) = ker\ T_nJ_1$. By the dual pair hypothesis, it follows then that $ker(T_nJ_2|T_nN) = ker\ T_nJ_2 \cap (ker\ T_nJ_2)^{\Omega}$, i.e., the dimension of the kernel of $T_nJ_2|T_nN$ equals the fiber dimension of the vector subbundle $ker\ TJ_2 \cap (ker\ TJ_2)^{\Omega}$ and so it is constant. Thus $J_2|N$ has constant rank. Since $J_2 : S \to P_2$ is a submersion, it is an open map and hence $J_2(N)$ is a submanifold of P_2 whose tangent bundle is $TJ_2(TN)$, by the rank theorem (see e.g., Abraham, Marsden, Ratiu [1983], theorem 3.5.18).

Since J_2 is Poisson, for $h : P_2 \to R$, $n \in N$, we have $X^2_h(J_2(n)) = T_nJ_2(X_{h \circ J_1}(n)) \in T_nJ_2(T_nN) = T_{J_2(n)}(J_2(N))$, where X^2_h denotes the Hamiltonian vector field on P_2 defined by h. Therefore $J_2(N)$ is a Poisson submanifold of P_2. In order to show it is symplectic, let $v \in T_{J_2(n)}(J_2(N))$, so that $v = T_nJ_2(w)$ for $w \in T_nN \subset T_nS$. Since S is symplectic, $w = X_F(n)$ for some $F : S \to R$ so that $v = T_nJ_2(X_F(n)) = X^2_{F \circ J_1}(n)$, i.e., any tangent vector to $J_2(N)$ is Hamiltonian in P_2 and therefore the Poisson manifold $J_2(N)$ has only one symplectic leaf, i.e., it is itself symplectic. Thus $J_2(N)$ is entirely contained in a symplectic leaf of P_2.

Conversely, let $L \subset P_2$ be a symplectic leaf, $p_2 \in L$, $s \in J_2^{-1}(p_2)$, and $p_1 = J_1(s)$. Let N be the connected component of $J_1^{-1}(p_1)$ containing s. Then $J_2(N)$ is contained in a symplectic leaf of P_2 by what we just proved and $p_2 = J_2(s) \in J_2(N)$ so that $L \supseteq J_2(N)$. ■

Proposition 1(ii) gives a correspondence between the symplectic leaves of P_1 and P_2 which is bijective locally and globally if J_1 and J_2 have connected fibers.

PROPOSITION 2. *Let P_i be Poisson manifolds $\dim P_i = 2n_i + k_i$, where $2n_i$ is the dimension of the generic symplectic leaf, $i = 1, 2$. Let $P_1 \xleftarrow{J_1} (S, \Omega) \xrightarrow{J_2} P_2$ be a full dual pair. Then $k_1 = k_2 = k$ and $\dim S = \dim P_1 + \dim P_2 = 2(n_1 + n_2 + k)$.*

PROOF. Recall that the generic symplectic leaf is a level set of the Casimir functions. So let $F_1, \ldots, F_{k_1} \in Cas(P_1)$ be such that $F_1 \times \cdots \times F_{k_1} : P_1 \to R^{k_1}$ has a regular value $\mathbf{x}$ for which $(F_1 \times \cdots \times F_{k_1})^{-1}(\mathbf{x})$ is a generic symplectic leaf in P_1. By Proposition 1(i), $F_i \circ J_1 = G_i \circ J_2$ for some $G_i \in Cas(P_2)$, $i = 1, \ldots, k_1$. Since both J_1 and J_2 are surjective submersions, $\mathbf{x}$ is also a regular value of $G_1 \times \cdots \times G_{k_1}$, which means that $(G_1 \times \cdots \times G_{k_1})^{-1}(\mathbf{x})$ contains a generic symplectic leaf of P_2. This says that $2n_2 + k_2 - k_1 \geq 2n_2$, whence $k_2 \geq k_1$. Reversing the roles of P_1 and P_2 we get the opposite inequality $k_1 \geq k_2$ and hence $k_1 = k_2$. If there are not enough global Casimirs, the argument above must be done locally, using Weinstein's Splitting Theorem [1983].

Finally, since J_1 and J_2 are submersions, we have $\dim S = \dim P_1 + \dim(\ker TJ_1) = \dim P_1 + \dim(\ker TJ_2)^\Omega = \dim P_1 + \dim S - \dim(\ker TJ_2) = \dim P_1 + \dim P_2 = 2(n_1 + n_2 + k)$. ■

PROPOSITION 3. *Let $P_1 \xleftarrow{J_1} (S, \Omega) \xrightarrow{J_2} P_2$ be a dual pair and $F : P_1 \to R$, $G : P_2 \to R$ be smooth functions. Then $\{F \circ J_1, G \circ J_2\} = 0$ on S.*

PROOF. If $s \in S$ and $v \in \ker T_s J_1$, then

$$\Omega(s)(X_{F \circ J_1}(s), v) = \mathbf{d}(F \circ J_1)(s) \cdot v = \mathbf{d}F(J_1(s)) \cdot (T_s J_1(v)) = 0$$

which shows that $X_{F \circ J_1}(s) \in (\ker T_s J_1)^\Omega = \ker T_s J_2$. Likewise $X_{G \circ J_2(s)} \in (\ker T_s J_2)^\Omega$ and hence $\{F \circ J_1, G \circ J_2\}(s) = \Omega(s)(X_{F \circ J_1}(s), X_{G \circ J_2}(s)) = 0$. ■

The main result of this paper is contained in the following.

THEOREM. *Let* $P_1 \xleftarrow{J_1} (S, \Omega) \xrightarrow{J_2} P_2$ *be a full dual pair and* F_i *be a completely integrable ring of functions on* P_i, $i = 1, 2$. *Let* F *be the ring of functions on* S *generated by* $J_1^* F_1 \cup J_2^* F_2$. *Then* F *is completely integrable on* S.

PROOF. By proposition 3, all functions in F Poisson commute. Let us show that F contains $n_1 + n_2 + k$ generically independent functions, where $dim\ P_i = 2n_i + k_i$, $2n_i$ being the dimension of the generic symplectic leaf of P_i and $k_1 = k_2 = k$ (see proposition 2). Let $f_1, \ldots, f_{n_1+k}$ be $n_1 + k$ generically independent functions in F_1 and let $g_1, \ldots, g_{n_2}$ be n_2 functions in F_2 which are generically independent on the generic symplectic leaves of P_2. Since J_1 and J_2 are surjective submersions, there is an open dense set $A \subset S$ such that for all $s \in A$, $\{\mathbf{d}(f_1 \circ J_1)(s), \ldots, \mathbf{d}(f_{n_1+k} \circ J_1(s))\}$ and $\{\mathbf{d}(g_2 \circ J_2)(s), \ldots, \mathbf{d}(g_{n_2} \circ J_2)(s)\}$ are linearly independent. Let V be the module of one-forms on A generated over the ring of functions on S by $\mathbf{d}(f_1 \circ J_1), \ldots, \mathbf{d}(f_{n_1+k} \circ J_1)$ and similarly let W be the module of one-forms on A generated over the ring of functions on S by $\mathbf{d}(g_1 \circ J_2), \ldots, \mathbf{d}(g_{n_2} \circ J_2)$. Then the associations $s \mapsto V(s)$, $s \mapsto W(s)$ define smooth subbundles of $T^*A = T^*S|A$, where $V(s)$, $W(s)$ denote the vector spaces obtained by evaluating each element of V and W, respectively at $s \in A$. By openness of non-intersection, the set $B = \{s \in A \mid V(s) \cap W(s) = \{0\}\}$ is open in A and hence in S. Let us show that B is also dense. If not, there would exist an open subset U of A and a one-form α on U such that $\alpha \in V \cap W$, $\alpha \neq 0$. This says, however, that $\alpha = \mathbf{d}(f \circ J_1) = \mathbf{d}(g \circ J_2)$ on U, for some $\mathbf{d}(f \circ J_1) \in V$ and $\mathbf{d}(g \circ J_2) \in W$. Therefore $\mathbf{d}(f \circ J_1 - g \circ J_2) \equiv 0$ in U and so adjusting g by adding a constant, it follows that $f \circ J_1 = g \circ J_2$ on U. In other words, $f \circ J_1 = g \circ J_2$ is a function on U which is constant on the fibers of J_1 and J_2 and hence by proposition 1(i), this function is a pull-back of a Casimir function on an open subset of P_2. Since J_2 is a surjective submersion, this would imply that $g|J_2(U)$ is a Casimir function on the open subset $J_2(U)$ of P_2; this is impossible, by the definition of W and A, thereby showing that B is dense in A and hence also in S. ■

COROLLARY 1. *Let* (S, Ω) *be a symplectic manifold on which* G *acts in a Hamilton manner with momentum map* $J : S \to \mathbf{g}^*$. *Suppose that* $\pi : S \to S/G$ *and* $J : S \to \mathbf{g}^*$ *are surjective submersions. Then* $S/G \xleftarrow{\pi} (S, \Omega) \xrightarrow{J} \mathbf{g}^*$ *is a full dual pair. If a* G*-invariant Hamiltonian system on* S *is non-commutatively integrable and if there is an integrable system on* $\mathbf{g}^*$, *then the original Hamiltonian system is commutatively integrable.*

PROOF. The first statement is a direct corollary of the reduction lemma; see Abraham and Marsden [1978], §4.3. The second statement is a direct consequence of the theorem. ■

COROLLARY 2. *In the hypothesis of Corollary 1, assume that the generic leaves of S/G are two-dimensional and that there is a completely integrable family of functions on g^*. Then any G-invariant Hamiltonian system on S is commutatively integrable.*

The second statement in Corollary 1 was first proved by Mishchenko and Fomenko [1978b]. Integrable systems on duals of Lie algebras appear for example from Euler equations or various involution theorems; if g is semisimple, such integrable families always exist. See Adler [1979], Kostant [1979], Mishchenko and Fomenko [1978a], Ratiu [1980], and Symes [1980] for a sample of such systems.

5. THE THREE POINT-VORTEX PROBLEM IN TERMS OF DUAL PAIRS

Recall from (2.11) that the momentum map of the $SE(2)$-action on S is given by

$$\mathbf{J}(\mathbf{x},\mathbf{y},\mathbf{z}) = (-(\Gamma_1 \|\mathbf{x}\|^2 + \Gamma_2 \|\mathbf{y}\|^2 + \Gamma_3 \|\mathbf{z}\|^2)/2,\ \Gamma_1 x_2 + \Gamma_2 y_2 + \Gamma_3 z_2,\ -\Gamma_1 x_1 - \Gamma_1 y_1 - \Gamma_3 z_1) \tag{5.1}$$

and that $\mathbf{J}$ is not equivariant. From (2.9) and (2.10) it follows that if we denote by $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3$, the group one-cocycle $\sigma : SE(2) \to se(2)^* = \boldsymbol{R}^3$ defined by $\mathbf{J}$ is given by

$$\sigma(e^{i\phi}, \lambda) = (-\Gamma \|\lambda\|^2/2, \Gamma\lambda_2, -\Gamma\lambda_1) \tag{5.2}$$

and the real-valued 2-cocycle $\Sigma : se(2) \times se(2) \to \boldsymbol{R}$ by

$$\Sigma((\xi, \lambda), (\eta, \mu)) = \Gamma(\lambda_2\mu_1 - \lambda_1\mu_2) \tag{5.3}$$

for $\lambda, \mu \in \boldsymbol{R}^2$, $\phi, \xi, \eta \in \boldsymbol{R}$. General theory says that $\mathbf{J}$ is equivariant relative to the Poisson bracket on $se(2)^*$ induced by the central extension of $se(2)$ by Σ (see Abraham and Marsden [1978], §4.2), i.e., relative to the bracket

$$\begin{aligned}\{f, h\}(\alpha, \lambda) &= \left\langle (\alpha, \lambda), \left[\frac{\delta f}{\delta(\alpha, \lambda)}, \frac{\delta h}{\delta(\alpha,\lambda)}\right]\right\rangle - \Sigma\left[\frac{\delta f}{\delta(\alpha, \lambda)}, \frac{\delta h}{\delta(\alpha,\lambda)}\right] \\ &= \lambda_1\left[\frac{\partial h}{\partial \alpha}\frac{\partial f}{\partial \lambda_2} - \frac{\partial f}{\partial \alpha}\frac{\partial h}{\partial \lambda_2}\right] + \lambda_2\left[\frac{\partial f}{\partial \alpha}\frac{\partial h}{\partial \lambda_1} - \frac{\partial h}{\partial \alpha}\frac{\partial f}{\partial \lambda_1}\right] \\ &\quad - \Gamma\left[\frac{\partial f}{\partial \lambda_2}\frac{\partial h}{\partial \lambda_1} - \frac{\partial f}{\partial \lambda_1}\frac{\partial h}{\partial \lambda_2}\right].\end{aligned} \tag{5.4}$$

It is easily seen that the Casimir functionals are arbitrary functions of

$$k(\alpha, \lambda) = \Gamma\alpha + \|\lambda\|^2/2 \tag{5.5}$$

so that the generic symplectic leaves are the circular paraboloids

$$\Gamma\alpha + \|\lambda\|^2/2 = \text{constant}. \tag{5.6}$$

Since the rank of the Poisson structure (5.4) is always two, there are no zero dimensional leaves. It is straightforward to check that the rank of (5.1) is 3 so that $\mathbf{J}$ is an open map. Let V be its open range endowed with the Poisson structure (5.4). We have proved the following.

PROPOSITION 4. *$T \overset{\phi}{\leftarrow} S \overset{\mathbf{J}}{\rightarrow} V$ is a full dual pair, where ϕ is given by (3.1), $\mathbf{J}$ by (5.1), T is endowed with the Poisson bracket (3.3) and V with the Poisson bracket (5.4).*

Let us now return to the three point vortex problem. The Hamiltonian H given by (2.3) is $SE(2)$-invariant and the leaves of T are all two-dimensional. To find an integrable system on V it suffices to take a function $f : V \to \mathbf{R}^3$ since the leaves of V are all two dimensional. Let us choose

$$f(\alpha, \lambda) = \alpha. \tag{5.6}$$

Thus f and the Casimir k given by (5.5) form an integrable family on V. By Corollary 2 in section 3, the Hamiltonian system given by H is commutatively integrable. Its commuting generically independent integrals are given by H, $f \circ \mathbf{J}$, $k \circ \mathbf{J}$ by the Theorem of section 3. Since

$$(f \circ \mathbf{J})(\mathbf{x}, \mathbf{y}, \mathbf{z}) = -(\Gamma_1\|\mathbf{x}\|^2 + \Gamma_2\|\mathbf{y}\| + \Gamma_3\|\mathbf{z}\|^2)/2 = -L^2/2, \quad \text{and}$$

$$\begin{aligned}(k \circ \mathbf{J})(\mathbf{x}, \mathbf{y}, \mathbf{z}) &= -\Gamma(\Gamma_1\|\mathbf{x}\|^2 + \Gamma_2\|\mathbf{y}\|^2 + \Gamma_3\|\mathbf{z}\|^2)/2 + [(\Gamma_1 x_1 + \Gamma_2 y_1 + \Gamma_3 z_1)^2 + (\Gamma_1 x_2 + \Gamma_2 y_2 + \Gamma_3 z_2)^2]/2 \\ &= -\Gamma L^2/2 + (Q^2 + P^2)/2\end{aligned}$$

it follows that H, L^2, $Q^2 + P^2$ are three generically independent commuting integrals, thus recovering the classical result.

BIBLIOGRAPHY

R. Abraham and J. Marsden, Foundations of Mechanics, Addison-Wesley, Reading, Mass., 1978.

R. Abraham, J. Marsden, and T. Ratiu, Manifolds, Tensor Analysis and Applications, Addison-Wesley, Reading, Mass., 1983.

M. Adler, "On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-de Vries type equations," Inventiones Math. 50 (1979), 219-248.

H. Aref and N. Pomphrey, "Integrable and chaotic motions of four vortices. I. The case of identical vortices," Proc. R. Soc. London A (1982) 359-387.

A. J. Chorin and J. Marsden, A Mathematical Introduction to Fluid Mechanics, Universitext, Springer-Verlag, 1979.

V. Guillemin and S. Sternberg, "On collective complete integrability according to a method of Thimm," Ergodic Theory and Dynamical Systems (1983a).

V. Guillemin and S. Sternberg, "Multiplicity free spaces," Journal Diff. Geom. (1983b).

G. Kirchhoff, Vorlesungen über mathematische Physik, Vol. 1, chapter 20, Teubner, Leipzig (1883).

B. Kostant, "The solution to a generalized Toda lattice and representation theory," Adv. in Math. 34 (1979), 195-338.

A. S. Mishchenko and A. T. Fomenko, "Euler equations on finite dimensional Lie groups," Math. USSR, Izvestija 12 (1978a), 371-389.

A. S. Mishchenko and A. T. Fomenko, "A generalized Liouville method for the integration of Hamiltonian systems," Funct. Anal. Appl. 12 (1978b), 113-131.

T. Ratiu, "Involution theorems," in Geometric Methods in Mathematical Physics (G. Kaiser and J. Marsden, eds.), Springer Lecture Notes 775 (1980), 219-257.

W. W. Symes, "Systems of Toda type, inverse spectral problems, and representation theory," Inventiones Math. 59 (1980), 13-51.

A. Weinstein, "The local structure of Poisson manifolds," Journ. Diff. Geom. 18 (1983), 523-557.

Malcolm Adams
Department of Mathematics
University of Georgia
Athens, GA 30602

Tudor Ratiu
Department of Mathematics,
University of California,
Santa Cruz, CA 95064

This page intentionally left blank

http://dx.doi.org/10.1090/conm/081/986269

Contemporary Mathematics
Volume **81**, 1988

ON A THEOREM OF ZIGLIN IN HAMILTONIAN DYNAMICS

David L. Rod[1]

ABSTRACT. This paper uses a theorem of Ziglin to demonstrate the non-integrability of a wide class of classical Hamiltonian systems with two degrees of freedom. The potential, when restricted to an axis, is a polynomial in one variable of degree 3 or 4. Certain geometric hypotheses and the general theory of elliptic functions allow us to bypass the detailed calculations of previous papers and obtain results on Hamiltonian systems involving many arbitrary parameters. An example of a Hamiltonian system is given whose flow is non-integrable at each energy $h > 0$, but for which the flow has a second independent real analytic (but not entire) integral at each $h < 0$.

INTRODUCTION. We work with the complex symplectic manifold (Σ,ω) where $\Sigma = \mathbb{C}^4 = \{(z_1,z_2,w_1,w_2)\}$ and the symplectic form $\omega = dz_1 \wedge dw_1 + dz_2 \wedge dw_2$. Let $\{\ ,\ \}$ be the Poisson bracket induced by ω, and for a holomorphic Hamiltonian $H\colon \Sigma \longrightarrow \mathbb{C}$, let X_H be the associated vectorfield. The image of a maximally continued non-equilibrium integral curve $\varphi = \varphi(t)$ with energy $h \in \mathbb{C}$ of X_H is a Riemann surface $\Gamma \subset \Sigma_h = H^{-1}(\{h\})$. The linearized equations along $\varphi(t)$ induce a linear differential equation on the (reduced) normal bundle $N = (T(\Sigma_h)|\Gamma)/T(\Gamma)$ of Γ called the (reduced) *normal variational equation* (NVE). Continuing a fixed fundamental system of solutions to NVE around inverses of loops based at $x_0 \in \Gamma$ gives the *monodromy group* of NVE as the image M of the representation $\rho\colon \pi_1(\Gamma,x_0) \longrightarrow S\ell(2,\mathbb{C})$ (direct continuation results in an antihomomorphism). $A \in M$ is *nonresonant* if no eigenvalue is a root of unity. We now give a version of a result of Ziglin [16], as formulated by Ito [9], that is appropriate for the applications of this paper.

1980 Mathematics Subject Classification (1985 Revision). 58F05.

[1]Author's research supported in part by the Natural Sciences and Engineering Research Council of Canada, Grant A8507.

ZIGLIN'S THEOREM. *Assume there is a meromorphic function* F *defined in some neighborhood* $U \subset \Sigma$ *of* Γ, *functionally independent of* H, *and satisfying* $\{F,H\} = 0$ in U. *Assume the monodromy group* M *contains a nonresonant element* A. *Then for any* $B \in M$ *the commutator* $(A,B) = B^{-1}A^{-1}BA$ *satisfies* $(A,B) = I$ (*identity*) *or* $(A,B) = A^2$, *with only* $(A,B) = I$ *possible if* B *does not admit* $\pm i$ *as eigenvalues. In particular,* X_H *has no meromorphic integral independent of* H *if there is a* $B \in M$ *such that* $I \neq (A,B) \neq A^2$.

In Ziglin [16] and Ito [9] the components of $\gamma(t)$ were elliptic functions, and Γ, therefore, had the topological structure of a punctured torus. This made the monodromy group M sufficiently computable for certain applications, but explicit calculations with these elliptic functions were still required to verify the nonresonance hypothesis. In the present paper such calculations are replaced by a simple geometric criterion which establishes the nonresonance (see also Ito [10] and Rod [13]). Then Ziglin's Theorem, with $\gamma(t)$ still elliptic, can be easily applied to a number of Hamiltonian systems having many parameters. A particular system with parameters (the Hénon-Heiles family) was first studied by Ito in [9], and this example was redone using more geometric methods in Ito's second paper [10]. In Section 2 we obtain these results as a corollary to Theorem 2.

In Section 1 a consequence of Moser's generalization of Liapunov's Theorem [11] is used to show how the eigenvalues of the linearized Poincaré map about a family of periodic orbits (parameterized by energy) change as these orbits limit into homoclinic/heteroclinic orbits. This will give us a set of energies at which the nonresonance hypothesis is satisfied. In Ito [10] the checking of this hypothesis is a consequence of an application of Siegel's general version of Liapunov's Theorem [15, Section 16] with results similar to the present paper. Sections 2 and 3 detail our applications to classical (kinetic plus potential energy) Hamiltonians with two degrees of freedom in the case that the potential, when restricted to an axis, is a polynomial in one variable of respective degree 3 or 4. These results are unaffected by adding an arbitrary term K to the Hamiltonian (see (1.1)) provided this term vanishes to order two when restricted to this axis.

In Remark 1(b) of Section 2 an example is given of a real analytic Hamiltonian whose flow at each energy $h > 0$ has a Smale horseshoe embedded in it, hence is not completely integrable at these energies, whereas at each $h < 0$ the flow has a second independent real analytic (but not entire) integral. This example illustrates the nature of the integrals that occur in the statement of Ziglin's Theorem.

For the basic facts about elliptic functions that we require in Sections 2 and 3, see [14].

1. EIGENVALUE CHANGE AND HOMOCLINIC/HETEROCLINIC ORBITS

In this section we establish notation and collect some basic facts required in Sections 2 and 3. The proof of Theorem 1 depends heavily on a number of results in [5] which follow from Moser's generalization of Liapunov's Theorem [11] as formulated in the two degree of freedom case by Conley in [7].

Consider on $(\mathbb{R}^4,\omega_0)$ the real analytic Hamiltonian

$$H(x_1,x_2,y_1,y_2) = (1/2)(y_1^2 + y_2^2) + W(x_1,x_2) + K(x_1,x_2,y_1,y_2), \tag{1.1}$$

where the function K is entire and vanishes to order two on the $x_2 = y_2 = 0$ plane, and $\omega_0 = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$ is the standard symplectic. We assume $W(x_1,x_2)$ is also entire and that $W(x_1,0)$ is a polynomial of degree 3 or 4 in x_1 with

$$W(0,0) = 0 \ , \quad (\partial W/\partial x_2)(x_1,0) \equiv 0. \tag{1.2}$$

When $\deg W(x_1,0) = 4$, we require the symmetry condition

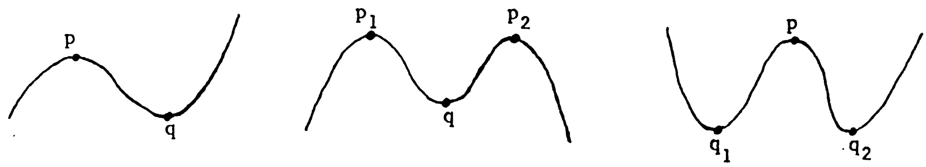
$$W(x_1,x_2) = W(-x_1,x_2) \quad \text{with} \quad (\partial^2 W/\partial x_1^2)(0,0) \neq 0. \tag{1.3}$$

The function $W(x_1,0)$ will possess at least one critical point $p = (x_1^*,0)$ such that

$$(\partial^2 W/\partial x_1^2)(p) < 0 \quad , \quad (\partial^2 W/\partial x_2^2)(p) > 0. \tag{1.4}$$

(Such a point will be a hyperbolic critical point of $W\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ since $(\partial^2 W/\partial x_1 \partial x_2)(p) = 0$ by (1.2).)

The above assumptions on W imply that the graphs of $W(x_1,0)$ for the two cases $\deg W(x_1,0) = 3$ or 4 are as in Figure 1. The variant of Figure 1(a) obtained by flipping this graph about the vertical axis can be converted back to Figure 1(a) by the linear symplectic transformation $(x_1,x_2,y_1,y_2) \longrightarrow (-x_1,x_2,-y_1,y_2)$ which preserves the above assumptions. The points labelled with p's and q's are critical points of $W(x_1,0)$, with the p's being hyperbolic critical points as in (1.4). The symmetry (1.3) implies q is the origin in Figure 1(b) and p is the origin in Figure 1(c). Without loss of generality we can assume q is the origin in Figure 1(a). By (1.2) we then see that in all cases the potential $W(x_1,x_2)$ lacks constant and linear terms in x_1, x_2.



(a) (b) (c)

$$\begin{cases} \deg W(x_1,0) = 3 \\ W(q) = 0 \\ W(p) = h^* > 0 \end{cases} \qquad \begin{cases} \deg W(x_1,0) = 4 \\ W(q) = 0 \\ W(p_1) = W(p_2) = h^* > 0 \end{cases} \qquad \begin{cases} \deg W(x_1,0) = 4 \\ W(p) = 0 \\ W(q_1) = W(q_2) < 0 \end{cases}$$

Figure 1

For Figures 1(a) and (b) let $0 < h < h^*$, and in Figure 1(c) take $h > 0$. Then (1.2) implies that there are periodic solutions

$$(1.5) \qquad \varphi_h(t) = (\pi_h(t),\ 0,\ \dot{\pi}_h(t),\ 0)$$

with energy h to the Hamiltonian vectorfield X_H associated to (1.1). Since $(\partial^2 W/\partial x_1 \partial x_2)(x_1,0) \equiv 0$ by (1,2), the linearized equations about $\varphi_h(t)$ decouple and take the form

$$(1.6) \qquad \begin{cases} \text{(a)}\ \ddot{\eta}_1 + (\partial^2 W/\partial x_1{}^2)(\pi_h(t),0)\eta_1 = 0 \\ \text{(b)}\ \ddot{\eta}_2 + (\partial^2 W/\partial x_2{}^2)(\pi_h(t),0)\eta_2 = 0\ . \end{cases}$$

Let $A(h) \in S\ell(2,\mathbb{R})$ be the linearized Poincaré map along π_h given by integrating the first order system equivalent to (1.6b). We can think of $A(h)$ as a linear symplectic map of the (x_2,y_2)-plane based at the origin of $\mathbb{R}^4$, and identify (1.6b) as the (reduced) normal variational equation (NVE) of the Introduction.

Note that the orbits π_h limit as $h \uparrow h^*$ to an orbit homoclinic to the critical point p in Figure 1(a), and in (b) to a pair of heteroclinic orbits connecting p_1 and p_2, whereas in Figure 1(c) the π_h limit as $h \downarrow 0$ to a pair of orbits homoclinic to p. In all cases the period of π_h tends to ∞ as h goes to its limiting value.

THEOREM 1. *Assume conditions* (1.1)-(1.4) *as summarized in Figure* 1. *Then the linearized Poincaré map* $A(h)$ *associated to* (1.6b) *is nonresonant except possibly on a set of measure zero of the energies* h *in the above intervals.*

PROOF. Theorems 2.1 and 2.4-2.8 of [5] show that trace $A(h)$ is a *non-constant* real analytic function of h in the above energy intervals when the critical points labelled by p's in Figure 1 are hyperbolic. The result then follows from this analyticity and the symplectic character of $A(h)$. (The connection to classical Sturmian oscillation theory is given in [5, Appendix B].)

Q.E.D.

2. APPLICATIONS: $\deg W(x_1,0) = 3$

Consider the potential

$$(2.1) \qquad W(x_1,x_2) = \sum_{2\le i+j\le 3} \gamma_{ij}\, x_1^{\,i}\, x_2^{\,j}$$

where the γ_{ij} are real constants with $\gamma_{30} \neq 0$. Assume that (1.2) and (1.4) hold for this potential (see Figure 1(a)), hence $\gamma_{11} = \gamma_{21} = 0$. Since the functions in (1.1) are entire, the Hamiltonian system can be complexified by replacing the variables $(x_1,x_2,y_1,y_2) \in \mathbb{R}^4$ by $(z_1,z_2,w_1,w_2) \in \mathbb{C}^4$ and letting the time $t \in \mathbb{C}$. Then at real energies $0 < h < h^*$ the first component $\pi_h(t)$ of the non-equilibrium solution (1.5) is an elliptic function with two independent (over the reals) periods $\omega_j(h)$, $j = 1,2$, and one pole $t_\infty = t_\infty(h)$ of order two which can be assumed to be *interior* to its period parallelogram (use curvilinear sides if necessary). We can take $\omega_1(h)$ real corresponding to the periodicity of $\pi_h(t)$ in real time. The meromorphic function $\pi_h(t)$ has a local expansion about the pole given by

$$(2.2) \qquad \pi_h(t) = a(t - t_\infty)^{-2} + \text{h.o.t.},$$

here h.o.t. denotes higher order terms. This orbit satisfies the following equation along the (complex) x_1-axis:

$$(2.3) \qquad \ddot{x}_1 + (\partial W/\partial x_1)(x_1,0) = \ddot{x}_1 + (2\gamma_{20}\, x_1 + 3\gamma_{30}\, x_1^{\,2}) = 0.$$

Putting (2.2) into (2.3) implies that the coefficient $a \neq 0$ is given by

$$(2.4) \qquad a = -2(\gamma_{30})^{-1},$$

where $\gamma_{30} \neq 0$ by assumption.

Now (1.6b) becomes

$$(2.5) \qquad \ddot{\eta}_2 + [2\gamma_{02} + 2\gamma_{12}\, \pi_h(t)]\eta_2 = 0.$$

In a neighborhood of t_∞ the equation (2.5) takes the following form (on setting $\gamma = \gamma_{12}(\gamma_{30})^{-1}$):

$$\ddot{\eta}_2 + [-4\gamma(t - t_\infty)^{-2} + \text{h.o.t.}]\eta_2 = 0. \tag{2.6}$$

THEOREM 2. *Let the potential* (2.1) *satisfy the conditions* (1.2) *and* (1.4) *as summarized in Figure* 1(a), *and set* $\gamma = \gamma_{12}(\gamma_{30})^{-1}$. *Then the Hamiltonian system* (1.1) *with potential* (2.1) *has no second integral* F *that is meromorphic in a neighborhood in* $\mathbb{C}^4$ *of the image* $\Gamma(h)$ *of* $\Psi_h(t)$ *and functionally independent of the Hamiltonian* H *in this neighborhood at energies* $h \in (0,h^*)$ *when* $(1 + 16\gamma) \neq (\text{odd integer})^2$, *except possibly on a set of measure zero of energies* h *in this interval.*

PROOF. Some of the excluded energies correspond to those in Theorem 1; the other excluded energies will be specified below. For all other energies the linearized Poincaré map $A(h)$ calculated along $\pi_h(t)$ for $t \in \mathbb{R}$ is nonresonant. Let $B(h)$ be the corresponding matrix obtained by analytic continuation along the other side of the period parallelogram of $\pi_h(t)$ of a fundamental matrix solution $\Psi(t,h)$ of (2.5) with $\Psi(0,h) = I$ (the 2×2 identity matrix) from $t = 0$ to $t = \omega_2(h)$. Then the commutator $C(h) = (A(h),B(h))$ can be interpreted as the result of analytic continuation in the complex t-plane of $\Psi(t,h)$ counterclockwise around the period parallelogram of $\pi_h(t)$ starting and ending at $t = 0$.

The eigenvalues of $C(h)$ can then be calculated by analytic continuation of $\Psi(t,h)$ in a positive sense around the regular singular point $t_\infty(h)$ of (2.5) as

$$\lambda_j = \exp(2\pi i\rho_j) \quad , \quad j = 1,2, \tag{2.7}$$

where the ρ_j are the roots of the indicial equation

$$\rho(\rho-1) - 4\gamma = 0 \tag{2.8}$$

of (2.6) (see [1, pp. 230-232]). (The expression (2.7) for the eigenvalues holds even when $(\rho_1-\rho_2)$ = integer.) Note that although the pole $t_\infty(h)$ depends on h, the eigenvalues λ_j are independent of h. If we further exclude those energies for which $C(h) = A^2(h)$ (recall from Theorem 1 that the eigenvalues of $A(h)$ change with h), we see that the solutions ρ to (2.8) force $\lambda_j \neq 1$ in (2.7) when $(1 + 16\gamma) \neq$ (odd integer)^2, hence $C(h) \neq I$, and the result follows from Ziglin's Theorem.

Q.E.D.

REMARK 1. (a) If F is entire or meromorphic on all of $\mathbb{C}^4$, then one does not need to exclude any energies in the interval $(0,h^*)$ in Theorem 2. For example, the complexification of entire real analytic F on $\mathbb{R}^4$ can be considered in this manner.

(b) As the following example shows, the flow of a Hamiltonian system can be integrable with independent real analytic (but not entire) integrals at some energies, but non-integrable at other energies. Let

(2.9) $$H(x_1,x_2,y_1,y_2) = (1/2)(y_1^2 + y_2^2) + (1/3)x_1^3 - x_1x_2^2.$$

Then it is known [6, Theorem 3.1] that at each energy $h > 0$ the flow of X_H admits a Smale horseshoe map and hence has no second real analytic integral independent of H (see [12, pp. 188–189] for the details on this conclusion). We now construct such an integral at each energy $h < 0$. Let $W(x_1,x_2) = [(1/3)x_1^3 - x_1x_2^2]$ be the potential, and let R be that region in the (x_1,x_2)-plane, containing the negative x_1-axis, with $W(x_1,x_2) < 0$. Given a solution $\varphi(t) = (x_1(t), x_2(t), y_1(t), y_2(t))$ with energy $h < 0$ whose (x_1,x_2)-plane projection lies in R, we have $y_1 = y_1(t,J)$ real analytic in t and the initial conditions $J \in \mathbb{R}^4$. For a given J_0 there is a unique time t_0 such that $y_1(t_0,J_0) = 0$ as follows from the values of $-\nabla W$ in R. Since $(dy_1/dt) = -(\partial W/\partial x_1) \neq 0$ in R, we have for J near J_0 a unique analytic $t = t(J)$ such that $y_1(t(J),J) \equiv 0$. We define the integral F in a neighborhood of J_0 by $F(J) = x_1(t(J),J)$, and note that when $h = 0$ we have $F(J) \longrightarrow 0$ and $t(J) \longrightarrow \infty$ as $J \longrightarrow (x_1^0,0,y_1^0,0)$ where $x_1^0 < 0$. F has a real analytic extension to all initial conditions J with $H(J) < 0$ on invoking the symmetry of W under rotations of the (x_1,x_2)-plane through angles $(2\pi/3)$ and $(4\pi/3)$ to pick up the other regions where $W(x_1,x_2) < 0$. The integral F has a continuous (but not real analytic) extension to the flow at energy $h = 0$. To demonstrate the independence of F from H when $h < 0$, let $\lambda(\varepsilon) = (\varepsilon,0,0,[h - W(\varepsilon,0)]^{1/2})$ be a curve in $\mathbb{R}^4$ with $\varepsilon < 0$. Then $(H \circ \lambda)(\varepsilon) = h$ and $(F \circ \lambda)(\varepsilon) = \varepsilon$ imply that ∇H is perpendicular to $(d\lambda/d\varepsilon)$ whereas ∇F is not perpendicular to $(d\lambda/d\varepsilon)$. Thus ∇H and ∇F are independent vectors along $\lambda(\varepsilon)$. The above arguments can be adapted to the flow in the unbounded components of $H^{-1}(\{h\})$ of many other Hamiltonians H.

We now apply Theorem 2 to the 2-parameter family of Hénon-Heiles Hamiltonians of the form (1.1) with potential

$$(2.10) \qquad W(x_1,x_2) = (1/2)(x_1^2 + x_2^2) + (a/3)x_1^3 + bx_1 x_2^2,$$

where $a \neq 0$ and b are real. The case $a = 1$ and $b = -1$ was originally investigated in [8].

COROLLARY 1. *For each fixed value of* $a \neq 0$ *and* b *in* (2.10) *we have the conclusions of Theorem 2 and Remark 1(a) at energies* $0 < h < h^*$ (*where* $h^* = h^*(a,b)$ *is specified in the proof below*) *provided* $(b/a) \neq 0$, $1/2$, $1/6$, $3/4$, 1.

PROOF. (1) Assume $(b/a) < (1/2)$. Then the critical point $p = (-a^{-1},0)$ satisfies $(\partial^2W/\partial x_1^2)(p) = -1$ and $(\partial^2W/\partial x_2^2)(p) = 1 - 2(b/a)$ and hence is a hyperbolic critical point for (2.10) with positive energy $h^* = (6a^2)^{-1}$. Then $\gamma = \gamma_{12}(\gamma_{30})^{-1} = (3b/a)$ and one need only check cases on the condition $(1 + 16\gamma) \neq (\text{odd integer})^2$.

(2) Assume $(b/a) > (1/2)$. As (b/a) passes $(1/2)$ two gradient lines $x_2 = \pm\mu x_1$ of (2.10) with $\mu = [2 - (a/b)]^{1/2}$ bifurcate off the x_1-axis. The restriction $W(x_1,\mu x_1)$ has critical points at $x_1 = -(2b)^{-1}$ and $x_1 = 0$. It is easily checked that $p = -(2b)^{-1} \cdot (1,\mu)$ is a hyperbolic critical point of (2.10) with positive energy $h^* = [1 - a/(3b)]\cdot(8b^2)^{-1}$ for $0 < (a/b) < 2$. The transformation $T\colon (x_1,x_2) \longrightarrow (\tilde{x}_1,\tilde{x}_2)$, given by

$$\begin{cases} \tilde{x}_1 = (1 + \mu^2)^{-1/2}\,(x_1 - \mu x_2) \\ \tilde{x}_2 = (1 + \mu^2)^{-1/2}\,(\mu x_1 + x_2)\ , \end{cases}$$

rotates the x_1-axis to the line $\tilde{x}_2 = \mu\tilde{x}_1$. A direct calculation for the transformed potential $(W \circ T)(x_1,x_2)$ yields $\gamma = 3(a-b)/(2b)$. Again, one need only check cases on the condition $(1 + 16\gamma) \neq (\text{odd integer})^2$.

Q.E.D.

REMARK 2. (a) Ito's method in [9] also shows that $(b/a) = (3/4)$ is a non-integrable case. He directly computes that trace $A(h)$ varies with h. We cannot obtain this case since the critical point p in (1) above is not hyperbolic at this parameter ratio. The cases $(b/a) = 0$, $1/6$, and 1 are known to be integrable, whereas $(b/a) = (1/2)$ is anomalous with computer evidence suggesting non-integrability [2].

(b) By Theorem 2 one can easily generalize the above corollary to include arbitrary coefficients in the quadratic terms of (2.10) provided conditions (1.2) and (1.4) are satisfied (see [2]). The non-integrability of (2.10) with $a = 1$ and $b = -1$ at energies $h > (1/6)$ (except possibly at a discrete set of such energies) was shown in [6] (see also [4]).

3. APPLICATIONS: deg $W(x_1,0) = 4$

Consider the potential

$$W(x_1,x_2) = \sum_{2\le i+j\le 4} \delta_{ij}\, x_1^{\,i}\, x_2^{\,j} \tag{3.1}$$

where the δ_{ij} are real constants with $\delta_{40} \neq 0$. Assume that (1.2)-(1.4) hold for this potential (see Figures 1(b) and (c)), hence $\delta_{11} = \delta_{21} = \delta_{12} = \delta_{31} = \delta_{13} = 0$ and $\delta_{20} \neq 0$. The following discussion is given at energies $0 < h < h^*$ for the case that the graph of $W(x_1,0)$ is given by Figure 1(b). A completely analogous discussion with identical conclusions holds for the energy range $h > 0$ in the case of Figure 1(c). Thus at energies $0 < h < h^*$ w.r.t. Figure 1(b) we see that the Hamiltonian system (1.1) with potential (3.1), complexified as in Section 2, admits non-equilibrium solutions (1.5) with $\pi_h(t)$ an elliptic function. π_h will have two independent periods $\omega_j(h)$, $j = 1,2$, and two poles, each of order one, which can be assumed to be interior to its period parallelogram. We can take $\omega_1(h)$ real corresponding to the periodicity of $\pi_h(t)$ in real time. The meromorphic function $\pi_h(t)$ has a local expansion about one of these poles $t_\infty = t_\infty(h)$ given by

$$\pi_h(t) = b(t - t_\infty)^{-1} + \text{h.o.t.} \tag{3.2}$$

Reasoning as in Section 2 the orbit $\pi_h(t)$ satisfies the analogue of (2.3) from which we compute that $b \neq 0$ satisfies

$$b^2 = -(2\delta_{40})^{-1}, \tag{3.3}$$

where $\delta_{40} \neq 0$ by assumption. Note that the two solutions for b correspond to the two poles and reflect the fact that the sum of the residues of an elliptic function in its period parallelogram must be zero.

The analogue of (2.5) is

$$\ddot{\eta}_2 + [2\delta_{02} + 2\delta_{22}\, \pi_h^{\,2}(t)]\eta_2 = 0. \tag{3.4}$$

Then (3.4) has a regular singular point at each of the two poles and takes the following form in a neighborhood of such a pole (on setting $\delta = -2\delta_{22}\, b^2 = \delta_{22}(\delta_{40})^{-1}$):

$$\ddot{\eta}_2 + [-\delta(t - t_\infty)^{-2} + \text{h.o.t.}]\eta_2 = 0. \tag{3.5}$$

Recall from Section 1 that our linearized Poincaré map along $\pi_h(t)$ could be considered as a linear map of the (x_2, y_2)-plane based at the origin of $\mathbb{R}^4$. We therefore take $\pi_h(0) = 0$. The symmetry (1.3) (see Figures 1(b) and (c)) then implies that $\pi_h(t + \omega_1(h)/2) = \pi_h(-t) = -\pi_h(t)$ for $t \in \mathbb{R}$, and hence for $t \in \mathbb{C}$. Thus the coefficients in (3.4) have basic real period $\omega_1(h)/2$.

THEOREM 3. *Let the potential* (3.1) *satisfy the conditions* (1.2)-(1.4) *as summarized in Figures 1(b) and (c) and assume* K *in* (1.1) *satisfies* $K(x_1,x_2,y_1,y_2) = K(-x_1,x_2,-y_1,y_2)$. *Set* $\delta = \delta_{22}(\delta_{40})^{-1}$. *Then the Hamiltonian system* (1.1) *with potential* (3.1) *has no second integral* F *that is meromorphic in a neighborhood in* $\mathbb{C}^4$ *of the image* $\Gamma(h)$ *of* $\varphi_h(t)$ *and functionally independent of the Hamiltonian H in each of the appropriate energy intervals for Figures 1(b) and (c) (as specified above) provided* $(1 + 4\delta) \neq (\text{odd integer})^2$, *except possibly on a set of measure zero of energies* h *in the respective intervals.*

PROOF. We use the notation in the proof of Theorem 2 except that $A(h)$ is now the analytic continuation of the fundamental matrix solution $\Psi(t,h)$ of (3.4) along the real time axis from $t = 0$ to $t = \omega_1(h)/2$. As before, $B(h)$ is the analytic continuation of $\Psi(t,h)$ along the other side of the period parallelogram of $\pi_h(t)$ from $t = 0$ to $t = \omega_2(h)$. Then the commutator $C(h) = (A(h),B(h))$ can be interpreted as the result of analytic continuation in the complex t-plane of $\Psi(t,h)$ counterclockwise around one-half of the period parallelogram of $\pi_h(t)$ starting and ending at $t = 0$.

The Hamiltonian H of (1.1) is invariant under the $\mathbb{Z}_2$-action $(x_1,x_2,y_1,y_2) \longrightarrow (-x_1,x_2,-y_1,y_2)$, and similarly for the complexified Hamiltonian. Ziglin in [16, Sections 1.5 and 4] has given the analogue of his theorem for the reduced Hamiltonian system $\hat{H}$ obtained through reduction by such a symplectic symmetry. He has shown that if H is integrable then so is $\hat{H}$. Applying Ziglin's Theorem to $\hat{H}$ is equivalent to showing that the commutator $C(h)$ calculated above does not equal I or $A^2(h)$. Now the eigenvalues of $C(h)$ can be calculated by analytic continuation of $\Psi(t,h)$ in a positive sense around one of the two regular singular points of (3.4). These eigenvalues are independent of h and are given by (2.7) where the ρ_j are now the roots of the indicial equation

$$\rho(\rho-1) - \delta = 0 \tag{3.6}$$

of (3.5). Excluding the same types of energies as were excluded in the proof of Theorem 2 (one can check from [5] that the eigenvalues of this new $A(h)$

change as specified in Theorem 1), the result follows from Ziglin's Theorem in the same manner as in the proof of Theorem 2.

Q.E.D.

We illustrate Theorem 3 with an example previously discussed in [3], [4, Section 8(b)], and [5, Section 6(c)]. The potential

$$W(x_1,x_2) = (1/2)(x_1^2 + x_2^2) - (1/2)x_1^2 x_2^2 \tag{3.7}$$

has gradient lines $x_2 = \pm x_1$. Along the line $x_2 = x_1$ there are two hyperbolic critical points $p_\pm = \pm(1,1)$ at energy $h^* = (1/2)$, and hence we have Figure 1(b) along this line. Let T be a rotation of the (x_1,x_2)-plane through an angle $(\pi/4)$ and call the new coordinates (x_1,x_2) again. Then

$$(W \circ T)(x_1,x_2) = (1/2)(x_1^2 + x_2^2) - (1/8)(x_1^4 - 2x_1^2 x_2^2 + x_2^4) \tag{3.8}$$

satisfies (1.2)-(1.4), and $\delta = \delta_{22}(\delta_{40})^{-1} = -2$ implies $(1 + 4\delta) = -7 \neq (\text{odd integer})^2$. Hence the system (1.1) with potential (3.7), and K symmetric as in Theorem 3, is non-integrable in the sense of Theorem 3 (recall Remark 1(a) in Section 2) at energies $0 < h < (1/2)$. For non-integrability results at energies $h > (1/2)$, see [3].

REFERENCES

[1] G. Birkhoff and G-C. Rota, "Ordinary Differential Equations," third edition, John Wiley and Sons, New York, 1978.

[2] Y.F. Chang, M. Tabor, and J. Weiss, Analytic structure of the Hénon-Heiles Hamiltonian in integrable and nonintegrable regimes, J. Math. Phys., 23 (1982), 531-538.

[3] R.C. Churchill, G. Pecelli, S. Sacolick, and D.L. Rod, Coexistence of stable and random motion, The Rocky Mountain J. of Math., 7 (1977), 445-456.

[4] R.C. Churchill, G. Pecelli, and D.L. Rod, A survey of the Hénon-Heiles Hamiltonian with applications to related examples, in "Como Conference Proceedings on Stochastic Behavior in Classical and Quantum Hamiltonian Systems" (G. Casati and J. Ford, Eds.), 76-136, Springer Lecture Notes in Physics, Vol. 93, Springer-Verlag, New York, 1979.

[5] R.C. Churchill, G. Pecelli, and D.L. Rod, Stability transitions for periodic orbits in Hamiltonian systems, Archive for Rational Mechanics and Analysis, 73 (1980), 313-347.

[6] R.C. Churchill and D.L. Rod, Pathology in dynamical systems (III); analytic Hamiltonians, J. Differential Equations, 37 (1980), 23-38.

[7] C. Conley, On the ultimate behavior of orbits with respect to an unstable critical point (I), J. Differential Equations, 5 (1969), 136-158.

[8] M. Hénon and C. Heiles, The applicability of the third integral of motion; some numerical experiments, Astronom. J., 69 (1964), 73-79.

[9] Hidekazu Ito, Non-integrability of Hénon-Heiles system and a theorem of Ziglin, Kodai Math. J., 8 (1985), 120-138.

[10] Hidekazu Ito, A criterion for non-integrability of Hamiltonian systems with nonhomogeneous potentials, J. Appl. Math. and Physics, 38 (1987), 459-476.

[11] J. Moser, On the generalization of a theorem of A. Liapounoff, Commun. Pure Appl. Math., 11 (1958), 257-271.

[12] J. Moser, "Stable and Random Motions in Dynamical Systems," Annals of Mathematics Studies, Vol. 77, Princeton Univ. Press, Princeton, New Jersey, 1973.

[13] D.L. Rod, On an example of Ziglin in Hamiltonian dynamics, in "Proceedings of the 1986 Annual Seminar on Oscillation, Bifurcation, and Chaos," (F.V. Atkinson, W.F. Langford, and A.B. Mingarelli, Eds.), 599-608, Conference Proceedings of the Canadian Mathematical Society, Volume 8, American Mathematical Society, Providence, Rhode Island, 1987.

[14] C.L. Siegel, "Topics in Complex Function Theory," Vol. 1 (Elliptic Functions and Uniformization Theory), Wiley-Interscience, New York, 1969.

[15] C.L. Siegel and J.K. Moser, "Lectures on Celestial Mechanics," Springer-Verlag, New York, 1971.

[16] S.L. Ziglin, Branching of solutions and nonexistence of first integrals in Hamiltonian mechanics: I and II, Functional Anal. Appl., 16 (1982), 181-189, and 17 (1983), 6-17.

DEPARTMENT OF MATHEMATICS AND STATISTICS
THE UNIVERSITY OF CALGARY
CALGARY, ALBERTA
CANADA T2N 1N4

Recent Titles in This Series

(*Continued from the front of this publication*)

133 **Alan Adolphson, Steven Sperber, and Marvin Tretkoff, Editors,** p-Adic methods in number theory and algebraic geometry, 1992
132 **Mark Gotay, Jerrold Marsden, and Vincent Moncrief, Editors,** Mathematical aspects of classical field theory, 1992
131 **L. A. Bokut′, Yu. L. Ershov, and A. I. Kostrikin, Editors,** Proceedings of the International Conference on Algebra Dedicated to the Memory of A. I. Mal′cev, Parts 1, 2, and 3, 1992
130 **L. Fuchs, K. R. Goodearl, J. T. Stafford, and C. Vinsonhaler, Editors,** Abelian groups and noncommutative rings, 1992
129 **John R. Graef and Jack K. Hale, Editors,** Oscillation and dynamics in delay equations, 1992
128 **Ridgley Lange and Shengwang Wang,** New approaches in spectral decomposition, 1992
127 **Vladimir Oliker and Andrejs Treibergs, Editors,** Geometry and nonlinear partial differential equations, 1992
126 **R. Keith Dennis, Claudio Pedrini, and Michael R. Stein, Editors,** Algebraic K-theory, commutative algebra, and algebraic geometry, 1992
125 **F. Thomas Bruss, Thomas S. Ferguson, and Stephen M. Samuels, Editors,** Strategies for sequential search and selection in real time, 1992
124 **Darrell Haile and James Osterburg, Editors,** Azumaya algebras, actions, and modules, 1992
123 **Steven L. Kleiman and Anders Thorup, Editors,** Enumerative algebraic geometry, 1991
122 **D. H. Sattinger, C. A. Tracy, and S. Venakides, Editors,** Inverse scattering and applications, 1991
121 **Alex J. Feingold, Igor B. Frenkel, and John F. X. Ries,** Spinor construction of vertex operator algebras, triality, and $E_8^{(1)}$, 1991
120 **Robert S. Doran, Editor,** Selfadjoint and nonselfadjoint operator algebras and operator theory, 1991
119 **Robert A. Melter, Azriel Rosenfeld, and Prabir Bhattacharya, Editors,** Vision geometry, 1991
118 **Yan Shi-Jian, Wang Jiagang, and Yang Chung-chun, Editors,** Probability theory and its applications in China, 1991
117 **Morton Brown, Editor,** Continuum theory and dynamical systems, 1991
116 **Brian Harbourne and Robert Speiser, Editors,** Algebraic geometry: Sundance 1988, 1991
115 **Nancy Flournoy and Robert K. Tsutakawa, Editors,** Statistical multiple integration, 1991
114 **Jeffrey C. Lagarias and Michael J. Todd, Editors,** Mathematical developments arising from linear programming, 1990
113 **Eric Grinberg and Eric Todd Quinto, Editors,** Integral geometry and tomography, 1990
112 **Philip J. Brown and Wayne A. Fuller, Editors,** Statistical analysis of measurement error models and applications, 1990
111 **Earl S. Kramer and Spyros S. Magliveras, Editors,** Finite geometries and combinatorial designs, 1990
110 **Georgia Benkart and J. Marshall Osborn, Editors,** Lie algebras and related topics, 1990
109 **Benjamin Fine, Anthony Gaglione, and Francis C. Y. Tang, Editors,** Combinatorial group theory, 1990
108 **Melvyn S. Berger, Editor,** Mathematics of nonlinear science, 1990
107 **Mario Milman and Tomas Schonbek, Editors,** Harmonic analysis and partial differential equations, 1990
106 **Wilfried Sieg, Editor,** Logic and computation, 1990

(See the AMS catalog for earlier titles)

This page intentionally left blank

ISBN 0-8218-5086-5

9 780821 850862